

Department of Business Analytics and Statistics
Haslam College of Business - The University of Tennessee
BZAN 542 – Data Mining Methods for Business Applications
MW 12:40 pm – 1:55 pm, Spring 2017 (HBB403)

Instructor:	Michel Ballings (PhD), Assistant Professor
Office:	249 Stokely Management Center
Phone:	(865) 974-1631
Email:	mb@utk.edu
Class sessions:	MW 12:40pm-1:55pm, January 1 - April 26, Final Exam: M 12:30pm-2:30pm, May 8 See course outline
Asking questions:	<ul style="list-style-type: none">• Office hours: MW by appointment or walk-ins 2pm-3pm• Online Q&A system (only about material): http://kddata.co/qa Please do not post the questions or answers of the Prep Quiz before the deadline (11:40AM on MW)
Technology:	Blackboard: http://bblearn.utk.edu Companion Website: http://kddata.co Videos: http://kddata.co/videos Q&A system: http://kddata.co/qa
Course materials:	Go to Blackboard > Course Materials
Prep quizzes, in-class assignments, at-home exercises	Go to Blackboard > Assignments
GTA:	Amil Williamson (awilli69@vols.utk.edu). Please contact Amil for grading questions only and add BZAN542 in the subject line.

Course Description

Each data mining project you will ever engage in involves 6 phases: business understanding, data understanding, data preparation, modeling, model evaluation, and model deployment. This course teaches you all the skills that you need to successfully complete these phases. We will work with technology (called 'R') that contains the latest and greatest

algorithms. To be able to use R you will need to learn how to program/code.

There are other tools out there that will allow you to point and click. SAS, SPSS and Jump come to mind. However, it takes time for these software providers to implement those point-and-click screens, and it takes time to implement new algorithms in a new programming language (e.g., at least 3 years in SAS). In addition, these platforms are closed, and you are not able to investigate the code to see how it works, to verify if it is even correct, or even improve it. Because being able to access the latest developments in the field will set you apart when you are looking for a job, I will spend considerable time teaching you how to code before we move on to the algorithms that will allow you to estimate models.

Course objectives

Students will learn

1. to program/code in R,
2. to process and manipulate data,
3. to understand and apply high performance data mining and machine learning algorithms,
4. to tune and evaluate models,
5. how to collaborate on data mining problems.

What's new in this iteration of the course?

	Fall 2014	Spring 2015	Spring 2016	Fall 2016	Spring 2017
Courseware					
No courseware expenses for students	✓	✓	✓	✓	✓
Write all notes on whiteboard, students copy	✓	✓			
Own Textbook (PDF)			✓	✓	✓
Own Videos				✓	✓
Course companion website				✓	✓
Q&A system				✓	✓
Class format					
Traditional classroom: In-class lecture	✓	✓	✓		
Flipped classroom: videos at home, exercises in class				✓	

Hybrid traditional-flipped classroom: videos and exercises at home, lecture-on- demand and exercises in class					✓
Grading					
Auto- drop 3 lowest scores			✓	✓	✓
Low frequency – high stake assignments and quizzes	✓	✓	✓		
High frequency – low stake assignments and prep-quizzes				✓	✓
Curving if necessary			✓	✓	✓
Bonus exercises			✓		✓
Getting used to the class format: Drop score of first prep-quiz, first optional at-home exercises, and first in-class assignment if you're not happy with it					✓
Learning styles					
Passive: Auditory and visual learning component	✓	✓	✓	✓	✓
Active: learning by doing component			✓	✓	✓
Group- based learning component			✓	✓	✓
Solitary learning component	✓	✓	✓	✓	✓
Self-paced learning component				✓	✓
Class-paced learning component	✓	✓	✓	✓	✓

