2022

Zero Textbook Cost Syllabus for CIS/STA 3920 (Data Mining for Business Analyitics)

Vinayak Javaly
CUNY Bernard M Baruch College

Zicklin School of Business – Baruch College – CUNY
Department of Computer Information Systems
CIS/STA 3920 PMWA – Data Mining for Business Analytics

Instructor: **Vinayak Javaly**	**Email:** **vinayak.javaly@baruch.cuny.edu**
Class Meeting Time: **Mon/Wed 5:50-7:05 PM**	**Classroom:** **B – Vert 5-175**
Office: **Classroom**	**Office Hours:** **Mon. 7:05-7:35 PM** **Wed. 5:20-5:50 PM** **and by appt.**

Overview

Data Mining is the process by which useful information is extracted from large amounts of data. This course is designed to provide students with the necessary tools and techniques to perform data mining and business analytics. This course is intended as an introductory module targeted at individuals who plan to work with data (modeling, data management) as well as towards those who will work with data scientists. While the course will primarily focus on modeling and evaluation, it will also include data preparation and examination. Modeling techniques covered include decision trees, clustering, and other methods. Emphasis will be placed on the entire context surrounding data mining, which includes the business problem, data processing, modeling, evaluation and deployment.

Course Objectives

- Learn commonly used data mining algorithms
- Understand how and why businesses use data mining
- Learn how to transform real-world data into datasets for modeling
- Use Python and tools to work with data and train different types of models
- Learn how to select a modeling technique to match the problem to be solved
- Become aware of the power and peril of data mining, machine learning & AI

BBA Learning Goals

- Analytical Skills
- Technological Skills
- Communication Skills: Oral and Written
- Civic Awareness and Ethical Decision-Making
- Global Awareness

Course Prerequisites

- CIS 2200 and STA 2000

Course Materials

- There is no required textbook for this course.
- Links to publicly available newspaper/magazine/journal articles, blog posts, research papers, YouTube videos, etc. will be used to supplement the lectures and slides.
    - **NOTE: It is the student's responsibility to review these references and treat this material the same as they would a textbook.**
- All class slides, notes and other reference materials will be posted in the course Blackboard site.

Course Software

- The course will use only free open-source software:
  - Python, pandas, numpy, scipy, scikit-learn, matplotlib, seaborn, Jupyter notebooks, etc.
  - These software packages are available for free, distributed under the permissive software license, and actively supported by their respective open-source communities.
- Students will download Anaconda Distribution, selecting the appropriate installer file for their computer's operating system.
  - Anaconda, Inc. packages and supports the popular open-source software listed above and provides a **Free** software tier for students, academics, and hobbyists.
    - Jupyter is a free open-source development environment (IDE) included with Anaconda.

Documentation Links

- Official Python.org site
- Anaconda - Python Data Science Platform
- Official Pandas site
- scikit-learn - Machine Learning in Python

Optional Reference Materials

Students are **not required** to purchase any of these textbooks. They are supplemental references for those students seeking a deeper understanding of the topics covered during the semester.

- Introducing Python: Modern Computing in Simple Packages 2nd Edition
- Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython 2nd Edition
- Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking 1st Edition
- Introduction to Data Mining 1st Edition
- Data Mining: Practical Machine Learning Tools and Techniques (Morgan Kaufmann Series in Data Management Systems) 4th Edition

Course Communications

- Blackboard will be used to post all courses materials and submit evaluation tasks.
- Students should frequently check Blackboard for any updates.
- Students should confirm that their Baruch email address is in blackboard so that announcements from Blackboard will be delivered promptly to your Baruch email account.

Course Methodology and Evaluation

- The course is structured around a combination of in-class lectures and code walkthroughs, individual homework assignments, and a group project.
- Students will be graded as follows:

Homework Assignments (3)	30 pts
Group Project	30 pts
Midterm Exam	15 pts
Final Exam	25 pts
Total	**100 pts**

- Late submission will **NOT** be accepted for any homework assignment or the group project.

Code Walkthroughs

- Throughout the semester, there will be approximately 25 in-class Python code walkthroughs delivered via Jupyter notebooks.
- The professor will use these code walkthroughs to reinforce course material presented in lectures.
- The Jupyter notebooks and related data files will be available through Blackboard.

