ITS450: Data Mining

Credit Hours: 3
Contact Hours: This is a 3-credit course, offered in accelerated format. This means that 16 weeks of material is covered in 8 weeks. The exact number of hours per week that you can expect to spend on each course will vary based upon the weekly coursework, as well as your study style and preferences. You should plan to spend 10-25 hours per week in each course reading material, interacting on the discussion boards, writing papers, completing projects, and doing research.

Faculty Information

Name: 
Phone: 
CSU-GC Email: 
Virtual Office Hours: 

Course Description and Outcomes

In this course, you will investigate various statistical methodologies for knowledge discovery within existing databases of information. You’ll learn about the preparation of data suitable for analysis from enterprise data warehouses using SQL and the documentation of results. You’ll also conduct a practical data-mining analysis project, using SPSS, to reinforce the concepts taught in the course.

Course Learning Outcomes:

1. Analyze how online analytical processing (OLAP) and data mining are utilized to obtain business intelligence (BI) for informed business decisions and plans.
2. Describe the various types of statistical approaches used for data mining of enterprise data warehouse databases in order to establish relationships in customer behavior.
3. Formulate the extract, transform, and load (ETL) processes used to refresh a data warehouse based on a Star Schema.
4. Appraise the purpose of denormalized relational database data stored in materialized views and the role of ad hoc queries.
5. Construct a simple data warehouse with appropriate denormalized data using SQL for input to an SPSS statistical analysis.
6. Perform and interpret a data mining analysis using SPSS.
7. Assemble the findings of a data mining analysis in a professional business oriented manner.
Participation & Attendance

Prompt and consistent attendance in your online courses is essential for your success at CSU-Global Campus. Failure to verify your attendance within the first 7 days of this course may result in your withdrawal. If for some reason you would like to drop a course, please contact your advisor.

Online classes have deadlines, assignments, and participation requirements just like on-campus classes. Budget your time carefully and keep an open line of communication with your instructor. If you are having technical problems, problems with your assignments, or other problems that are impeding your progress, let your instructor know as soon as possible.

Course Materials

Required:


**All non-textbook required readings and materials necessary to complete assignments, discussions, and/or supplemental or required exercises will be provided within the course itself. Please read through each course module carefully.**

Course Schedule

Due Dates

The Academic Week at CSU-Global begins on Monday and ends the following Sunday.

- Discussion Boards: The original post must be completed by Thursday at 12 midnight MT and Peer Responses posted by Sunday 12 midnight MT. Late posts may not be awarded points.
- Mastery Exercises: Students may access and retake mastery exercises through the last day of class until they achieve the scores they desire.
- Critical Thinking Activities: Assignments are due Sunday at 12 midnight MT.

<table>
<thead>
<tr>
<th>Week #</th>
<th>Readings</th>
<th>Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>• Chapters 1 &amp; 2 in <em>Introduction to Data Mining</em></td>
<td>• Discussion (25 points)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Mastery (10 points)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Critical Thinking (50 points)</td>
</tr>
<tr>
<td>2</td>
<td>• Chapter 3 in <em>Introduction to Data Mining</em></td>
<td>• Discussion (25 points)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Mastery (10 points)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Critical Thinking (60 points)</td>
</tr>
<tr>
<td>3</td>
<td>• Chapters 4 &amp; 5 in <em>Introduction to Data Mining</em></td>
<td>• Discussion (25 points)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Mastery (10 points)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Critical Thinking (60 points)</td>
</tr>
<tr>
<td>4</td>
<td>• Chapters 6 &amp; 7 in <em>Introduction to Data Mining</em></td>
<td>• Discussion (25 points)</td>
</tr>
<tr>
<td>Assignment</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
<td></td>
</tr>
</tbody>
</table>
| 5 | Mastery (20 points)  
Chapters 8 & 9 in *Introduction to Data Mining*  
Discussion (25 points)  
Mastery (10 points)  
Critical Thinking (60 points)  
| 6 | Mastery (20 points)  
Discussion (25 points)  
Mastery (10 points)  
Critical Thinking (60 points)  
| 7 | Mastery (20 points)  
Discussion (25 points)  
Mastery (10 points)  
Critical Thinking (60 points)  
- Chapter 10 in *Introduction to Data Mining* |
| 8 | Mastery (20 points)  
Discussion (25 points)  
Mastery (10 points)  
Portfolio (350 points)  

**Assignment Details**

This course includes the following assignments/projects:

**Module 1**

**Critical Thinking: Data Quality, Measurement and Data Collection (50 Points)**

Provide an overview of data quality as it relates to measurements and data collection. How is data quality maintained? What are some of the different types of measurements, and is there a better type of measurement for different types of data? Describe some of the issues that can arise as a result of the collection process. How is the data handled? What are some of the common sources of inaccuracies? Use these questions as a framework for your responses.