Hybrid traditional-flipped classroom

The flipped classroom is a relatively new teaching method, where students watch a pre-recorded lecture online before coming to class to free up in-class time for practice problems. The idea is that this should promote student success more than the traditional classroom (in-class lecture, and at-home exercises) because students then have immediate access to the instructor when they are doing the most difficult part of learning something new: application of the newly acquired knowledge. I have implemented the flipped classroom for one semester and I found that students' grades were higher than ever. However, students provided the feedback that they really wanted their instructor to be there when they were acquiring the new knowledge, as opposed to only when they were applying new knowledge. This finding is consistent with literature (Hotle, S., Garrow L.A., *Flipped and micro-flipped classrooms*, February

2015, *ORMS Today*, 42, 1, p12-13). Hotle and Garrow recommend a hybrid traditional-flipped classroom in which all aspects of learning (knowledge acquisition, and knowledge application) happen both at home and during class. As such the advantages of both teaching formats are combined. Figure 1 is a visual representation of the hybrid traditional-flipped classroom that we will use in this class.

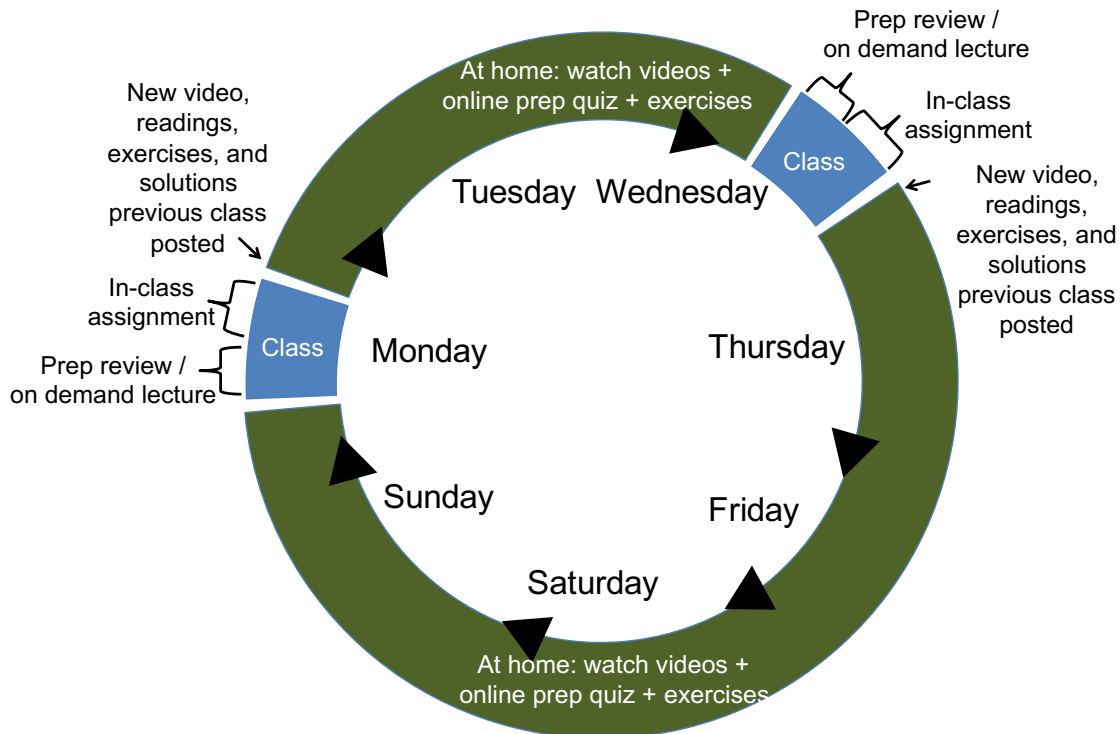


Figure 1: The hybrid classroom format for BZAN542

Step 1 (at home): videos and readings

You acquire new content at home through videos and readings. Some videos are short and some are long. Sometimes multiple videos are assigned. The idea of the videos is that you can skip material that is too easy, and pause material that is too hard.

Step 2 (at home): optional exercises

After watching the video(s) and completing the readings there is a set of exercises that will help you understand the material at a deeper level. These exercises are relatively easy, and optional, but highly recommended. They are a great way to prepare for the in-class assignment and help you master the material.