Exercises

- There will be 2 after-class exercises (not graded) using Python.
- Though these exercises are not graded, students are encouraged to complete them in order to be ready for the homework assignments and group project.
- These exercises can be completed individually or jointly with other students.
- Instructions for the exercises will be made available through blackboard.

Homework Assignments

- There will be 4 after-class homework assignments using Python.
    - Students are required to submit 3 homework assignments.
    - If students submit all 4 homework assignments, the highest 3 grades will be used.
- These homework assignments must be completed individually.
- Instructions for the homework assignments will be made available through blackboard.

Group Project

- There will be 1 group project using Python.
- The project must be completed in small groups (4 people).
- The project must be presented in person in-class during week 14.
- Instructions for the project, including the project grading rubric, will be made available through blackboard.
- The last 2 or 3 classes (classes 26-28) are reserved for project presentations. The actual schedule will be determined after class rosters are finalized.

Exams

- There will be 2 in-class exams: a midterm exam in the middle of the semester, and a final exam during finals exam week.
- The exams will be concepts based, in short answer format.
- The exams will cover material from the course lectures, slides, reading assignments, in-class code walkthroughs, exercises, homework assignments and the group project.
- The exams will be closed-book, no-notes, no computer, no-phone.
- The instructor reserves the right to curve the scores for the exams, if deemed necessary.
- Exams will be held during specific time slots in-person / in-class. **Please check the tentative course schedule for exam time and make the necessary plans as early as possible.**

Students with Disabilities

Baruch has a continuing commitment to providing reasonable accommodations for students with disabilities. Like so many things this fall, the need for accommodations and the process for arranging them have been altered by COVID-19 and the safety protocols currently in place. Students with disabilities who may need some accommodation to fully participate in this class should contact Student Disability Services as soon as possible at disability.services@baruch.cuny.edu or call 646/312-4590. Please refer to Baruch College Student Disability Services.

Baruch College Counseling

At Baruch, we acknowledge that as a student, you are balancing many demands. During the semester, if you start to experience personal difficulties or stressors that are interfering with your academic performance or day to day functioning, please consider seeking free and confidential support at the **Baruch College Counseling Center**. For more information or to make an appointment, please visit their website at Baruch College Counseling Center or call 646-312-2155. If it's outside of business hours (Monday-Friday 9-5pm) and you need immediate assistance, please call 1-888-NYC-WELL (888-692-9355). If you are concerned about one of your classmates, please share that concern by filling out a Campus Intervention Team form at Campus Intervention Team (CIT).

Final Letter Grades

Letter grades are calculated automatically according to the Official Grading System of Baruch College. No round up for the final score.

Letter Grade	Final Score
A	93.0–100
A-	90.0–92.9
B+	87.0–89.9
B	83.0–86.9
B-	80.0–82.9
C+	77.0–79.9
C	73.0–76.9
C-	70.0–72.9
D+	67.0–69.9
D	60.0–66.9
F	below 60.0

Course Policies

- It is very important to attend class regularly.
- During class, students should refrain from engaging in any kind of disruptive behavior. Disruptive behavior can negatively affect the class environment as well as the educational experience for fellow students. Any type of disruptive behavior will be dealt according to the *Classroom Management Guidelines* established by the Baruch College.
- Late submission will not be accepted for any homework assignments or any component of the project.
- Students have the right to ask why they received a certain grade on an assignment within one week after the grade is posted. If you received a grade in error I will correct it.
- Let me know about any problems or issues (such as missing class, long term illnesses, job related problems, etc.) as soon as possible. *If you come to me at the end of the semester about a problem you had earlier in the semester, I will not be able to help.*
- Show up **ON TIME** for exams. No make-ups will be given for missed exams.
- The final exam must be taken in the time slot posted in the college bulletin.
- Behavior during exams is expected to conform to Baruch College guidelines. Any form of cheating or communications with other students or any other incident of improper behavior will be dealt according to the guidelines established by the College.

Academic Integrity Statement

The CIS Department fully supports Baruch College's policy on Academic Honesty, which states, in part:

> "*Academic dishonesty is unacceptable and will not be tolerated. Cheating, forgery, plagiarism, and collusion in dishonest acts undermine the college's educational mission and the students' personal and intellectual growth. Baruch students are expected to bear individual responsibility for their work, to learn the rules and definitions that underlie the practice of academic integrity, and to uphold its ideals. Ignorance of the rules is not an acceptable excuse for disobeying them. Any student who attempts to compromise or devalue the academic process will be sanctioned.* "

Academic sanctions in this class will range from an F on the assignment to an F in this course. A report of suspected academic dishonesty will be sent to the Office of the Dean of Students. Additional information and definitions can be found at Baruch College Academic Honesty.

