ITCS 6156/8156, DSBA 6156: Applied Machine Learning

Spring 2022

Siddharth Krishnan

Contact Information:

Email:	skrishnan at uncc dot edu
Office:	Woodward Hall 403B
Office Hours:	Tuesdays 4:00 pm – 5:00 pm via Zoom (check Canvas for details)

TA:	TBA
Email:	TBA
Office Hours:	TBA

Course Description:

Prerequisite: ITCS 6150 & some programming experience, preferably in Python or permission of the instructor.

The era of Big Data has arrived. From healthcare to retail to manufacturing to e-commerce, big data is now pervasive. Technological advances in data collection have resulted in massive datasets that demand methods for analytics & visualization. It is not a far stretch of the imagination to claim that most IT professionals today, to be successful, must have a working understanding of data science. Therefore, in this course we will introduce some of the fundamental concepts of data science/analytics that will catalyze the student's skills in gleaning actionable insights using techniques (taught in the course) from mining massive datasets, machine learning, and data visualization. The class discusses how to apply some fundamental methods like decision trees, bayesian classifiers, SVMs, random forests, and some advanced techniques like semi-supervised learning, neural networks, and deep learning on real-world datasets, including data preparation, model selection, and evaluation. We will learn a range of applications like social network analysis, natural language processing, sports analytics, time-series analytics, and automated machine learning. Apart from applying models, we will also discuss software development tools and industry best practices relevant to productionizing machine learning models, particularly how to develop scalable models that can be deployed on a cloud infrastructure. Our class will include guest lectures from data scientists from Bank of America and ESPN. Prior iterations of the class have resulted in successful IEEE research publications as well (read: Extracting Cryptocurrency Price Movements From the Reddit Network Sentiment) (Links to an external site.).

Objectives of the Course:

Upon completion of this course, students should be able to:

- Tackle practical programming issues underlying storing, retrieving, accessing, and processing large datasets.
- Implement machine learning methods and apply them to large scale datasets.
- Analyze real-world datasets to develop predictive and descriptive applications.

Schedule of Topics, Exams & Presentations:

We will cover the following topics: (**tentative**)

- Matplotlib and Visualization
- Data Preprocessing
- Supervised Learning, Model Complexity, and Model Validation
- Linear Models – Regression, Classification, SVMs
- Trees, forests, and ensembles
- Gradient descent and boosting
- Calibration and imbalanced data
- Model evaluation, model interpretation, and feature selection
- Parameter tuning and automatic machine learning
- Dimensionality reduction, clustering, and mixture models
- Semi-supervised Learning
- Recommender systems
- Business applications of predictive modeling and mining health records
- Natural language processing, graph mining, and social media analytics
- Word and document embeddings
- Neural Networks
- Graph representational learning and graph convolutional networks
- Time series analytics
- Sports analytics, and other topics

Instructional Method:

Materials presented in this course will be covered through lectures, video-lectures, quizzes, homework assignments, and a project. Most of the concepts will be covered through the use case-based hands-on experience. In the end, students will have both the theoretical understanding of machine learning concepts and concrete experience of putting such concepts and principles into practice.

Textbooks:

There is **no required textbook** for the class, however the following reference books are *highly recommended*:

1. **"*Hands-On Machine Learning with Scikit-Learn & TensorFlow" 1st Edition, Aurelien Geron, O'Reilly Publishing, 2017***
2. *"Applied Predictive Modeling"*, Kuhn, Max and Johnson, Kjell, Springer 2016 (e-book available through library login)
3. "*Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*" 1st Edition, by Wes McKinney, O'Reilly Publishing, 2013 (latest edition is also ok)
4. ***"Mining Massive Datasets"* by Jure Leskovec, Anand Rajaraman, Jeffery Ulman, 2nd Edition, Cambridge University Press, 2015**
5. *"Machine Learning: A Probabilistic Perspective"* Murphy, Kevin, MIT Press, 2012
6. *"Deep Learning"* Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron, MIT Press, 2016

Course Requirements and Grading Policy:

Group project	35%
Homeworks (4-5)	50%
Quizzes	10%
Class Discussion	5%

Group Project Grading Breakdown:

1st written submission – project description & presentation	25%
2nd written submission – method selection & preliminary results	35%
Final write-up & project presentation	40%

Additional Policies:

Attendance:

Students are expected to attend all class meetings and to arrive before the class starts. Class topics are integrated, with each week building on prior weeks. Failure to attend or to arrive on time can adversely affect both individual performance, ability to contribute to the group project, and the earned letter grade. If a student misses a class due to work or other reasons, it is their responsibility to get notes from peers; instructors do not hold extra repeat class sessions.

Grade Discussions:

The instructor and TA will discuss grades only in-person or via Zoom (and not via e-mail) and only with the student (not with parents, spouses, etc). Office hours are listed in the syllabus.

Academic Integrity:

All students are expected to adhere to the UNC Charlotte Code of Student Academic Integrity (http://legal.uncc.edu/policies/ps-105.html) as specified in the current Catalog (http://catalog.uncc.edu/). Among other things, this code forbids cheating, fabrication or falsification of information, multiple submission of academic work, plagiarism, abuse of academic materials, and complicity in academic dishonesty.

Diversity Statement

No student will be discriminated against in the class based upon age, race, nationality, religion, sexual orientation, gender identity/expression, veteran's status, country of origin, or group affiliation. Likewise, all participants in this class will be expected to respect other members who fall into these categories. Any student who does not behave in a respectful manor with their classmates will be withdrawn from the class.

Special Needs:

If you have a documented disability and require accommodation in this course, contact Disability Services, Fretwell 230, phone: 687 4355 voice/TDD) the first week of the semester. Information about available services may be found at http://legal.uncc.edu/policies/ps-51.html. Accommodations for learning will be arranged by that office and communicated to the Instructor. If you speak English as a second language, please inform the instructor.

Religious Accommodation:

It is the obligation of students to provide faculty with reasonable notice of the dates of religious observances on which they will be absent by submitting a Request for Religious Accommodation Form to their instructor prior to the census date for enrollment for a given semester http://legal.uncc.edu/policies/ps-134.html . The census date for each semester (typically the tenth day of instruction) can be found in UNC Charlotte's Academic Calendar (http://registrar.uncc.edu/calendars/calendar.htm) .

Inclement Weather:

University Policy Statement #13 states the University is open unless the Chancellor announces that the University is closed. The inclement weather hotline number to call is 704-786-2877. **In the event of inclement weather, check your email the morning of class.** The instructors will use their best judgment as to whether class should be held understanding that some of you commute from far away and the instructors will notify you by email if class is cancelled.