You can make the exercises without turning them in. If you do turn them in they will be graded according to a bonus system. The TA will assign a binary grade of either 100% or 'exempt'. Exempt means that the grade will not be taken into account towards your overall grade. This is the characteristic function that will be used to determine your final grade:

$$\text{Binary grade} := \begin{cases} 100\%, & \text{if } \text{real} - \text{valued grade} \geq 70\% \\ \text{exempt}, & \text{if } \text{real} - \text{valued grade} < 70\% \end{cases}$$

There are two reasons why I am using this function. First, I want to reward you for making the exercises. If we would not have that 70% threshold, then you would be less inclined to try the exercises because there is always the chance of having a low score, and that would defeat the purpose of a bonus system. Second, assigning a binary grade as opposed to a real-valued grade, and without any downside risk for the student, makes grading a lot more efficient for the TA. As you can imagine, there will be a lot of grading in this course and this is one way to make it a win-win situation for both the students and the TA. As a final note, this system is designed to reward a reasonable effort to complete the exercises. Please be considerate of the TA's workload and try to only send in your answers if you feel you have a reasonable chance of getting it right in 70% of the exercises. If you choose to send your exercises in for grading, the window for sending them in closes one hour before the corresponding class meeting.

The submission URL for the exercises will be specified in the exercise document.

Step 3 (at home): prep- quiz

After watching the video(s) and completing the reading (and optionally making the at-home exercises) there is a prep- quiz. Prep Quizzes are short duration (10 minutes) timed online quizzes (Blackboard) that should be attempted only after completing the at-home assigned readings, and/or video lecture viewings. The prep-quizzes are auto-graded. The window of completion closes one hour before the corresponding class meeting. The Prep Quizzes are considered individual assignments and no help or advanced notice should be given or received prior to or when taking them. Due to their wide window for completion, there are no opportunities to make up missed Prep Quizzes, but the grading calculation will automatically drop your three scores.

Step 4 (in class): on-demand lecture or review of exercises

Class will be used for two purposes: (1) reviewing the prep quiz, material, previous in-class assignment, or anything else the class deems necessary (e.g., an on-demand lecture), and (2) an in-class group assignment. I will start class by asking you if you have any questions, or if you want me to review something. The goal of the at-home part is to have you generate a list of questions that you want to ask during class. If you prefer, you can also post them on the Q&A website (<http://kddata.co/qa>). I will assume a certain degree of preparation from your part, meaning that I am not expecting questions that would repeat everything that was said in the videos or readings. I am expecting questions that pertain to the material, and have me go over specific parts of the material that were unclear. I hope you can understand that it would be unfair to students that have made the effort to prepare for class, for me to basically rehash the video during class, and have them lose time in the process. If I would do that, students will stop preparing for class altogether. We have a maximum of 30 min for step 4 (i.e., until 1:10pm).

Step 5 (in class): group-based assignment

This step will take start at 1:10pm the latest. If step 4 takes less time, you will have more time for this step. Before class you will receive an email specifying your group number and assigned seat location. I will walk around during class to help you with the assignment. In-class assignments are to be completed during class meetings.

Collaboration is a crucial component in the learning experience. Therefore, In-Class Gradeables need to be completed in class unless you have permission from the instructor to take it from home. This rule exists because

- it is unfair towards your colleagues not to show up: their group will be smaller and each individual will have to carry a bigger load
- collaboration is a crucial component in the learning experience
- not showing up hinders course organization and the TAs work in preparing the group configurations

Missing in-class assignment results in a grade of 0. You can turn that 0 into an exempt grade if you have a pre-approved written excuse. A maximum of 2 pre-approved excuses does not require any documentation from your part such as a doctor's note. After those 2 excuses documentation will be required as specified in the College's and University's rules and regulations. Pre-approval requires you to send an email 2 hours before the start of the class.

The grading calculation also automatically drops the lowest 3 scores, of the scores that remain after any preapproved drops. This is to accommodate students that miss class without a preapproved excuse. Everybody will get those three drops automatically at the end of the course.

