Course Information
Fall 2020 – 8W1: August 24 - October 16
Class Meeting Time/Location – We will meet in room 124 / Hall Park B on Thursdays from 6:30 CST to 9:20 CST. The class will be a hybrid course, a mix of in-person and online content delivery.

Instructor Contact & Communication
- Dr. John Garcia, DBA, MS, CAP, CPA
- Office hours: Monday, Wednesday, Friday 4 pm – 6 pm (by appointment).
  - I am happy to meet with students for academic advising, help with assignments, or just to chat.
  - Using this link, you can schedule time with me.
- John.Garcia2@unt.edu / (469) 296-8426 (call or text)
  - We will also be using Microsoft Teams to communicate with each other.
- You can also send me anonymous feedback through this survey form.

About the Professor
Welcome to Data Analytics 2! I am Dr. John Garcia, the professor for this course. Before joining the UNT faculty in January 2020, I was a Finance & Analytics executive at Toyota, where I worked for 15 years in various finance and analytics roles. Before Toyota, I worked at Ernst & Young for six years. I hope to bring my 21 years of industry experience to the classroom and provide you not only the theoretical background but also the practical implications that you will see in business. Like most of my students, I have not followed the traditional academic path, which I think enables me to see both the academic and practitioner viewpoints, thereby helping me add practical, real-world, meaning to the textbook material. The variety of career and academic experiences that we all bring will also provide the foundation for exciting course discussions.

I have a Doctorate degree in Finance & Analytics from Creighton University, a Masters in Predictive Analytics from Northwestern University, a Masters in Accountancy from the University of Notre Dame, and I received my Bachelor’s degree in business from Cal Poly in San Luis Obispo, CA.

I am excited to have you in this course, and I look forward to learning more about you and your career goals. Together we will explore a variety of data mining and predictive modeling techniques, learn about how and when to use them, interpret the outputs of the analysis, and describe the results in ways that will help us or others take appropriate actions to achieve the desired outcomes or goals. I believe data science is also part art as it is impossible to consider data divorced from people, and understanding how to influence people is more art than science. Thus, throughout the course, we will also discuss the art of data science. I look forward to our learning journey!

Course Pre-requisites, Co-requisites, and/or Other Restrictions
Data Analytics 1
Required Materials

The required textbook is listed below. There is also a recommended book (which you can access for free via UNT). We will be using SAS JMP Pro (instructions on how to install will be provided in Canvas).


Recommended Resource


Guidelines for Success

While unforeseen events do happen that can make college life and achievement difficult, generally speaking success is a choice. In order to help yourself and others succeed, avoid distractions during class (like cell phones, iPads, social media, or games), and utilize the resources at your side. You have many valuable resources: textbooks, college services, each other, and myself. Don’t hesitate to ask for help and always communicate. Be sure to read your assigned readings, be punctual, and save all your assignments (and back them up!) Follow these guidelines and you’ll be well on your way.

Course Description

As organizations look for ways to leverage data to create value, analytics has become an important source of competitive advantage for businesses. This course extends the concepts developed in Data Analytics I by providing a hands-on introduction to the collections of predictive modeling techniques used to extract patterns and trends from data. The topics covered include data manipulation, predictive analytics, and decision making under uncertainty. The course includes hands-on work with data and the SAS JMP statistical software package. By the end of the course, you will able to identify situations concerning the applicability of the predictive modeling techniques, employ the techniques to derive results, interpret the results and comprehend the limitations, if any, of the final outcome.

Course Objectives

By the end of the course, students should be able to:

1. Define and explain the business analytics process (problem definition; data preparation; technical analysis and modeling; evaluation of results; implementation and deployment).
2. Investigate data sets, identifying appropriate transformations, creating new variables, and interpolating missing values.
3. Identify and fit appropriate supervised data mining models, such as multivariate regression, logistic regression, non-linear regression, partial least squares regression, k-nearest neighbor, discriminant analysis, naïve Bayes, neural networks, and decision trees.
4. Understand and utilize unsupervised models, including principal components analysis and cluster analysis.
5. Develop analytic solutions to practical business problems using the JMP statistical programming language to transform data into knowledge.
Course Topics

1. Survey of multivariate analysis methods
2. Overview of regression models (multiple linear regression, partial least squares regression, penalized regression, nonlinear, generalized linear models (logistic, poisson, negative binomial), regression trees, time series analysis)
3. Overview of classification methods (logistic regression, neural nets, discriminant analysis, decision trees, Naïve Bayes, k-nearest neighbor, discriminant analysis)
4. Imputations, data pre-processing techniques, and dimension reduction (Principal component analysis)
5. Cluster Analysis
6. Automated Machine Learning

Course Requirements

Your final grade will be determined based on the assignments noted in the table below. The total number of points received will be divided by the total possible number of points.

