PREDICTIVE MODELING

August 19th – December 5th [Fall 2024]

Syllabus

DR. LILY POPOVA ZHUHADAR

Email: Lily.popova.zhuhadar@wku.edu

Office Locations: Center for Applied Data Analytics, Grise Hall, First Floor, GH 128.

Guidelines for Requesting Zoom Sessions in BDAN 420

Dear Students,

As you are aware, our BDAN 420 course is conducted entirely online, and there are no in-person meetings. To ensure you have the support you need throughout the semester, I am available for Zoom sessions to discuss any course-related issues you may encounter.

Please follow the steps below to request a Zoom meeting:

1. Send your request via email to Lily.popova.zhuhadar@wku.edu.
2. Use the following format for the subject line of your email: "BDAN 420 - Request for Zoom Meeting".
3. In the body of your email, clearly outline the reason for your request, providing as much detail as possible to assist in addressing your concerns promptly.

I strive to respond to all student requests within 24 hours. To ensure timely assistance, especially with homework-related inquiries, please send your request at least three days before the assignment's due date. Each homework assignment is designed with a minimum of one week to complete, giving you ample time to reach out for help if necessary.

Thank you for your attention to this process, and I look forward to assisting you in achieving a successful semester.

Course Description

This course teaches the fundamentals of data mining and data science. It is a practical ***Project-based*** course. It introduces a machine learning framework with business applications using ***the 7-Step Business Analytics Process Framework***. It provides a survey of supervisedmachine learning algorithms and techniques--including regression, classification, regularizationand reduction, tree-based methods, and fully-connected, convolutional, and recurrent neural networks.

Students also learn modern techniques of unsupervised learning where they identify groups of individuals and groups of variables with cluster analysis and block clustering. In addition, they explore relationships among categorical variables with association rules and detect anomalies using autoencoders and probabilistic learning.

Machine Learning Professional Certification

At the end of the semester, students will be able to pass the RapidMiner Machine LearningProfessional Certification. Examination Guide can be Accessed from this link.

Capstone Final Project Overview

Instead of the final exam, students work on a Capstone project. This project is the culmination of the business data analytics program. It provides students an opportunity to demonstrate their strategic business thinking, communication, and consulting skills. These business cases are across various industries, and application areas illustrate strategic advantages of analytic. Students are assigned to a specific theme to generate business plans and project execution plans.

Dual-major students can choose Capstone projects closely related to their first major withan analytics modeling focus. While the capstone theme falls under various domains such as (business analytics, finance analytics, marketing analytics, health analytics, sports analytics, etc.,) students, work individually on projects within their preferred analytics focus.

Learning Objectives

At the end of the semester, students should be able to

- Critically identifying which types of algorithms and methods are better in solving the final Capstone project.
- Productively presenting a solution to analytical problems in the Business Analytics Final Project.

What Are Student Deliverables?

- At the end of the Course, students will submit a detailed 10+ page project report; and deliver a final presentation as a video pitch presentation.

Examples of Capstone Analytics Modeling Theme

1) Health Analytics Theme

Students apply machine learning techniques to health-related data. Projects related to predicting a heart attack, assessing mobile health heart rate App, which factors best determine whether a patient will receive a flu shot immunization? Predict which of these factors might have a relationship with mortality, billing issues and fraud and abuse, develop a model for detecting hospital readmissions.

2) Marketing Analytics Theme

Students apply machine learning techniques to applications in digital marketing -- including applications of analytics in marketing. Students examine methods for understanding customerpreferences, market segments, and competitive brands and products. They address problemsin new product design and pricing -- highlighting the impacts of promotion and advertising.

3) Financial Analytics Theme

Students apply machine learning techniques to finance--including risk analytics, financial fraud detection, and taxation analytics.

4) **Sports Analytics Theme**

Students utilize predictive modeling and presentation graphics, showing real-world implications in the sports industry. To create a predictive model, they apply machine learning to review athletic performance measurements, including speed ability, agility, and strength. Also, they develop models related to sports team performance. They apply modeling methods in studying player and team valuation.

5) **Web and Network Data Analysis Theme**

Students analyze information from the web and search performance metrics, including random graphs and small world. They compute network metrics, exploring structure and associations in information and social networks.