The submission URL for the assignment will be specified in the assignment document.

Grades in Blackboard

If you have no score (--) for a prep-quiz or in-class assignment this means you did not take the quiz or were absent. This is equivalent to a 0, and will be counted as such in the average/total.

If you were excused for not doing the in-class assignment you will get an Exempt Grade. This will look like this in the grade center:



Taking Prep Quizzes in Blackboard

Internet Explorer and Safari can cause trouble such as not showing all questions. Therefore, only use Firefox or Chrome to take the Prep Quizzes

Dropbox

I often get requests for extensions on deadlines because somebody's computer crashed with all their work trapped inside. To avoid these problems please sign up (free) for www.dropbox.com and download the app. The app will create a folder on your computer, and everything you do in that folder will sync to the cloud, so you can access it from another computer. Put the files you are working on in that folder.

Q&A system

Whenever you have a question about the material, post it on <http://kddata.co/qa> so your colleagues can learn from it.

Make sure to ask and answer questions in a polite and respectful way to keep the site functional for everybody. The site will not automatically post your name or username whenever you post a question or answer, but I can see who posted the question or answer in the system. I appreciate it if you post your name whenever you post something, but you don't have to.

First register here. Use your official school email address, but DO NOT use your school password, as this will be visible for me. Use a password that is easy to remember and write it down, as you will be using it a lot. You only need to register ones for both the videos and Q&A part of kddata.co.

<http://kddata.co/signup474.php>

List of questions, and asking new questions:

<http://www.kddata.co/qa/>

Videos, readings, solutions

To be able to watch the videos first sign up here. Use your official school email address, but DO NOT use your school password, as this will be visible for me. Use a password that is easy to remember and write it down, as you will be using it a lot. You only need to register ones for both the videos and Q&A part of kddata.co.

<http://kddata.co/signup474.php>

- Videos: <http://kddata.co/videos>
- Readings: <https://goo.gl/boiuCx>
- Solutions to in-class assignment: <https://goo.gl/VgPGt0>
- Solutions to at-home exercises: <https://goo.gl/LdjoDA>
- Solutions to prep- quizzes: <https://goo.gl/gWesQa>
- All in one folder: <https://goo.gl/PxGmj1>

Exams

There are two midterm exams and a final. Any conflict with the exam dates needs to be discussed with me before the scheduled date to avoid being counted as absent.

Missing an exam results in a 0. There will be an opportunity to take the exam at a later time if you have a well- documented pre-approved excuse (e.g., doctor's note, tow truck receipt, police note, note of athletic department) with a date on it. If not pre-approved by email by the instructor, all excuses, even well documented ones, are subject to being denied. Note that even pre-approved excuses can be withdrawn if there is no documented excuse after the absence. Please do everything you can to avoid missing exams.

Grade components and grading

Item	Percentage	Comments
Preparation Quizzes and optional bonus exercises	15%	Each Prep Quiz gets equal weight. We will drop your lowest three grades automatically.
In-Class Gradeables	30%	Each In-Class Gradeable gets equal weight. We will drop your three lowest grades automatically.
Midterm1	15%	
Midterm 2	15%	
Final Exam	25%	

Final letter grades will be assigned as follows:

Grade	Suggested range (%)	Grade Points
A	92-100	4.0
A-	90-91.99	3.7
B+	88-89.99	3.3
B	82-87.99	3.0
B-	80-81.99	2.7
C+	78-79.99	2.3
C	72-77.99	2.0
C-	70-71.99	1.7
D+	68-69.99	1.3
D	62-67.99	1.0
D-	60-61.99	0.7
F	0-59.99	0.0

I do not round up. Once a grade has been communicated the student has 72 hours to make an appointment and review the quiz/exam with the GTA. After 72 hours grades cannot be changed anymore.

Academic & Professional Integrity:

Professionalism implies respect and courtesy for others in our classroom setting and chosen business profession. We expect our students to maintain the highest standards of professionalism in the classroom, in group and team settings, in the greater university community, and in related public settings. All that you do and say, and the way you present yourself either elevates or diminishes your professional image in the eyes of others.