<table>
<thead>
<tr>
<th>Assignments</th>
<th>Points Possible</th>
<th>Percentage of Final Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quizzes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 quizzes based on the course materials @ 10 pts. each</td>
<td>150 points</td>
<td>15.0%</td>
</tr>
<tr>
<td>6 in-class quizzes @ 5 points each</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Discussions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Six discussions boards @ 15 pts. each</td>
<td>95 points</td>
<td>9.5%</td>
</tr>
<tr>
<td>Self Introduction @ 5 pts</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Final Exam</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehensive exam covering the material covered during the course.</td>
<td>100 points</td>
<td>10.0%</td>
</tr>
<tr>
<td><strong>Project Assignments</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eleven projects @ 30 pts. each</td>
<td>330 points</td>
<td>33.0%</td>
</tr>
<tr>
<td><strong>Final Project</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>325 points</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total Points Possible</strong></td>
<td>1000 points</td>
<td>100%</td>
</tr>
</tbody>
</table>

Grading

Course grades will be assigned as follows (100% — 89.5%, A; 89.5% — 79.5%, B; 79.5% — 69.5%, C; 69.5% — 60%, D; <60%, F).

Course Assignment, Examination, and or Project Policies

**Quizzes** (150 points, 12 asynchronous quizzes at 10 points each / 6 in class quizzes at 5 points each)
- There will be a quiz for each module. Quizzes will be worth 10 points each and may be taken only once.
- The quizzes will be multiple questions designed to reinforce the textbook content.
- Quizzes need to be completed by the due date as indicated on the course schedule.
- Once you start the quiz, you will have 30 minutes to complete it.
- There will also be 6 in-class live quizzes covering the readings and required videos from each week. The in-class quizzes will be completed using iClicker and consists of multiple choice questions. Students choosing to attend remotely will be able to complete it remotely. If you are unable to join the class synchronously and miss the quiz, you will be awarded the average for all your other quizzes for up to 1 missed quiz (missed quiz 2+ will earn a 0).
Discussions (95 points, 6 at 15 points each & self-introduction at 5 points)

- There will be six video group discussions on the posted weekly article assignment. Each video discussion cannot be longer than 5 minutes, with the minimum as follows: ADTA 5230 – 2 minutes, IPAC 4230 - 1.5 minutes. A grading rubric will be provided in Canvas.
- These are reflective in nature and are designed for you to share your thoughts and experiences related to the topic presented.
- There will be pre-reading assigned with each journal assignment that will be provided by your professor.
- It is expected that your responses be thoughtful, well structured, and show your understanding of the topic being discussed.
- It is required that you post your initial response by Wednesday and that you provide a substantive response to at least two members of your group by the end of day Saturday.
- We will be using Flipgrid for course discussion. Flipgrid is a video-based discussion board. You will video your response and then can reply to your peers via a video.

Project Assignments (330 points, 11 at 30 points each)

- To help students gain hands-on experience in applying statistical learning techniques using JMP, this course will use twelve individual project assignments.
- It is recommended that you start these projects after having completed the JMP tutorials and in-class exercises.
- You may start the assignment at any time before the due date.
- You will have 3 hours to complete the JMP individual project assignment from the time you start it.

Final Project (325 points / 250 points for the final report / 75 points for the presentation).

- The project will be conducted individually.
- Also, part of the final project is a 10-minute presentation.
- The minimum requirement for the paper will be a minimum of:
  - ADTA 5230: 12 pages, double-spaced, 1-inch margins, using Arial or Times Roman 12-point font.
  - IPAC 4230: 8 pages, double-spaced, 1-inch margins, using Arial or Times Roman 12-point font.
- The submitted research paper should also include a separate cover page (not part of the page minimum) that includes your names and the title of your paper. References should use APA formatting.
- The paper is due Wednesday, October 14th at 11:59 pm CST & the presentation is due October 15th, at 6 pm CST. Late papers/presentations will not be accepted.
- The paper will be submitted for grading via software that checks for plagiarism. Plagiarism is a violation of the Student Code of Conduct and will be handled per university policy.
- A grading rubric will be provided in Canvas.

