University of La Verne

COMPUTER SCIENCE & COMPUTER ENGINEERING PROGRAM

Central Campus, Spring 2022

CMPS 386 INRODUCTION TO DATA MINING - 1514

COURSE INFORMATION:

Units: 4.0 Credit Hours
Pre-Req.: CMPS 372 and MATH 201
Schedule Types: Lecture/Seminar
Requirements: Computer Science B.S. Elective
Class Location: Synchronous Zoom online meetings
Course Time: Lecture/Seminar: Wednesday 6:50 – 10:00 p.m.

INSTRUCTOR INFORMATION:

Instructor: Professor Jozef Goetz Ph.D.
Office: Zoom
E-mail: JGoetz@laverne.edu
Phone: (909) 448-4663
Office Hours: M: 2:30 – 4:30 p.m. on Zoom/WebEx by appointment at https://ulvadvising.as.me/jgoetz.

COURSE DESCRIPTION:

Data Mining has wide applications in many different fields in business, science, engineering, education, and many more. The course provides a broad and interactive overview of a rapidly growing field. This explosive growth in data has generated an urgent need for new techniques and tools that can intelligently and **automatically extract hidden patterns** in large databases. **Database mining** is not only **evolutionary** but also **revolutionary**. It is a **new** discipline lying at the interface of statistics, database technology, pattern recognition, machine learning, AI and other areas. The automated, prospective analyses offered by database mining are far beyond the analyses of past events provided by typical of decision support systems.

The course provides a broad and interactive overview of a rapidly growing field. The objective of this course is **focused on data mining,** with an **emphasis on data visualization using Excel, SQL, NoSQL, Python, and R programming**. **Case studies** and hands-on activities incorporate **real-world data sets** and allow students the opportunity to exercise their new skills**. The importance of** data visualization **for communication purposes is explored, as are** the **processes of data cleansing, clustering,** and **classification. Excel, SQL, NoSQL, Python, and R programming all receive in-depth treatments, supplemented with hands-on exercises.**

Topics include: Data Mining, Machine Learning, Databases SQL and NoSQL, Data Visualization, Programming Data Analytics, Data Preprocessing and Cleansing, Clustering, Classification, Predictive Analytics, Data Association, Mining Text and Images. Tools: JupyterLab, MySQL Workbench or phpMyAdmin, MongoDB, RapidMiner, Weka.

COURSE OBJECTIVES:

a. Specific outcomes of instruction:

- Demonstrate an understanding data mining and data visualization
- Perform key data-mining and machine-learning operations
- Understand databases
- Use data visualization to create charts
- Create charts to visualize data using Excel and HTML
- Perform complex SQL queries.
- Perform NoSQL query operations
- Use Python and R to perform common machine-learning and data-mining operations

- Demonstrate data cleansing, data clustering, data classification
- Demonstrate predictive analysis
- Demonstrate data association

b. Outcomes addressed by the course:

Course Contribution	Student Learning Outcomes
*	1. Ability to **analyze** a complex computing problem and to **apply** principles of computing and other relevant disciplines to identify solutions (**AA**).
*	2. Ability to **design, implement, and evaluate** a computing-based solution to meet a given set of computing requirements in the context of the program's discipline (**DIE**).
	3. Ability to **communicate effectively** in a variety of professional context (**CE**).
	4. Ability to recognize professional responsibilities and make informed judgment in computing practice based on **legal** and **ethical** principles (**LE**).
	5. Ability to function effectively as a member or **leader** of a **team** engaged in activities appropriate to program's discipline (**LT**).
*	6. Ability to **apply** computer science **theory** and software **development** fundamentals to produce computing-based solutions (**ATD**).
	7. Ability to apply **security principles** and practices to maintain operations in the presence of **risks** and **threats (SPRT)**.

📖 **REQUIRED TEXT:**
[1] Kris Jamsa, Introduction to Data Mining and Analytics, **2021**, Jones & Barlett Learning, ISBN-13: 9781284180909.

