Course Syllabus [UPDATED]

Course Information

Sections: CSCI 349 01
Place: BRKI 066
Time: Section 01: MWF 2:00-2:52pm

Instructor Information

Name: Brian R. King
Office: Breakiron 169
Phone: 570-577-3427
Office Hours: E-mail for an appointment
E-mail: brian.king@bucknell.edu
URL: http://www.eg.bucknell.edu/~brk009

Course Description

This course offers an introduction to data mining concepts and techniques. The focus will be on the understanding, implementation, and application of the most important data mining algorithms used today. Topics include: data cleaning and preprocessing (munging), data mining algorithms and methods including association analysis, classification, and cluster analysis. Basics of data visualization will be discussed. Depending on the time available, we will explore emerging methods and challenges in data mining and "big data".

Course Objectives

The following list identifies the primary objectives of the course. This course aims to teach students how to:

1. Understand the breadth of opportunities in this burgeoning field; understand the logic behind the increased interest in data mining in numerous domains
2. Learn essential terminology to be able to speak intelligently about data mining and data science
3. Learn foundational material in statistics and probability that is critical to understanding methods discussed
4. Understand how real-world entities are represented in data
5. Explore the most common tasks faced by data miners for understanding patterns in data, including classification, discrimination, association, and cluster analysis
6. Explain common challenges that are faced when dealing with "big data", including filtering and cleaning of large data sets, data analysis, storage and retrieval, and explore how these challenges are met
7. Learn how to use real-world data mining software to explore large datasets
8. Gain experience in data mining by applying methods learned using a current, popular data mining framework, working through numerous labs, assignments and a final project

Topics

We will explore fundamental methods used in data mining. Topics will include:

- Background and history of data mining
- Definitions, and use of the methods today
- Essential statistics and probability
- **Data munging / wrangling - preprocessing and cleaning**
- **Mining frequent patterns / association analysis**
- **Classification models**
- **Clustering**
- (Time permitting) Advanced classification models, outlier detection, stream and sequence mining, biological data, other advanced topics

Course Materials

Textbook(s): **REQUIRED:**
None

Moodle: Moodle will be the primary point of communication and transfer of material between the student and professor, found at http://moodle.bucknell.edu. You should plan on checking Moodle on a daily basis.

Computer: You will need a laptop to complete the work in a timely fashion. Though the Linux labs will have much of the software available, history with this course has repeatedly demonstrated that students do the best when they simply install everything on their laptop. A laptop is ***strongly*** recommended.

Software: This course will primarily use Python 3.7, and will have a ***heavy*** reliance on a wide range of Python packages and modules used in data mining, and more broadly speaking, data science.

The first lab will discuss the software tools used, and require you to install them on your own laptop. All tools used are freely available for students. You are responsible for ensuring all software is configured correctly on your system.

Fortitude: You *must* have fortitude to do well in this course. *Synonyms*: courage, bravery, strength of mind, toughness of spirit, firmness of purpose, persistence, resilience, backbone, etc. FACT: your learning is entirely up to you. No matter what your knowledge is coming into this course, there is plenty more that you can learn. You will get as much or as little out of this course as you want, depending on your fortitude. You *will* have challenges, and fortitude will get you through them. You might come across things you know and discover boredom. Fortitude will make you dig deeper out of your own desire to maximize your skills, *and* your future.

Slack: This course will use the Slack app for all communication of material related to the course. If you have any questions about the course, material, labs, assignments, etc. you must use Slack to communicate your questions. I will not respond to e-mail. You are strongly encouraged to utilize Slack to help each other through difficult problems, as long as you are not posting answers, files or any other content that represents an obvious answer to a problem. You must sign up for Slack ASAP using the URL provided in Moodle.

Real-time Polling: We will use a polling / quizzing app in class. You will need to be in class to respond to the polls. Details on signing up for the service will be given in lecture.