6) **Natural Language Processing Theme**

Students use recent developments in computational linguistics and machine learning to work with unstructured and semi-structured text from online sources and document collections to develop applications including text classification, search, recommendation systems, sentiment analysis, and topic modeling.

Expectations and Goals

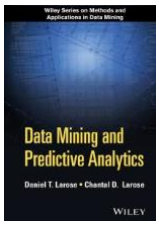
Throughout the semester, this course' conceptual coverage is complemented with application case studies (examples of both successes and failures), as well as simple, hands-on tutorials.

Graded Work

- Online Homework (70%): Homework will be posted on Blackboard. Therefore, you will always need to access Blackboard, regularly throughout the semester (approximately once or twice weekly). It is your responsibility to access the assignment in adequate time to complete it before the due date/time.
- Final Project (30%): A Capstone

Grades will be posted on Blackboard. Please consider the online gradebook as a courtesy to you, subject to errors given various upgrades and shifts in the software. I reserve the right to make gradebook corrections to keep it consistent with the syllabus so that your grade reflects true performance, not software or user error. If you see something that doesn't make sense, please alert me.

Textbook (optional)



DATA MINING AND PREDICTIVE ANALYTICS

DANIEL T. LAROSE & CHANTAL D. LAROSE

ISBN-13: ISBN-13: 978-1118116197 | **ISBN-10:** 1118116194| Copyright © 2019 | ***Published by WILEY***

Course Outline

PART I - DATA PREPARATION

Module 1: An Introduction to Data Mining and Predictive Analytics
Module 2: Data Preparation
Module 3: Preparing to Model the Data

PART II - A QUICK REVIEW OF STATISTICAL MODELING

Module 4: Statistical Modeling - Linear Regression

- A Review of Linear Regression, Least-Squares Estimates, Coefficient of Determination, Standard Error of the Estimate, Correlation Coefficient, Outliers, Regression Equation, Verifying the Regression Assumptions.

PART III - CLASSIFICATION

Module 5: Decision Trees

- What is a Decision Tree?
- Requirements for Using Decision Trees
- Classification and Regression Trees- C4.5 Algorithm
- Decision Rules
- Comparison of the C5.0 and CART Algorithms Applied to Real Data Confusion Matrix

Module 6: Logistic Regression

- Example of Logistic Regression
- Maximum Likelihood Estimation
- Interpreting Logistic Regression Output
- Odds Ratio and Relative Risk
- Interpreting Logistic Regression for a Dichotomous Predictor
- Interpreting Logistic Regression for a Polychotomous Predictor
- Interpreting Logistic Regression for a Continuous Predictor
- Assumption of Linearity
- Validating the Logistic Regression Model

Module 7: Model Evaluation Techniques

- Model Evaluation Techniques for the Estimation and Prediction Tasks
- Model Evaluation Measures for the Classification Task
- Accuracy and Overall Error Rate
- Sensitivity and Specificity
- False-Positive Rate and False-Negative Rate
- Proportions of True Positives, True Negatives, False Positives, and False Negatives
- Misclassification Cost Adjustment to Reflect Real-World Concerns
- Decision Cost/Benefit Analysis

Module 8: Graphical Evaluation

- Review of Lift Charts and Gains Charts
- Lift Charts and Gains Charts Using Misclassification Costs
- Response Charts
- Profits Charts
- Return on Investment (ROI) Charts

PART IV Auto Machine Learning (AutoML)

Module 9: Automated Machine Learning (AutoML)

- Deep Learning
- Predicting customer responses to marketing
- Customer acquisition and reducing churn
- Predicting customer satisfaction
- Product and service recommendations

Course Prerequisite

Before attending this course, you should have a Personal Computer to install some software packages, in order to do your assignments, during the semester.

Required Software

- A significant amount of time that students spend completing their assignments will involve the use of data science mining software (will be listed on Blackboard).
- Students will be able to learn the basic features of one or more of these through training videos that are posted in Blackboard, self-directed studies or by using available resources on Blackboard.