✍ **EVALUATION AND GRADING:**
The course grade will be calculated as follows:

Programming projects, quizzes and lab	**50%**
Midterm	**25%**
Final Exam	**25%**
TOTAL	**100%**

Final course grades will be assigned as follows:

94 – 100 = **A**	90 – 93 = **A-**	87 – 89 = **B+**
84 – 86 = **B**	80 – 83 = **B-**	77 – 79 = **C+**
74 – 76 = **C**	70 – 73 = **C-**	67 – 69 = **D+**
64 – 66 = **D**	0 - 63 = **F**	

NATURE OF ACTIVITES IN THE CLASS

1. **Time spend outside of class:**
In order to gain genuine knowledge and skills you should know that for every **one credit hour** in which you enroll, you need to spend **approximately two to three hours** outside of class **studying** and **working** on **assignments** for the course. So, students should plan to work **at least 6 - 12 hours** per week outside of class. You need to be aware that approximately **33.4%** of your learning will take place in class with the remaining **66.6%** at home.

2. ☺**Collaboration:**

One of the goals **of studying at the university is to learn how to learn. Learning is a long life process. One of the** computer-science educational methods is an **Extreme Learning** method. Extreme Learning integrates **problem-based learning, pairing learning, collaborative learning** practices to help students **gain more hands-on** experience and **in-depth knowledge** on specific topics. **Collaborative** learning in pairs allows **students to open interaction, educate each other** and **share ideas, knowledge** and **experience.**

Guidelines:

a. You should use the **Extreme Learning** method by giving each other technical support, help understand the assignment and brainstorm general solution but each student must submit **your own detailed project solution.**
b. Each member of the group project should be able to explain any part of the submission, and **not just be able to explain "his or her" part.**

3. 🔔 👆**Attendance and Participation:**

Required and verified. Attendance and class participation are extremely important in this course. You should **notify the instructor in advance of your absence** from the scheduled course meeting. **It is essential** that you **attend all lectures and labs to succeed in the course.** Regardless of excuse, absences in excess of **three week classes** will result in the automatic exclusion of the student from that class and you will receive a **grade of F.** If you are absent from class, it is your responsibility *to make-up* any missed classes and check on announcements made while you were absent. **It is essential that you attend all lectures and labs to succeed in the course.**

You have to **read** corresponding sections in the **textbook,** which will be covered at the **next** class meeting. In addition to that, **after** each lecture/lab session you should study sections in the **textbook one more time** and **run all corresponding programs.**

4. 🔔 **Timeliness:**

You are expected to be in your seat and ready to begin class promptly at the start of each class. **Don't leave the class before class ends.** When students do that, it **negatively affects** the **whole class.** It is **distracting** and **rude, and sends a message** that the **material is easy, which is not true.** Schedule your day such that you may manage contingencies (such traffic, doctor appointments, etc.) when they occur. The instructor maintains the discretion to mark you absent for all or part of the class in the event you fail to be timely and prompt.

5. ☺ **Class Contribution:**

Class Contribution in the form of presentation your final project, **comments** that relate to material in the text and **answering a question** asked by the professor or another student counts for **extra points** of your grade in this course. These are the behaviors to avoid:

- not listening
- pretending to be listening while texting or cruising online
- speaking without being recognized
- making fun or otherwise berating something said by another person.

6. ◘ **Quizzes:**

Brief quizzes will be given during the semester. The content will relate to the material covered in the lectures and assigned readings. Please **attend class regularly and keep up** with course material. **No-makeup quizzes** are allowed.

6. ✍**Lab, home and project assignments**:

Several lab and project assignments will be given over the course of the semester. An electronic version of the project assignments can be downloaded from the course's website. All assignments will be graded on a scale from 0 to 20 after presenting the assignments to the instructor. Expect one to two quick questions to show your

understanding. You will receive a **score of zero** if **falsified input/output** that doesn't much the source code or submissions that are plagiarized or that violate the collaboration guidelines.

You will need to turn in your projects according to the description found in **1_Project Submittals.doc** at **http://classes.jgspectrum.com/classes/386_S22/Guidelines/.** Please do not attempt to **turn in any lab assignment by email. No credit will be given for such work.** Each project assignment will be posted on **OneDrive** under **386_Introduction_to_Data_Mining/S22_386_YourName/PROJ_ch# or ASS#**. The **PROJ_ch#** directory should contain only **project word document** and **project directory** with all source files for the current chapter project.
Class and **project assignments** are the **key** to your **success**. Don't expect to learn or have a good grade if you miss classes and/or home assignments. You will **build** your **knowledge** and **skills based** on the **previous classes** and **project assignments**. You will need to **create** and submit the **final** project of your choice. At the end of the semester you will present your final project to the class.