Course Schedule

A real-time course schedule will be maintained as a Google Sheet, linked in Moodle. Get yourself accustomed to referring to this schedule. This page will be your go-to place for keeping yourself up to date with lab and assignment due dates, course material covered, midterm and final exam dates, and so on. It is *your* responsibility to make sure you check the course schedule on a regular basis.

Grades

The grades **you** *earn* in this course are based upon **your** individual performance (with the exception of team projects, where your project grade depends on the collaborative work of the team). The following table lists the various criteria that will be used to assess your final grade in this course:

Category	Weight
Labs/Assignments	25%
Exams (2)	40%
Project	15%
Quizzes	10%
Journal Reflections	5%
Professionalism	5%

Updated in light of the campus shutdown for Spring 2020:

Category	Weight
Labs/Assignments	40%
Midterm	30%
Project	20%
Journal Reflections	5%
Professionalism	5%

Your final grade will be assigned as follows:

A >= 93%	[Exemplary. You impressed me. Superior achievement.]
A- >= 90%	[Outstanding]
B+ >= 87%, B >= 83%, B- >= 80%	[High Pass, Above Average]
C+ >= 77%, C >= 73%,	[Pass, Average work, Satisfactory]
C- >= 70%	[Below Average]
D >= 60%,	[Low Pass, Unsatisfactory, Barely passing]

F < 60% [Failure, Unacceptable]

Readings
Reading assignments will be given on a regular basis, taken from a variety of resources online, including online text books, individual chapters made available to the class, web sites, and relevant paper and periodical articles of interest. It is expected that you will complete reading assignments by the date specified. If the due date is on a class date, then you should complete the reading *before* class starts. Do NOT get behind.

Be prepared to discuss the reading material in class, or handle the occasional unannounced quiz related to the assigned reading.

Exams
There will be two exams – one in the middle of the semester, and one at the end. Exams will be given in class, and the final exam time slot will be used for final project presentations.

Quizzes
Random, unannounced quizzes *will* be given during class. The likelihood of a random quiz will increase as attendance decreases. They will be short, and designed to help you gauge your own understanding of the material; they may also be used to encourage class attendance if necessary, and to be sure you are keeping up with the material. Quizzes may be on Moodle, or using an online polling / quiz system.

Labs and Assignments
Labs and assignments are designed to give you a chance to explore a wide range of topics that are critical to your comprehension of the issues found in data mining. Labs are designed to be more "hands-on" with the exercises, and assignments are designed to be more thought provoking, challenging, and to give you a chance to apply what you have learned. Some assignments may require writing a critical answer to a problem; others may require you to write code to compute answers. As usual, pay attention to Moodle, as this will be where new labs and assignments are posted.

All labs and assignments will be submitted by pushing a Python Notebook .ipynb file to your Git repository.

Project
There will be one major project designed to give you and a team member an opportunity to apply the methods you are learning about toward the exploration and analysis of a real-world dataset. The project will culminate in a short paper and a presentation of your results at the end of the semester.

Course Policies

You are responsible to keep track of your daily commitments inside and outside of the classroom. If you do not keep up with the material, you will quickly become overwhelmed. Do not fall behind! Please do not wait until the last minute to start your work. As you know, computers around the campus are shared resources. Network problems can happen. You have obligations from other classes, clubs, sports, and other various organizations. These and related excuses will not be accepted as reasons for late work.

Late Work
Lab and Assignments: Your grade on a lab will be deducted 10% of the original value for each day that your work is turned in late, up to a maximum of 3 days. No late work is accepted after 3 days.

Projects: This is a substantial part of the course, and will be turned in at the end of the semester. Therefore, no late work will be allowed after the project deadlines.

Missed Quiz
There are no make-ups allowed on missed quizzes. If you miss a quiz, you receive a 0 for that quiz. If more than one quiz is given during the semester, the lowest quiz grade is dropped from your final grade calculation.