Furthermore, each student is responsible for abiding by the policies and honor code set forth in UT Hill Topics regarding academic integrity. Cheating of any sort including plagiarism will not be tolerated and will result in a grade of F for the course (at the instructor's discretion) and a charge of academic dishonesty against the student(s). It is recognized that any student has the right to appeal a grading decision of an instructor and/or penalties resulting from a charge of academic dishonesty.

UT Honor Code Statement:

"As a student of the University, I pledge that I will neither knowingly give nor receive any inappropriate assistance in academic work, thus affirming my own personal commitment to honor and integrity."

Use of Laptops in the Classroom:

We will be programming almost each class. So please bring your laptop at all times. I do recommend refraining from utilizing laptops or other electronic devices such as tablets or cell phones for non-class purposes.

Disability Statement

Any student who feels s/he may need an accommodation based on the impact of a disability should contact the instructor privately to discuss specific needs. Please contact the Office of Disability Services at 865-974-6087 in Hoskins Library to coordinate accommodations for students with documented disabilities.

UT Inclement Weather Policy

The University of Tennessee, Knoxville, will remain open except in the most severe weather conditions. The chancellor (or appointed representative) may officially close or suspend selected activities of the university because of extreme weather conditions. When a decision to close is reached, campus and local radio and TV stations will be notified and the notice will be posted on the front page www.utk.edu.

In the event of a delayed opening, the chancellor (or appointed representative) will determine a specific time of opening and that information will be distributed to the campus community through the local media and via the front page of www.utk.edu. All faculty and staff are expected to report to their specific work location by the set opening time. Students are expected to report to their regularly scheduled class only if there are 30 or more minutes remaining in the session.

Preliminary Course outline – Topics for next class will be updated before current class starts

Session	Date (2017)	Pre-class readings (book) and pre-class video(s) to watch	Prep Quiz & optional at- home exercises & in- class assignment?
1	01/11	<ul style="list-style-type: none"> Discuss Syllabus and class format 	No
2	01/18	<ul style="list-style-type: none"> Topics: History of Data Mining, Business Understanding, CRISP-DM and Software, Objects Videos: 1-5 Book Sections: 1.1, 2.1, 2.2, 2.3.3.1, 2.3.3.2 (book version*: 2016/11/23/14:13:55) 	Yes-1
3	01/23	<ul style="list-style-type: none"> Topics: Subsetting Video: 6 Book Section: 2.3.3.3 (book version*:.) 	Yes-2
4	01/25	<ul style="list-style-type: none"> Topics: Memory management, reading data Videos: 7, 8 Book Section: 2.3.3.4, 2.3.3.5 (book version*:.) 	Yes-3
5	01/30	<ul style="list-style-type: none"> Topics: Data exploration Video: 9 Book Section: 2.3.3.6 (book version*:.) 	Yes-4
6	02/01	<ul style="list-style-type: none"> Topics: Installing and loading packages, Missing values, Getting help, Time window, Code flowchart Videos: 10- 14 Book Sections: 2.3.3.7, 2.3.3.8, 2.3.3.9, 2.4.1, 2.4.2 (book version*:.) 	Yes-5
7	02/06	<ul style="list-style-type: none"> Topics: operators, dummy variables, conditional processing Video: 15 - 17 Book Section: 2.4.3.1, 2.4.3.2, 2.4.3.3 	Yes-6