You should be **ready to show** project solution to your **professor** at the very **beginning** of class**.** Please check the **Assignment.doc** every time for all assignment specifications. The **Assignment.doc** serves as a **staring point** to any assignment solution.

7. ☻ **Make-up and late assignments**.
No credit will be given for assignments turned in after the due day specified in Assignment.doc. Assignments MUST be submitted **before class begins** on the due date**. No-makeup assignments are allowed. Do not get left behind**. Unless extraordinary circumstances can be documented, no assignments will be accepted after the beginning of class on the day the assignment is due. **No assignments will be accepted after they have been handed back or reviewed in class.**

8. ✍ **Midterm and Final Exams:**
There will be two exams to complete the course work and obtain a grade for the course. **There will be no make-ups for the midterm and final examinations**. If you are absent from a **midterm** and have a **valid excuse**—an illness, a death in your family, injury or another equally compelling reason—the weight of your final will be increased by the weight of the midterm. You must provide **adequate** and **verifiable** documentation. Without a valid excuse, you will receive a **zero score for** the **midterm** and the final's weight will remain unchanged. A missed **final** will be dealt with according to University regulations on incompletes and withdrawals. Midterm and final **exams** will cover specified chapters.

9. 🗁 **Course material:**
All handouts, my syllabus, guidelines, lecture notes, links and assignments will be posted at **http://classes.jgspectrum.com/**. Click the top menu item **Classes** you will see a folder labeled **CMPS 386: Introduction to Data Mining**, and you will find all **CMPS 386** documents there. You may copy them to your computer.

10. ♫ **Email Policy:**
I usually reply to emails that require a fast answer within 24-36 hours on weekdays. I will not reply to email messages that are unclear or disrespectful. Please include your **class name** and **section** in the **subject field** and a **salutation** (e.g. Dear Professor Goetz), so that it is clear that the message is not junk mail and deleted.
Students must check their e-mail messages on a daily basis. I will only use your Laverne e-mail address.

11. 𝅘𝅥𝅮 **ZOOM OR WEBEX ETIQUETTE & TIPS**
Face to face online: This teaching modality expects students to be highly motivated and disciplined.

1. I ask you the **webcam** be continuously **on** for attendance purpose.
2. **Mute yourself** to avoid background noises that can disrupt the session, or to avoid embarrassing "hot mic" moments.

3. Identify the icon gesture to **"raise hand" digitally**. Don't assume you can unmute yourself to speak unless you have been given permission by the instructor verbally or in writing (in the syllabus).
4. **Speak only if prompted** or **appropriate**.
5. **Keep your focus on** the **camera and maintain eye contact on the screen** - this shows you are **attentive** and **engaged**.
6. **Limit facial expressions** that give away negative reactions.
7. ***Electronic Devices:***
   a. **You need to get into mood of thinking and studying, not into a mood of texting or checking your email.** So, b*efore class begins*, **turn off cell phones**. The **cell phone vibrating** or a **student texting** can be **very distracting to those around the student,** including **the faculty.** Please don't use **cell phones, e-mails, keyboards, browsers** etc. **during lectures** unless the instructor asks you. **Your desktop/laptop** is to be **used only** for the purpose of lab exercises, taking notes and your **tablet/phone** for reading the textbook. **No recording devices** are allowed.
   b. **Note:** Students **who use their mobile phones** during class lectures tend to write down less information, **recall less information**, and **perform worse** on a multiple-choice test than those students who abstain from using their mobile phones during class (p.251). **Reference**: Kuznekoff. J. H. and Titsworth, S. (2013). The impact of mobile phone usage on student learning. *Communication Education, 62*(3), 233-252.
8. **Classroom Behavior:**
   a. Everyone is expected to maintain a **courteous** and **respectful manner during lecture** or **student activities. Do not sleep, text, chat with your neighbors**, or work on assignments for other classes.
   b. **No clicking keyboard while lecturing**. **Please don't leave the class meeting during lectures.** All the above activities are **very disruptive** to others in class. Students who do not demonstrate appropriate classroom behavior will be asked to leave and receive an absence.
   c. **Patience** and **attention** to detail are important to succeed in Data Mining course.
9. **Requirements:**
   a. Every time students should **save your work in CMPS 386 directory on your desktop/laptop** and to your cloud drive (OneDrive or Google drive or Dropbox) or your email in the case **your desktop/laptop crashes.**

Patience and **attention** to detail are important to succeed in Apps Development.