Missed Exam
No makeup exam will be given without proper documentation justifying your extenuating circumstances. If you must miss a class due to a documented illness, accident, death in family, other serious mitigating circumstances, or official university representation, then the student must:

- Inform me that formal documentation is coming from a physician, coach or other authority prior to the absence if possible, but no later than the first class period after the excused absence
- Bring the original documentation justifying your extenuating circumstance to me as soon as possible. I may ask to keep a copy for my own records.
- Work with me to schedule make up assignments and/or exams as soon as possible

Failure to provide adequate documentation that justifies your extenuating circumstances will result in a zero for the exam. The documentation must be supplied in the first class after you return, or at a time agreed upon by you and the professor. The makeup must be scheduled to occur within a week after you return.

Class Attendance
In addition to standard lectures, classes will consist of group exercises, open discussions on challenging topics, demonstrations, videos, and other activities (if they make sense for the topic at hand.) Many lectures will be conducted as labs, where you will work on exercises for the majority of the lecture. In fact, some classes will be "flipped" where a video of the lecture will be recorded for you to watch on your own, to allow more time in class to work on exercises. Therefore, regular and punctual attendance is **expected**. If you miss class, it is your responsibility to determine what material was missed. It is very important that you attend class since the lecture content will be drawn from a variety of sources.

Even if the class starts at an ungodly hour of 8am, you are still expected to be in class on time. Drink coffee.

Class Participation
The class will be taught using a combination of lecture, labs, hands-on work, videos, Q/A sessions, quizzes, polls, discussions, and small group exercises. You are expected to take part in all classroom activities. You are highly encouraged to ask questions, especially if you do not understand something. If you have a question, there is a good chance someone else has the same question as well, and other students may have answers! Please feel free to contribute to the questions and comments of other

students. Active and cooperative participation will help with your understanding of the material. Do not be shy!

Collaboration
All exams, quizzes and homework assignments and projects are **individual exercises unless explicitly stated otherwise.** Generally, I am a believer in the value of working together on material that is challenging. Great learning can be had in these circumstances. However, I also know that, unfortunately, communication sometimes turns into division of work in the name of trying to be efficient. Sometimes a strong member often gives the answers to others. At this stage of your academic career, I should not need to tell you how much this harms your learning, and is cheating! If you can't work together in a way that is helping partners equally learn the material while each partner is completing actual work individually, then do not communicate with anyone or anything besides the professor about the assignment. You are not to email anyone the assignment, post it on public space, solicit help from Stack overflow or any other online forums, or discuss via any means (in person, text, im, through your pet, etc.) about the assignment with anyone who is not the professor. If you have any questions whatsoever about anything related to any work in this class, talk to the professor!

Academic Engagement and Credit Hours

The following statement is reprinted as a reminder of the time you are expected to put into this course (as well as all of your courses at Bucknell):

> ***Courses at Bucknell that receive one unit of academic credit have a minimum expectation of 12 hours per week of student academic engagement. Student academic engagement includes both the hours of direct faculty instruction (or its equivalent) and the hours spent on out of class student work. Half and quarter unit courses at Bucknell should have proportionate expectations for student engagement.***

The above statement is certainly true of this course. You will have weeks that will meet or exceed the minimum criteria, *especially* if you fall behind. Plan your time accordingly. If you do, you will have very few problems completing course material, and you will maximize your learning of the material and enjoy the course.

Academic Responsibility

Read the following sentences carefully. A passing grade in the course can only be achieved by meeting all of these following criteria:

- All individual work in this course is to be your own work, and only your own work
- All team work in this course is to be the work of the team, and only the team
- Use of and/or copying any resources other than those provided by the instructor is deemed unacceptable
- Copying any work that is not your own, for any aspect of this course, is plagiarism. I will report it to the Board of Review for Academic Dishonesty

A plagiarism checker may be used throughout the semester on submitted work. Be assured that it accomplishes far more than checking for renaming variables, changing comments, and moving code around. It has been used in recent years to successfully catch a large number of cheaters in our courses. All have been reported, and all were found guilty, with the report being carried with them throughout their Bucknell career.