Final exam (100 points)

- There will be a final exam. You will have 150 minutes to complete the exam.
- The exam will be a mix of multiple questions and problems you solve using JMP and will be an open book and notes exam.
- The exam will be open in Canvas from Wednesday, October 14th through Friday, October 16th, 11:59 pm. You must log on to Canvas and take the exam during this time window.

Late Assignment Policy

All work for this course is due no later than 11:59 pm on the designated due. Any assignment submitted after that time will receive a highest possible score of up to 60% through 48 hours past the deadline. Additional points may be deducted when the assignment is graded based on the quality of the work submitted. No points will be awarded for assignments turned in 48 hours or more past the due date. Please don’t lose valuable points this semester by turning in work late. Late work is subject to the penalty described above unless previously approved by the instructor.
Class Participation

Students are encouraged to login regularly to the online class site and attend class virtually or in person. Students are also required to participate in all class activities such as discussion boards, chats, and synchronous sessions.

COVID-19 Impact on Attendance

While attendance is expected as outlined above, it is important for all of us to be mindful of the health and safety of everyone in our community, especially given concerns about COVID-19. Please contact me if you are unable to attend class because you are ill, or unable to attend class due to a related issue regarding COVID-19. It is important that you communicate with me prior to being absent so I may make a decision about accommodating your request to be excused from class.

If you are experiencing cough, shortness of breath or difficulty breathing, fever, or any of the other possible symptoms of COVID-19 (https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html) please seek medical attention from the Student Health and Wellness Center (940-565-2333 or askSHWC@unt.edu) or your health care provider. UNT also asks that you contact the UNT COVID Hotline at 844-366-5892 or COVID@unt.edu for guidance on actions to take due to symptoms, pending or positive test results, or potential exposure. While attendance is an important part of succeeding in this class, your own health, and those of others in the community, are more important.

The UNT fall schedule requires this course to have fully remote instruction beginning November 28th. Additional remote instruction may be necessary if community health conditions change or you need to self-isolate or quarantine due to COVID-19. Students will need access to a webcam and microphone to participate in fully remote portions of the class. Information on how to be successful in a remote learning environment can be found at https://online.unt.edu/learn.

Statement on Face Covering

Face coverings are required in all UNT facilities. Students are expected to wear face coverings during this class. If you are unable to wear a face covering due to a disability, please contact the Office of Disability Access to request an accommodation. UNT face covering requirements are subject to change due to community health guidelines. Any changes will be communicated via the instructor.

Turnitin Notice

Turnitin is used as a tool to assist students in their scholarly writing to address plagiarism issues. All works submitted for credit must be original works created by the scholar uniquely for the class. It is considered inappropriate and unethical, particularly at an advanced undergraduate/graduate level, to make duplicate submissions of a single work for credit in multiple classes, unless specifically requested by the instructor. It is also considered inappropriate and unethical to work together on individual assignments or share work that is to be created on an individual level. Work submitted at the senior/graduate level is expected to demonstrate higher-order thinking skills and be of significantly higher quality than work produced at the lower undergraduate levels. It is recommended that students use the Turnitin resource to ensure their work is free of copyright issues prior to the final submission of their projects.

Grades of Incomplete

Grades of Incomplete will only be given per university policy as outlined by the Office of the Registrar.

Additional UNT Policies

Please review and familiarize yourself with the additional UNT policies outlined in the Canvas site.
Course Calendar

Below is a tentative schedule. Should any change become necessary, it will be announced in class as well as in the announcements sent via the UNT email. It is the student’s responsibility to check for changes in the schedule.