Your well-written paper should be 2-3 pages in length and formatted per CSU-Global APA format. Include at least two credible outside references, and one citation from the course textbook, totaling three. The outside references may include credible sources in print or from the Internet. The CSU-Global library is a great place to get sources!
Module 2

**Critical Thinking: Dimension Reduction Methods (60 Points)**

Discuss the differences between dimensionality reduction based on aggregation and dimensionality reduction using techniques such as PCA and SVD. Discuss the different techniques in both cases and how they impact the data. Is one method more accurate than the other? (Hint: Make sure to review Appendix B in the back of the textbook.)

Your well-written paper should be 2-3 pages in length, in CSU-Global APA format. Include at least two credible outside references, and one citation from the course textbook, totaling three. The outside references may include credible sources in print or from the Internet. The CSU-Global library is a great place to get sources!

Module 3

**Critical Thinking: Classifier Analysis (60 points)**

Consider the task of building a classifier from random data, where the attribute values are generated randomly irrespective of the class labels. Assume the data set contains records from two classes, “+” and “−”. Half of the data set is used for training while the remaining half is used for testing.

- Suppose there are an equal number of positive and negative records in the data and the decision tree classifier predicts every test record to be positive. What is the expected error rate of the classifier on the test data?
- Repeat the previous analysis assuming that the classifier predicts each test record to be positive class with probability 0.8 and negative class with probability 0.2.
- Suppose two-thirds of the data belong to the positive class and the remaining one-third belong to the negative class. What is the expected error of a classifier that predicts every test record to be positive?
- Repeat the previous analysis assuming that the classifier predicts each test record to be positive class with probability 2/3 and negative class with probability 1/3.

Respond to each situation and support your response with analysis. Your well-written findings should be 1-2 pages in length and formatted according to CSU-Global APA guidelines.

Module 5

**Critical Thinking: Clustering Algorithms (60 Points)**

Using the attached image (found on the Week 5 Assignments page), respond to the following:

This image shows a clustering of a two-dimensional point data set with two clusters: the left cluster of somewhat diffuse points and the right cluster whose points are more closely compact. To the right of the compact cluster, notice that there is a single point (marked by an arrow) that belongs to the diffuse cluster, whose center is farther away than that of the compact cluster. Explain in detail how this is possible using an Expectation-Maximization algorithm, but not using a K-means algorithm. Examine the steps of each algorithm and identify how the Expectation-Maximization algorithm could catch the outlier point, as well as how K-means cannot.

Your well-written findings should be 1-2 pages in length and formatted according to CSU-Global APA guidelines. Include at least two credible outside references, and one citation from the course textbook, totaling three. The outside references may include credible sources in print or from the Internet. The CSU-Global library is a great place to get sources!
Module 6

Critical Thinking: SQL, Data Warehouse, and Preparing the Analysis (60 points)

For this Critical Thinking assignment, you will Install and configure a SQL server, attach the sample AdventureWorks database to the SQL server, prepare and extract data from the AdventureWorks database, and then report on your findings and propose a plan to analyze the data.

See the Week 6 Assignment Page for complete details.

Module 7

Critical Thinking: Anomaly Detection Method (60 points)

Compare and contrast the different techniques for anomaly detection that were presented in section 10.1.2 of your text. In particular, identify circumstances in which the definitions of anomalies used in the different techniques might be equivalent or situations in which one might make sense but another would not. Be sure to consider different types of data.

Your well-written findings should be 2-3 pages in length and formatted according to CSU-Global APA guidelines. Include at least two credible outside references, and one citation from the course textbook, totaling three. The outside references may include credible sources in print or from the Internet. The CSU-Global library is a great place to get sources!