Consider yourself thoroughly warned. DON'T CHEAT!

If you have any question in regard to whether the activity you are doing at any time might be a violation of these rules above, then just assume it is, and don't do it! If you are struggling in the course, see the professor or a TA. Do not consult other resources via the Internet, via e-mail or other means of communication online. Do not consult anyone but the instructor or TA.

Bucknell University Honor Code

See https://www.bucknell.edu/academics/academic-responsibility-support/academic-responsibility for the honor code. As a student and citizen of the Bucknell University community:

1. I will not lie, cheat, or steal in my academic endeavors.
2. I will forthrightly oppose each and every instance of academic dishonesty.
3. I will let my conscience guide my decision to communicate directly with any person or persons I believe to have been dishonest in academic work.
4. I will let my conscience guide my decision on reporting breaches of academic integrity to the appropriate faculty or deans.

Failure to adhere to the Bucknell University Honor Code will result in immediate failure of the course.

Professionalism

Professionalism is defined by many criteria, most of which are common sense. In particular, mutual respect between you, your classmates, and the instructor are expected. This should not need to be discussed, and yet I continue to witness repeated violations of proper classroom behavior nearly every semester. The following is a list of rules expected out of every student in this course. Most are common sense. Please do not violate proper classroom decorum. Failure to do so will result in a 0 for your professionalism grade.

- When you bring your computer or tablet to class, you are expected to use it ONLY for your course work. In other words, do not surf the net, use any IM or chat programs, play games, read blogs, update or check your Facebook, Twitter, LinkedIn, Pinterest, Tumblr, Instagram, Vine, Snapchat, Reddit, or any other social media or other related accounts, or do any other activity with the computer that is not **directly** related to this course. You may think you are being subtle, but chances are that other people (including the professor) are noticing and are being distracted. It is downright rude and disrespectful. Your most glorious, ever emerging social status among all of your connections in the interwebs can wait to be checked and/or updated. If you are caught, you will be asked to leave.
- With respect to your mobile device / smart phone: in addition to the above expectation, phones must be completely silenced during class. This includes muting any vibrate setting. Do not use it to send / receive any text messages. (You can wait until class is over. The message will still be there.) If you can't resist the temptation, give the phone to the

instructor before class starts. Don't worry – he or she will return it. If you must take an emergency call during class time, quietly leave the room before answering the phone.

- If you are recording any portion of the class, you must get my explicit permission before doing so.
- Do not use the classroom to catch up on your sleep. Do not bring a pillow to class. Sleeping during class is extremely rude and disrespectful to the instructor, and disruptive to others around you… especially when you snore. Sleeping results in a 0 in your professionalism grade. Snoring results in a negative professionalism grade. The louder the snore, the more negative the grade.

There is no lower limit regarding how low your professionalism grade can go.

Your professionalism is determined by your response to all of the above criteria outlined above and throughout this entire syllabus. Professionalism is worth 5% of your grade. Stellar professional behavior may help decide borderline grades in cases where superior professionalism has been observed. A borderline grade is defined as a grade that is within 0.5 points of a letter grade change. For example, a student with an 86.6 in the course (letter grade of ”B”) who has a record of perfect attendance, active participation in discussions, respectful behavior among their peers and the instructor, is found helpful during group exercises, completes their homework in a timely fashion, does not sleep in class, does not text in class, etc… may, at the discretion of the professor, receive a ”B+.”

As an adult, you are responsible for your education at Bucknell. Therefore, you need to be the one to weigh the consequences of all of your decisions that may affect your performance in this course. Make your choices accordingly.

Students with disabilities

Any student who needs an accommodation based on the impact of a disability should contact Heather Fowler, Director of the Office of Accessibility Resources at hf007@bucknell.edu, 570-577-1188 or in room 212 Carnegie Building, who will coordinate reasonable accommodations for students with documented disabilities.