		(book version*.)	
8	02/08	<ul style="list-style-type: none"> Topics: timing code, loops Video: 18, 19 Book Section: 2.4.3.4, 2.4.3.5 (book version*.) 	Yes-7
9	02/13	<ul style="list-style-type: none"> Topics: applying functions to lists, aggregating, merging Video: 20, 21, 22 Book Sections: 2.4.3.6, 2.4.3.7, 2.4.3.8 (book version*.) 	Yes-8
10	02/15	<ul style="list-style-type: none"> Topics: Dates, Creating functions, Basic Version Control Video: 23, 24, 25 Book Sections: 2.4.3.9, 2.4.3.10 (book version*.) 	Yes-9
11	02/20	<ul style="list-style-type: none"> Topics: Text mining Video: 26, 27, 28 Book Section: 2.4.3.11 (book version*.) 	Yes-10
12	02/22	<ul style="list-style-type: none"> Topics: data.table package Video: 29 Book Section: 2.4.3.12 (book version*.) 	Yes-11
13	02/27	<ul style="list-style-type: none"> Midterm 1: everything (readings, videos, and solutions to in-class gradeables and prep quizzes) from sessions 2 to 12. You can use the book, internet, videos, notes, solutions to ICG and quizzes. 	EXAM
14	03/01	<ul style="list-style-type: none"> Topics: training, validation, test set and Naïve Bayes Videos: Modeling, 30, 31 Book Sections: 2.5 (first two paragraphs), 2.5.1, 2.5.2 (book version*.) 	Yes-12
15	03/06	<ul style="list-style-type: none"> Topics: Logistic regression Video: 32 Book Section: 2.5.3 (book version*.) 	Yes-13
16	03/08	<ul style="list-style-type: none"> Topics: Neural networks Video: 33 Book Section: 2.5.4 (book version*.) 	Yes-14
17	03/20	<ul style="list-style-type: none"> Topics: K-nearest neighbors Video: 34 	Yes-15

		<ul style="list-style-type: none"> • Book Section: 2.5.5 (book version*.) 	
18	03/22	<ul style="list-style-type: none"> • Topics: Decision trees • Video: 35 • Book Section: 2.5.6 (book version*.) 	Yes-16
19	03/27	<ul style="list-style-type: none"> • Topics: Ensemble methods: Bagged trees • Videos: 36, 37 • Book Section: 2.5.2, 2.5.9, 2.5.10 (book version*.) 	Yes-17
20	03/29	<ul style="list-style-type: none"> • Topics: Random forest • Video: 38 • Book Section: 2.5.11 (book version*.) 	Yes-18
21	04/03	<ul style="list-style-type: none"> • Midterm 2: everything (readings, videos, and solutions to in-class assignments and prep quizzes) from sessions 14-20. You can use the book, internet, videos, notes, solutions to ICG and quizzes. 	EXAM
22	04/05	<ul style="list-style-type: none"> • Topics: Adaptive boosting • Video: • Book Section: (book version) 	Yes-19
23	04/10	<ul style="list-style-type: none"> • Topics: Support Vector Machines • Video: • Book Section: (book version) 	Yes-20
24	04/12	<ul style="list-style-type: none"> • Topics: Rotation Forest • Video: • Book Section: (book version) 	Yes-21
25	04/17	<ul style="list-style-type: none"> • Topics: Performance evaluation of binary classification models • Video: • Book Section: (book version) 	Yes-22
26	04/19	<ul style="list-style-type: none"> • Topics: Understanding classifier performance (bias-variance tradeoff) • Video: • Book Section: (book version) 	Yes-23
27	04/24	<ul style="list-style-type: none"> • Topics: Understanding classifier performance (bias-variance tradeoff) • Video: • Book Section: (book version) 	Yes-24
28	04/26	<ul style="list-style-type: none"> • Topics: Model deployment 	Yes-25

		<ul style="list-style-type: none"> • Video: • Book Section: (book version) 	
30	05/08	12:30pm-2:30pm, Final exam: everything (readings, videos, and solutions to in-class assignments and prep quizzes) from sessions 22 to 28. You can use the book, internet, videos, notes, solutions to ICG and quizzes.	EXAM

- **Book version can be found on p2 of the book. Make sure to read the specified version or later.*
- *The course outline for the next class will be up to date before the start of the class session*

The descriptions and timelines contained in this syllabus are subject to change at the discretion of the professor. Updates (except updates to the course outline) are always highlighted in yellow and stay that way for two weeks.