Module 8

Portfolio Project: Data Mining with SPSS (350 points)

The Portfolio Project, due at the end of Week 8, is a statistical analysis using one or more of the statistical analysis approaches presented in Modules 4 and 5. These various approaches are designed to produce business intelligence to resolve problems and enable management to make informed business decisions.

You will be conducting these analyses based on your evaluation and preparation of the data in Module 6 using the IBM SPSS statistical analysis software. This week’s lecture contains the information you need for downloading and using the SPSS software. Having completed your analysis for the Week 6 Critical Thinking assignment, you will refine and present your findings in the Portfolio Project and associate them to business knowledge that can be used to enhance the decision-making abilities of the organization.

Your Portfolio Project submission should include the following sections:

- **Introduction:** Introduce the organization, the business problem that you defined in the Week 6 Critical Thinking assignment, and the data that was presented to you.
- **Data Preparation:** Discuss the data and any quality and accuracy concerns that you discovered. Describe your approach to processing the data to improve the chances of discovering new knowledge from the existing data.
- **Analysis Approach:** Describe your chosen analysis approach. Include your assessment of how effective the method was in discovering new relevant knowledge within the data.
- **Findings:** Present and discuss your findings as the result of your analysis. Include any supporting data, graphics, tables, charts, etc., along with your analysis of the findings. Discuss any correlations, associations or patterns that you identified in the data.
- **Application:** Briefly discuss your findings as they pertain to the business problem. How might the organization use this new knowledge to make informed business decisions to resolve the business problem?
• **Discussion**: Discuss the limitations of your data mining process in discovering useful business intelligence. Additionally, discuss other research or analysis methods. What methods might another data mining expert consider using should the organization need further analysis on this data and business problem?

Your well-written findings should be 6-8 pages in length and formatted according to CSU-Global guidelines. Include at least five credible outside references, and one or more citations from the course textbook, to support your analysis process. The outside references may include credible sources in print, or from the Internet, screenshots, images, etc. (Be sure to refine your graphics.) The CSU-Global library is a great place to get sources!

**Course Policies**

**Course Grading**

- 20% Discussion Participation
- 10% Mastery Exercises
- 35% Critical Thinking Activities
- 35% Final Portfolio Paper

<table>
<thead>
<tr>
<th>Grading Scale and Policies</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
</tr>
<tr>
<td>A-</td>
</tr>
<tr>
<td>B+</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>B-</td>
</tr>
<tr>
<td>C+</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>D</td>
</tr>
<tr>
<td>F</td>
</tr>
</tbody>
</table>

**In-Classroom Policies**

For information on late work and Incomplete grade policies, please refer to our [In-Classroom Student Policies and Guidelines](#) or the Academic Catalog for comprehensive documentation of CSU-Global institutional policies.

**Academic Integrity**

Students must assume responsibility for maintaining honesty in all work submitted for credit and in any other work designated by the instructor of the course. Academic dishonesty includes cheating, fabrication, facilitating academic dishonesty, plagiarism, reusing /re-purposing your own work (see *CSU-Global Guide to Writing and APA Requirements* for percentage of repurposed work that can be used in an assignment), unauthorized possession of academic materials, and unauthorized collaboration. The CSU-Global Library provides information on how students can avoid plagiarism by understanding what it is and how to use the Library and Internet resources.

**Citing Sources with APA Style**

All students are expected to follow the *CSU-Global Guide to Writing and APA Requirements* when citing in APA (based on the APA Style Manual, 6th edition) for all assignments. For details on CSU-Global APA style, please review the APA resources within the CSU-Global Library under the “APA Guide & Resources” link. A link to this document should also be provided within most assignment descriptions on your course’s Assignments page.
Netiquette
Respect the diversity of opinions among the instructor and classmates and engage with them in a courteous, respectful, and professional manner. All posts and classroom communication must be conducted in accordance with the student code of conduct. Think before you push the Send button. Did you say just what you meant? How will the person on the other end read the words?

Maintain an environment free of harassment, stalking, threats, abuse, insults or humiliation toward the instructor and classmates. This includes, but is not limited to, demeaning written or oral comments of an ethnic, religious, age, disability, sexist (or sexual orientation), or racist nature; and the unwanted sexual advances or intimidations by email, or on discussion boards and other postings within or connected to the online classroom.

If you have concerns about something that has been said, please let your instructor know.