Religious Accommodations

Please refer to Baruch College Religious Accommodations.

Assessment

BBA Learning Goals	Significant Part of Course	Moderate Part of Course	Minimal Part of Course	Not Part of Course
Analytical skills	☒	☐	☐	☐
Technological skills	☒	☐	☐	☐
Oral communication skills	☐	☒	☐	☐
Written communication skills	☐	☒	☐	☐
Civic awareness and ethical decision-making	☐	☐	☒	☐
Global awareness	☐	☐	☒	☐

Course Learning Goals	BBA learning goals	Delivery Mechanism
Learn commonly used data mining algorithms	Analytical skills Technological skills	Class lectures and slides Code walkthroughs
Understand how and why businesses use data mining	Analytical skills Civic awareness and ethical decision-making Global awareness	Class lectures and slides Newspaper/magazine/journal articles, blog posts, research papers, YouTube videos, etc.
Learn how to transform real-world data into datasets for modeling	Analytical skills Technological skills	Class lectures and slides Code walkthroughs Exercises Homework assignments Group project
Use Python and tools to work with data and train different types of models	Analytical skills Technological skills	Code walkthroughs Exercises Homework assignments Group project
Learn how to select a modeling technique to match the problem to be solved	Analytical skills Technological skills	Code walkthroughs Exercises Homework assignments Group project
Be able to interpret results to your audiences without relevant background	Analytical skills Technological skills Oral communication skills Written communication skills	Group Project • In-class project presentation • Final project report
Become aware of the power and peril of data mining, machine learning & AI	Civic awareness and ethical decision-making Global awareness	Class lectures and slides Data mining in real life case studies

Course Schedule

	Primary Topics	Supporting Topics
Class 1 (8/29)	Course overview Introduction to data mining	Class development environment review Industry, Careers
Class 2 (8/31)	Key concepts	Big data Data science Artificial intelligence
Class 3 (9/7)	Python	Jupyter notebooks

Class 4 (9/12)	Pandas	scikit-learn
Class 5 (9/14)	Data mining terminology Introduction to machine learning	Training, validation, test sets
Class 6 (9/19)	Supervised learning Linear regression	Regression analysis
Class 7 (9/21)	Linear regression (continued) Logistic regression	Calculating modeling errors Regularization
Class 8 (9/28)	Logistic regression (continued) Model performance metrics	Cost functions Gradient descent
Class 9 (9/29)	Exploratory data analysis Data preparation	Plotting data Encoding variables
Class 10 (10/3)	Data preparation (continued)	Outliers Data transformations
Class 11 (10/12)	Distance functions K-nearest neighbors (KNN)	KNN alternatives
Class 12 (10/17)	Text mining	Word clouds Word embeddings
Class 13 (10/19)	Midterm exam	
Class 14 (10/24)	Recommender systems	Collaborative filtering
Class 15 (10/26)	Recommender systems (continued)	Collaborative filtering
Class 16 (10/31)	Recommender systems (continued)	Collaborative filtering Matrix factorization Singular value decomposition
Class 17 (11/2)	Recommender systems (continued)	Content-based filtering
Class 18 (11/7)	Decision tree	Overfitting Dimensionality reduction Principal Component Analysis
Class 19 (11/9)	Decision tree (continued) Ensemble methods	Hyperparameters Cross validation
Class 20 (11/14)	Random forest	Bootstrapping
Class 21 (11/16)	Support vector machine (SVM)	SVM kernel functions
Class 22 (11/21)	Unsupervised learning Clustering	Contrast with supervised learning
Class 23 (11/23)	Clustering – K-Means Association rule mining	Hierarchical clustering DBSCAN
Class 24 (11/28)	Market basket analysis	Apriori algorithm
Class 25 (11/30)	Naïve Bayes	Biases in machine learning
Class 26 (12/5)	Additional topics OR Project presentations	
Class 27 (12/7)	Project presentations	
Class 28 (12/12)	Project presentations	
Class 29 (TBD)	Final exam	