<table>
<thead>
<tr>
<th>Week</th>
<th>Topic / Reading</th>
<th>Assignments</th>
</tr>
</thead>
</table>
| Week 1<br>Aug 24<br>(Module 1) | *Course Overview, Introduction to Data Mining, Data Visualization and Dimension Reduction*  
Course overview, Syllabus review, and the CRISP-DM analytics process. Read the following chapters in Data Mining for Business Analytics (DMBA) and the recommended Applied Predictive Modeling (APM).  
**DMBA Required:**  
1. Chapter 1-2: Introduction & Overview of Data Mining  
2. Chapter 3: Data Visualization  
3. Chapter 4: Dimension Reduction  
**APM Recommended:**  
1. Chp 3 – Data Pre-Processing (p. 27-50)  
2. Chp 4 – Over Fitting and Model Tuning (p. 61-80) | Quiz #1 – Chps 1-2  
Quiz #2 – Chps 3-4  
Self Introduction  
Project #1 - EDA |
| Week 2<br>Aug 31<br>(Module 2) | *Linear, Partial Least Squares Regression, & Generalized Linear Model Extensions (Logistic, Poisson & Negative Binomial, Penalized Regression)*  
**DMBA Required:**  
1. Chapter 5: Evaluating Predictive Performance  
2. Chapter 6: Multiple Linear Regression  
3. Chapter 10: Logistic Regression  
**APM Recommended:**  
1. Chp. 5 – Measuring Performance in Regression Models (p. 95-100)  
2. Chp. 11 – Measuring Performance in Classification Models (p. 247-266)  
3. Chp. 6 – Linear Regression and Its Cousins (p. 101-122) | Class Quiz #1 (on week 2 required material)  
Project #2 – Linear Regression  
Project #3 – Logistic Regression  
Quiz #3 – Chp 5-6  
Quiz #4 – Chp 10  
Discussion #1 |
| Week 3<br>Sep 7<br>(Module 3) | *k-Nearest Neighbors (k-NN) & Naïve Bayes Classifier*  
**DMBA Required:**  
1. Chapter 7: k-Nearest Neighbors  
2. Chapter 8: The Naïve Bayes Classifier  
**APM Recommended**  
1. Chp. 13.5-13.5 – Nonlinear Classification Models (p. 350-358)  
*Extra Credit – KNN & NB (10 pts) / due 9/13* | Class Quiz #2  
Project #4 KNN  
Project #5 NB  
Quiz #5 – Chp 7  
Quiz #6 – Chp 8  
Discussion #2 |
<table>
<thead>
<tr>
<th>Week</th>
<th>Topic / Reading</th>
<th>Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Week 4</strong>&lt;br&gt;&lt;br&gt;Sep 14&lt;br&gt;(Module 4)</td>
<td><strong>Decision Trees and Tree Ensembles</strong>&lt;br&gt;DMBA Required:&lt;br&gt;1. Chapter 9: Classification and Regression Trees&lt;br&gt;APM Recommended&lt;br&gt;1. Chp. 8 – Regression Trees and Rule-Based Methods (p. 173-192)</td>
<td>Class Quiz #3&lt;br&gt;Project #6 – DT&lt;br&gt;Project #7 – Tree Ensembles&lt;br&gt;Quiz #7 – Decision Trees&lt;br&gt;Quiz #8 – Tree Ensembles&lt;br&gt;Discussion #3</td>
</tr>
<tr>
<td><strong>Week 5</strong>&lt;br&gt;&lt;br&gt;Sep 21&lt;br&gt;(Module 5)</td>
<td><strong>Neural Networks &amp; Discriminant Analysis</strong>&lt;br&gt;DMBA Required:&lt;br&gt;1. Chap 11: Neural Nets&lt;br&gt;2. Chap 12: Discriminant Analysis&lt;br&gt;APM Recommended:&lt;br&gt;1. Chp. 13.2 – Neural Networks (p. 333 – 338)&lt;br&gt;2. Chp. 12.3 – Linear Discriminant Analysis (p. 287-297)</td>
<td>Class Quiz #4&lt;br&gt;Project #8 – Neural nets&lt;br&gt;Project #9 - Discriminant Analysis&lt;br&gt;Quiz #9 – Chp. 11&lt;br&gt;Quiz #10 – Chp. 12&lt;br&gt;Discussion #4</td>
</tr>
<tr>
<td><strong>Week 6</strong>&lt;br&gt;&lt;br&gt;Sep 28&lt;br&gt;(Module 6)</td>
<td><strong>Cluster Analysis</strong>&lt;br&gt;1. DMBA Chap 14: Cluster Analysis</td>
<td>Class Quiz #5&lt;br&gt;Project #10– Cluster Analysis&lt;br&gt;Quiz #11 – Chp. 14&lt;br&gt;Discussion #5</td>
</tr>
<tr>
<td><strong>Week 7</strong>&lt;br&gt;&lt;br&gt;Oct 5&lt;br&gt;(Modules 7)</td>
<td><strong>Time Series, Regression Forecasting, Smoothing Methods</strong>&lt;br&gt;DMBA Required:&lt;br&gt;1. Chap 15: Handling Time Series&lt;br&gt;2. Chap 16: Regression-Based Forecasting&lt;br&gt;3. Chap 17: Smoothing Methods</td>
<td>Class Quiz #6&lt;br&gt;Project #11 – Time Series&lt;br&gt;Quiz #12 – Chp. 15-17&lt;br&gt;Discussion #6</td>
</tr>
<tr>
<td><strong>Week 8</strong>&lt;br&gt;&lt;br&gt;Oct 12&lt;br&gt;(Module 8)</td>
<td><strong>Presentations &amp; Final Project</strong>&lt;br&gt;Project Presentations and Paper&lt;br&gt;<em>Complete Final Exam between October 14th and 16th</em></td>
<td>Final Grp. Project Report &amp; Presentation&lt;br&gt;Final Exam</td>
</tr>
</tbody>
</table>
SCHOLARLY EXPECTATIONS