Good luck in your course!

12. Tentative schedule (subject to change):

Date	Week No.	Topic	Reading Chapter
Feb 2	**1**	**Syllabus** review. Data Mining and Analytics Lab Exercises	ch1
Feb 9	**2**	Machine Learning Lab Exercises	ch2
Feb 16	**3**	Databases and Data Warehouses Lab Exercises	ch3
Feb 23	**4**	Data Visualization Lab Exercises	ch4
March 2	**5**	Keep Excel in Your Toolkit Lab Exercises	ch5

March 9	**6**	Keep SQL in Your Toolset Lab Exercises	ch6
March 16	**7**	Keep SQL in Your Toolset Lab Exercises	ch6
March 23	**8**	**Midterm**	chapters above
03/28 - 04/03		**Spring Break**	
April 6	**9**	NoSQL Data Analytics Lab Exercises	ch7
April 13	**10**	Programming Data Mining and Analytics Solutions Lab Exercises	ch8
April 20	**11**	Programming Data Mining and Analytics Solutions Lab Exercises	ch8
April 27	**12**	Data Preprocessing and Cleansing Lab Exercises	ch9
May 4	**13**	Data Clustering Lab Exercises	ch10
May 11	**14**	Classification Lab Exercises	ch11
May 18	**15**	Predictive Analytics and Data Association Lab Exercises	ch12, 13
May 25	**16**	**Final: Wednesday 6:50 pm**	above chapters

13. ☝PLAGIARISM POLICY:

Students are encouraged to collaborate, discuss and debate course concepts. It is all right to ask someone else about how to solve a problem, but **it is not all right to copy somebody's code or give a code**. Any cases of someone **turning in work that is not originally theirs** will be dealt with by **assigning zeros to both parties involved.** Each student is responsible for performing academic tasks in such a way that **honesty** is not in question**.**

There is a "zero tolerance" approach to academic dishonesty. Appropriate disciplinary action may include, but is not limited to **giving student an F** on the assignment/project/quiz/exam and/or in the course and/or recommending expulsion. The dean may place on probation, suspend, or expel any student who violates the academic honesty policy. (See ULV catalog for details).

14. ☝SOCIAL JUSTICE AT LA VERNE:

The Social Justice Incident Report Form is available to any University of La Verne community member wishing to report an incident of social injustice or discrimination (these may be acts that promote hate, fear, intimidation, unfair treatment, or oppression against an individual or a group). **Please note that reports can be submitted anonymously. Prior to submitting a social justice form, consider** if the reason is academic (classroom related) or something beyond that as all classroom related issues should be taken up with the Chair of the Department. The social justice incident/issue may be a non-emergency or emergency incident and can be **reported to an agency** (e.g. 911, La Verne Police Department, or University of La Verne Campus Safety Office). More information and the **online reporting forms** can be found on the web page of the Office of Diversity and Inclusivity or using the link below: https://cm.maxient.com/reportingform.php?UnivofLaVerne&layout_id=25.

15. ☝REMOTE COURSE PRIVACY:
    It is an invasion of privacy and a violation of the course policies for anyone to **record and/or distribute** another class participant's photographs, videos, screenshot saves, or any other method for capturing an image or audio, moving or still, with or without sound, without the participant's written consent. This policy does not apply to the University's or professor's recording of the synchronous portion of the course.

16. ☝ INCLUSION:
    The act of creating environments in which any individual or group can be and feel welcomed, respected, supported, and valued to fully participate and bring their full, authentic selves to work. An inclusive and welcoming climate embraces differences and offers respect in the words/actions/thoughts of all people.

Registration in this course **and acceptance** of this **syllabus** constitutes acknowledgement by **holder that student has read** and **agrees** to the p**rovisions** of the **foregoing** agreement between student and professor.