Disclaimer

- This syllabus is subject to change at the discretion of the instructor. Any such changes will be clearly communicated by email and/or on Blackboard.
- The scope, timing, and due date/time of any assignments, or any other required work may be adjusted by the instructor as needed to maximize learning opportunities for students and/or better serve the goals of the course. The syllabus may likewise be modified at the discretion of the instructor. Any adjustments will be communicated to students in class and on Blackboard with as much advance notice as possible.

Expectations

Students are expected to:

- Read the scheduled assignments each week
- Complete the assigned homework, and quizzes posted on Blackboard, by the due date
- Students are expected to complete each test, exam, homework, and all other assignments independently. The student's submissions must represent his or her individual work, and citations must be provided where content from other sources is referenced.

Course Policies

- The professor reserves the right to make changes to the syllabus or schedule as necessary; it is the student responsibility to be aware of these changes by periodicallychecking the syllabus on Blackboard Course Site.
- Grades will be posted promptly on Blackboard after each assignment has beencompleted.
- Problems with Blackboard should immediately be reported to the Information Technology Services at (270) 745-7000.
- Assignments are due and must be uploaded to the appropriate location by the announced due date. Your assignments must be uploaded to the correct place toreceive a grade.
- A malfunction of your personal computer is not a reason for not completing a test orassignment on time.
- Don't email your assignment to me unless I instruct you to.
- It is the student's responsibility to check the site to make sure the assignment upload was successful. However, if you don't see it immediately you should wait at least a couple of hours to give the online system time to process the upload. If the upload wasnot successful, repeat the process. If you are still having problems, contact the appropriate tech support service for assistance.
- No "extra credit" activities will be provided so please do not ask.
- The course policies apply to all students. Students experience personal illnesses, a family crisis, work schedule problems, automobile trouble, and similar "life" situationsevery semester. These are not typically unusual circumstances and do not warrant exceptions to course policies. Good planning and proactive efforts on the part of the student can avoid many of the problems that arise from these situations.

Assignment Evaluation Rubric

Criteria	10 – Outstanding	9 – Proficient	8 – Basic	7 (or lower) - Below Expectations
OBJECTIVE				
Completed assignment per requirements	All portions of the assignment, including presentations, data preparation, and visualizations were attempted and submitted.	This is a pass / fail component. All or no points are awarded.		
The analysis is appropriate and sufficient	The data analysis used is appropriate, correct, and sufficient to support the findings.	The data analysis is appropriate but minor issues may be present or enhancements may be needed.	The data analysis is related but not sufficient to support the findings, or significant data analysis issues prevent a clear reading of the results.	The data analysis has little or no relation to the topic being explored, errors will lead to incorrect conclusions, and/or data issues make the findings unusable.
Headers, directions, citations, and visual cues are given as guides	A clear direction is provided. Citations are consistently and correctly employed to inform and guide.	Headers, footers, and instructions are present, but visual cues may be missing or could be improved.	The user must self-discover functionality. Headers and footers may be missing. Difficult to know what to do.	The user has little or no indication of how to engage. Directions are missing on clear. Missing headers and footers for context and meaning.
SUBJECTIVE				
Analysis & Viz are clean, clear, concise, captivating	The data analysis & data Viz used are well represented; the visualization is clear, clean, concise, and captivating.	Aspects of data analysis & data Viz are apparent; opportunity exists for further enhancement.	Multiple aspects of data analysis & data Viz are missing, or have not been well addressed in the visualization.	Significant or complete disregard for the guidance present in the lectures, resulting in poor data analysis & visualization.
Attractiveness and attention to design and details of craft	Proper grammar and spelling are used, and choice of position, size, and emphasis integrate elements into a visually appealing and engaging whole.	The visualization shows thought and planning, and most aspects work in harmony. May exhibit minor issues with spelling, alignment, or sizing mismatched with importance.	Visualization appears sloppy and may be difficult to understand as a coherent whole. Multiple issues with spelling, font consistency, positioning, or other distracting characteristics.	Little or no apparent thought or given and visualization comes across as disorganized. May be visible through numerous spelling or grammar issues.
The visualization is usable and actionable	The visualization is targeted to the audience, the story is evident, and the conclusion or action required is clearly apparent. No additional interpretation is needed.	There is a clear message or story conveyed, but the action or conclusion that should be drawn is not definitive. May require interpretation.	The visualization suggests some possibilities but does not lead to clarity of understanding and therefore action is not possible.	No apparent message or relevancy to the user; no actions can nor should be taken based on the analysis.
Quality, integrity, and impact of the findings and analysis	The analysis shows a level of quality, integrity, and competency that makes the data analysis & data viz impactful, generating a high level of trust.	The overall conclusions of the analysis seem to be sound, with support by anecdotes or additional evidence.	The analysis shows a trend or suggests a result, but is not trustworthy because of errors in the process, omission, or scope.	The analysis appears to be poorly conducted, greatly compromising the integrity of some or all of the data analysis & visualization.
The overall effectiveness of communication and presentation	The presentation is delivered in a convincing way that demonstrates confidence, competency, and thoroughness.	Delivery provides a strong argument and is well supported; minor details should be vetted and affirmed.	The presentation and communication leave concerns or lingering lack of clarity. Work required to review and confirm.	The communication and presentation result in confusion and a low level of confidence in the analysis, requiring a significant or complete re-do.