Copyright Notice
Some or all of the materials on this course web site may be protected by copyright. Federal copyright law prohibits the reproduction, distribution, public performance, or public display of copyrighted materials without the express and written permission of the copyright owner, unless fair use or another exemption under copyright law applies. Additional copyright information may be located at: http://policy.unt.edu/policy/08-001.

UNT Code of Student Conduct
You are encouraged to become familiar with the University's Code of Student Conduct and the Policy of Academic Integrity (Links to an external site.) found on the Dean of Students website. The Dean of Students Office (opens in a new window) (Links to an external site.) enforces the Code. The Code explains what conduct is prohibited, the process the DOS uses to review reports of alleged misconduct by students, and the sanctions that can be assigned. When students may have violated the Code they meet with a representative from the Dean of Students Office to discuss the alleged misconduct in an educational process. The University's expectations for student conduct apply to all instructional forums, including University and electronic classroom, labs, discussion groups, field trips, etc.

Of particular interest are the following terms:

• **Cheating** – intentionally using or attempting to use unauthorized materials, information, or study aids in any academic exercise. The term academic exercise includes all forms of work submitted for credit or hours.
• **Plagiarism** – the deliberate adoption or reproduction of ideas, words, or statements of another person as one’s own without acknowledgement.
• **Fabrication** – intentional and unauthorized falsification or invention of any information or citation in an academic exercise.
• **Facilitating academic dishonesty** – intentionally or knowingly helping or attempting to help another to violate a provision of the institutional code of academic integrity.

The policies contained on the course website apply to this course. In addition, you are expected to adhere to the ADTA Academic Integrity Policy outlined below. If you have questions regarding any of the information presented regarding academic integrity, please feel free to contact me.

Academic Integrity
All works submitted for credit must be original works created by the scholar uniquely for the class. It is considered inappropriate and unethical, particularly at the graduate level, to make duplicate submissions of a single work for credit in multiple classes, unless specifically requested by the instructor. Work submitted at the graduate level is expected to demonstrate higher-order thinking skills and be of significantly higher quality than work produced at the undergraduate level.

ADTA Academic Integrity Policy

<table>
<thead>
<tr>
<th>Occurrence</th>
<th>Minor Assignments (e.g., Discussions, Homework, and Journals)</th>
<th>Major Assignments (e.g., Exams and Projects worth more than 10% of your grade)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Warning</td>
<td>1. First written warning</td>
<td>1. Written warning</td>
</tr>
<tr>
<td></td>
<td>2. Min. 20% deduction</td>
<td>2. Min. 15% deduction</td>
</tr>
<tr>
<td>2nd Warning</td>
<td>1. Second written warning</td>
<td>1. Second written warning</td>
</tr>
<tr>
<td></td>
<td>2. Min. 50% deduction</td>
<td>2. Min. 50%</td>
</tr>
<tr>
<td></td>
<td>3. Inform academic advisor during Dept. Meeting</td>
<td>3. Inform academic advisor during Dept. Meeting</td>
</tr>
<tr>
<td>3rd Warning</td>
<td>1. Written Letter</td>
<td>1. Written Letter</td>
</tr>
<tr>
<td></td>
<td>2. Min. 0 grade for that assignment</td>
<td>2. Min. 0 grade for that assignment</td>
</tr>
</tbody>
</table>