Predictive Modeling (BDAN 420)
SYLLABUS

Associate Professor: *Dr. Lily Popova Zhuhadar*	**Office:** *GH 226*
Email: *lily.popova.zhuhadar@wku.edu*	

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, regularization and 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 Learning Professional 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 with an 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 Posters Competition***.

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 poster presentation.

Examples of Capstone Analytics Modeling Theme

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.

Marketing Analytics Theme

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

Financial Analytics Theme

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

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.

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.

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

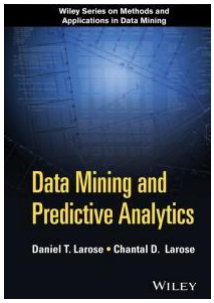
Throughout the semester, this course' conceptual coverage is complemented with application case studies (examples of both successes and failures), as well as simple, hands-on tutorials.

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

- What Tasks Can Accomplish
- Description
- Estimation
- Classification
- Clustering
- Prediction
- Association

Module 2: Data Preparation

- Why do We Need to Preprocess the Data?
- Data Cleaning
- Handling Missing Data
- Identifying Misclassifications
- Graphical Methods for Identifying Outliers
- Data Transformation
- Min–Max Normalization, Z-Score Standardization
- Transformations to Achieve Normality
- Numerical Methods for Identifying Outliers
- Transforming Categorical Variables into Numerical Variables
- Binning Numerical Variables
- Reclassifying Categorical Variables
- Removing Variables that are not Useful

Module 3: Preparing to Model the Data

- Supervised Versus Unsupervised Methods
- Statistical Methodology and Data Mining Methodology
- Cross-Validation
- Overfitting
- Bias–Variance Trade-Off
- Balancing the Training Data Set
- Establishing Baseline Performance

PART II - A QUICK REVIEW OF STATISTICAL MODELING

Module 4: Statistical Modeling - Linear Regression

- A Quick 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
- Hands-On Analysis Using Logistic Regression

Module 7: Neural Networks

- Input and Output Encoding
- Neural Networks for Estimation and Prediction
- Simple Example of a Neural Network
- Sigmoid Activation Function
- Back-Propagation
- Gradient-Descent Method
- Back-Propagation Rules
- Example of Back-Propagation
- Termination Criteria
- Learning Rate
- Momentum Term
- Sensitivity Analysis
- Application of Neural Network Modeling

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 CLUSTERING

Module 9: Hierarchical and k-Means Clustering

- Hierarchical Clustering Methods
- Single-Linkage Clustering
- Complete-Linkage Clustering
- k-Means Clustering
- Example of k-Means Clustering at Work
- Behavior of MSB, MSE, and Pseudo-F as the k-Means Algorithm Proceeds
- Application of k-Means Clustering Using SAS Enterprise Miner
- Using Cluster Membership to Predict Churn

PART V ASSOCIATION RULES

Module 10: Market Basket Analysis

- Affinity Analysis and Market Basket Analysis
- Data Representation for Market Basket Analysis
- Support, Confidence, Frequent Itemsets, and the a Priori Property
- Generating Frequent Itemsets
- Association Rules are Easy to do Badly
- How can we Measure the Usefulness of Association Rules?
- Local Patterns Versus Global Models

PART VI Auto Machine Learning (AutoML)

Module 11: 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.

Adjustment to Assignments, Schedule, and Syllabus

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 periodically checking the syllabus on Blackboard Course Site.
- Grades will be posted promptly on Blackboard after each assignment has been completed.
- 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 to receive a grade.

- A malfunction of your personal computer is not a reason for not completing a test or assignment 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 was not 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" situations every 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.

Email Etiquette

- Each student must have a WKU email address and Blackboard account for participation. Email and Blackboard are considered official communication tools for this course. 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, please ensure 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, I usually 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 include BDAN-420 as well as your last name in the subject line of the e-mail.

Academic Dishonesty

- Students who commit any act of academic dishonesty will receive from the instructor a failing 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 few words is also plagiarism,” for more details, refer to https://www.wku.edu/undergraduatecatalog.*

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.

University Policies

ADA ACCOMMODATIONS

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.

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.

COVID-19

All students are strongly encouraged to get the COVID-19 vaccine. In accordance with WKU policy, all students must call the WKU COVID-19 Assistance Line at 270-745-2019 within 4 hours of testing positive for COVID-19 or being identified as a close contact to someone who has tested positive. The COVID Assistance Line is available to answer questions regarding any COVID-19 related issue. This guidance is subject to change based on requirements set forth by federal, state, and local public health entities. Please refer to the Healthy on the Hill website for the most current information.
www.wku.edu/healthyonthehill

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/heretohelpemail.php. If you need immediate help, please visit the Counseling Center in 409 Potter Hall or call the 24-hour emergency help line at 270-745-2548.

Big Red Backpack

The Big Red Backpack program is the title of WKU's partnership with Barnes & Noble College for their First Day Complete initiative. Big Red Backpack is a course materials delivery program that ensures students have their required course materials on the first day of class and may reduce the costs of materials for many of their courses. Enrollment in the program occurs automatically when students register for classes. Students may opt out of the program each semester, as long as they choose to opt out for all of their classes.

Students MUST carefully and completely follow all instructions regarding the Big Red Backpack Program. Those instructions will be provided by:

1. An email to all Topper email addresses from the Big Red Backpack Program one month prior to the start of classes.
2. The Big Red Backpack Program site: https://www.wku.edu/bigredbackpack/.

Students should refer to the Big Red Backpack Program site referenced above for opt-out deadlines and any additional details.

How to get Certified in

Machine Learning Professional by RapidMiner?

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.

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