Email Etiquette

- Each student must have a WKU email address and Blackboard account for participation. Email and Blackboard are considered official communication tools for thiscourse. Thus, announcements and reminders are posted/sent frequently and will be sent only to official WKU email addresses. If you use a different email address, pleaseensure that your WKU email is forwarded correctly. It is your responsibility to check WKU email and Blackboard regularly and to be aware of all reminders and announcements.
- E-mail will generally be the best way to contact me. During normal business hours, Iusually respond to e-mail fairly quickly (unless I am in a meeting/class/etc.).
- Please consider e-mail as a method of professional correspondence. Also, when you contact me via e-mail, make sure that you use the following format for the subject line of your email: "**BDAN 420 - Request for Zoom Meeting**".

Academic Dishonesty

- Students who commit any act of academic dishonesty will receive from the instructor afailing grade F in this course. This rule applies to any act that involves cheating and plagiarism.
- Cheating will not be tolerated. Note that cheating might involve but not limited to the following *actions: "To represent written work taken from another source as one's own is plagiarism. Plagiarism is a serious offense. The academic work of a student must be his/her own. One must give any author credit for source material borrowed from him/her. To lift content directly from a source without giving credit is a flagrant act. To present a borrowed passage without reference to the source after having changed a fewwords is also plagiarism," for more details, refer to https://www.wku.edu/undergraduatecatalog.*

University Policies & Resources

ADA Compliance

In compliance with university policy, students with disabilities who require academic and/or auxiliary accommodations for this course must contact the Student Accessibility Resource Center located in Downing Student Union, Room 1074. The SARC can be reached by phone number at 270-745-5004 [270-745-3030 TTY] or via email at sarc.connect@wku.edu. Please DO NOT request accommodations directly from the professor or instructor without a faculty notification letter (FNL) from the Student Accessibility Resource Center. Students with an FNL should NOT email the FNL to the instructor requesting a signature. Instead, students must first meet with the instructor during scheduled office hours to discuss how the accommodations requested will be served in this course. The Blackboard Ally tool has been enabled within the Blackboard course site that allows you to have access to different formats of course files, including HTML, readable PDF, electronic braille, ePub, and audio (mp3). You may review the Ally for Students video regarding how to access these alternative formats.

Research Appointments with your Personal Librarian

At WKU Libraries, a personal librarian is available for every program on campus, plus Special Collection librarians and archivists. Our goal is to save you time and help you be successful on term papers and research projects by showing you what you need to know to get started and be successful. Start your research by scheduling an appointment with your Personal Librarian (you can find them listed online here) through email or by calling (270)745-6125.

Writing Center Assistance

The Writing Center on the Bowling Green campus will have writing tutors available to offer advice to current WKU students on any stage of their writing projects. In-person tutoring is available in Cherry Hall 123 from 10:00 AM to 4:00 PM Monday through Friday and in the Cravens Commons (at the horseshoe-shaped reference desk) from 5:00 PM to 9:00 PM on Sunday through Thursday evenings. Making appointments, as well as instructions and how-to videos are available through the website.

Walk-in feedback may be available, and students may also request feedback on their writing via email or arrange a real-time Zoom conference to discuss a paper. Students may also get short writing questions answered via email line sent to writingcenter@wku.edu by entering Quick Question in the subject line. The WKU START Centers will be offering writing tutoring sessions via Zoom as well as in person in their Glasgow and Elizabethtown locations. More information is available at WKU Start Center.

WKU Counseling and Testing Center

The university experience should be challenging, but not overwhelming. To this end, the WKU Counseling Center provides a variety of services to help strengthen students' capacity to tolerate distress, form healthy relationships, and seek healthy expressions of their ideals and values. If you would like to speak with someone, you may contact WKU's Counseling and Testing Center at 270-745-3159 or use their Here To Help service at https://www.wku.edu/heretohelp/.

Peer Tutoring Services

The Learning Center (TLC) provides free tutoring services helps students enhance their academic performance and sharpen their skills to be successful WKU graduates.

Mental Health Support Group

A Free Support/Recovery Group for current WKU students struggling with anxiety, depression, or other mental health issues is available. This is an opportunity for students to support, connect with, and encourage others struggling with mental health issues and is led by two National Alliance on Mental Illness (NAMI) certified facilitators who are in mental health recovery themselves. This group has the support of the WKU Counseling and Testing Center and is a not a substitute (rather a supplement) for therapy or medication. No formal mental health diagnosis is required for students to attend. The group is sponsored by the NAMI-Bowling Green Chapter.

Title IX/ Discrimination & Harassment

Western Kentucky University (WKU) is committed to supporting faculty, staff and students by upholding WKU's Sex and Gender-Based Discrimination, Harassment, and Retaliation (#0.070) and Discrimination and Harassment Policy (#0.2040). Under these policies, discrimination, harassment and/or sexual misconduct based on sex/gender are prohibited. If you experience an incident of sex/gender-based discrimination, harassment and/or sexual misconduct, you are encouraged to report it to the Executive Director, Office of Institutional Equity/Title IX Coordinator, Ena Demir, 270-745-6867 or Title IX Investigators, Michael Crowe, 270-745-5429 or Joshua Hayes, 270-745-5121. Please note that while you may report an incident of sex/gender-based discrimination, harassment and/or sexual misconduct to a faculty member, WKU faculty are "Responsible Employees" of the University and MUST report what you share to WKU's Title IX Coordinator or Title IX Investigator. If you would like to speak with someone who may be able to afford you confidentiality, you may contact WKU's Counseling and Testing Center at 270-745-3159.

Regular and Substantive Interaction

In the pursuit of excellence and in accordance with the stipulations set forth by the U.S. Department of Education, all distance education courses, including this one, are designed to foster regular and meaningful interaction between students and faculty. These interactions play a crucial role in not only enhancing your learning experience but also ensuring that you have the requisite support and guidance throughout the course. For a comprehensive understanding of the concept of 'Regular and Substantive Interaction' in the context of distance learning at Western Kentucky University (WKU), we invite you to visit the dedicated webpage titled 'Regular and Substantive Interaction in Online and Distance Learning'. This course is constructed with this principle at its core, ensuring an interactive learning environment that supports your academic growth and progress. Throughout the term, you will have ample opportunities to connect, collaborate, and communicate with faculty and peers. Details regarding the structure, frequency, and modes of interaction for this course will be available on our Blackboard site, allowing you to prepare for and participate fully in these important aspects of your learning experience.

After finishing this course, you can go over the content of this course via Blackboard and then

get ready to take the certificate. Note that there is a free course from RapidMiner that covers both theory and hands-on practice with the basic techniques for building correctly validated Machine Learning Models. It will train you on some of the most common model types and how to build RapidMiner processes to train and evaluate those models.

This course can help you prepare for the Machine Learning Professional Exam. It doesn't cover the answers to all questions on the exam; instead, it asks you to take ownership for learning, understanding and practicing in the topics that this certificate will examine you on. To prepare you to answer questions in the Machine Learning Professional Exam. Here are some useful links to consider before taking the Exam:

- Free Training Course: https://academy.rapidminer.com/learn/course/machine-learning-professional/introduction/welcome?page=1
- Examination Guide can be accessed from this link