INFO-I-421
Applications of Data Mining
Indiana University School of Informatics and Computing, Indianapolis
Fall 2016

Section No.: 23500  Credit Hours: 3
Time: Tuesday and Thursday 4:30-5:45pm
Location: ET 015
First Class: August 8-23-2016
Instructor: Meeta Pradhan,
PhD in Chemical Engineering,
MS in Chem-Informatics
MS in Chemical Engineering
Assistant Research Professor
Office Hours: Tuesday and Thursday 2:00- 3:30pm, or by Appointment
Office: WK 306, Walker Plaza Building
719 Indiana Avenue, Indianapolis, IN 46202 [map]
Phone: (317) 278-0148 (Office)
Email: mpradhan@iupui.edu
Prerequisites: I123

OFFICIAL COURSE DESCRIPTION
Data is big, powerful, valuable and everywhere. As the amount of data grows the importance of gleaning useful information from data also increases. Data mining is the process of analyzing data from different perspectives and summarizing it into useful information. Through the engagement in this course, students get a deeper understanding of the effective uses of data, the hidden information and the knowledge in the data for scientific analysis, problem-solving and decision making.

EXTENDED COURSE DESCRIPTION
Data mining is a dynamic and fast growing field that incorporates the fields of statistics and computer science. Through the engagement in this course students will get the deeper understanding of the data mining concepts, algorithms, to solve the real world data-driven problems. Students will be introduced to data mining software’s R for analyzing different data. Students will develop a concrete understanding of the fundamentals and principles of data mining. On completion of this course a student will get hands on experience of analyzing a large amount of data and data types from different resources.
This is the brief, official course description approved by remonstrance.
**Course Outcomes**

Following are the two learning objectives of this course:

Learning objective 1: Data mining is an iterative processes and the two primary aims of data mining are prediction and description. In this objective the students will first learn the prediction techniques and then understand the use of different description methods to unravel the patterns in the data.

Learning objective 2: The students will work on one project in a group to get the hands-on experience of using different techniques. The project will help them to learn how to prepare, process, understand, analyze and present the data.

**Core Competencies**

The core competencies of this course include the following:

1. Core competency 1: Analyze the problem; Implement different techniques, Evaluate and Refine them to analyze the problem;
2. Core competency 2: Work Collaboratively; Discuss and Disseminate

**COURSE ORGANIZATION AND STRUCTURE**

The course is organized into three primary Themes: (A) Understanding the data (B) Extracting the Knowledge from the data ( C) Interpreting the knowledge to solve the real world problems.

Key themes will be addressed each week and each theme is accompanied by its own set of specific goals and outcomes.
Required Text(s):
Title: Introduction to Data Mining
Author(s): Tan Pang-Ning, Steinbach Michael, Kumar Vipin
Edition: 1st Edition
Publisher: Pearson Education

Reference Book
Practical Data Science with R
Nina Zumel and John Mount

Additional Readings:
Other reading material will be provided at the end of each class.
By Ian H Witten, Eibe Frank and Mark A Hall.

SOFTWARE, TECHNICAL REQUIREMENTS, AND SUPPORT
Students are responsible for making all necessary provisions for accessing course-related resources, including those that are available in Oncourse <http://oncourse.iu.edu>

In order to access licensed information resources from off-campus computers, students may be required to Install IU’s VPN client onto your home computer. This allows your
computer to access the IUPUI pool of resources by adopting an IUPUI IP address so that IP-restricted resources recognize your computer as an IU-affiliated machine.

Excellent technical assistance is available from the following:

- **KnowledgeBase Online Q & A:** [http://kb.iu.edu](http://kb.iu.edu)
- **ITHelp Live:** [https://ithelplive.iu.edu/](https://ithelplive.iu.edu/)
- **Phone support:** 317-274-4357, 24 hours a day, 7 days a week
- **Email:** [ithelp@iu.edu](mailto:ithelp@iu.edu)

The R project for Statistical Computing - Data Mining with open sources machine learning are freely available. Please become familiar with the tools and resources available to you via **IUanyWhere**: [https://iuanyware.iu.edu/vpn/index.html](https://iuanyware.iu.edu/vpn/index.html)

**IUanyWare** is a client virtualization (CV) service available to Indiana University students, faculty, and staff. With IUanyWare, you can use a web browser or mobile app to run certain IU-licensed software applications without having to install them on your computer or mobile device.

You'll need to do some initial setup, and configure cloud storage if you wish to store files remotely (e.g., on your IU Box or SharePoint My Site account); see:

For information on software currently available through IUanyWhere and in the IUB and IUPUI STCs, see the current software list at: [https://stcweb.stc.indiana.edu/Public/Software/current.cfm](https://stcweb.stc.indiana.edu/Public/Software/current.cfm)

**Please note that IUanyWhere resets your account at the end of each school year.** If you had an account last spring 2014, it will have been reset for this fall. Details at: [https://kb.iu.edu/d/bdlb](https://kb.iu.edu/d/bdlb)

**Required Software**

- **R** ([http://www.r-project.org](http://www.r-project.org)), a language and environment for statistical computing and graphics. R supports all the data analytics methods covered in the course. R is flexible, fully programmable, and preferred by research statisticians, which means that new algorithms are often implemented in R first. Unlike most other powerful statistical packages, R is free of cost.

**Optional Software**

- **Rcommander** ([http://www.rcommander.com](http://www.rcommander.com)), a graphical user interface (GUI) for statistics using R. Rcommander is free of cost and operates on Windows, IOS (Mac), and Linux operating systems.
• Rattle ([http://rattle.togaware.com](http://rattle.togaware.com) or [http://cran.r-project.org/web/packages/rattle/](http://cran.r-project.org/web/packages/rattle/)), a GUI for data mining using R. Rattle is free of cost and operates on Windows, IOS (Mac), and Linux operating systems.

• RStudio ([http://www.rstudio.org](http://www.rstudio.org)), an integrated development environment (IDE) for R. RStudio is free of cost and operates on Windows, IOS (Mac), and Linux operating systems.

PRINCIPLES OF UNDERGRADUATE LEARNING (PUL)

Learning outcomes are assessed in the following areas:

1. Knowledge and skills mastery
   - Major emphasis
2. Critical thinking and good judgment
   - Moderate emphasis
3. Effective communication
   - Some emphasis
4. Ethical behavior

Learning Outcomes:

Upon completion of this course, students will

<table>
<thead>
<tr>
<th>PGPL</th>
<th>Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EQMFP</td>
</tr>
<tr>
<td>1. Understanding the different data and their types</td>
<td>1, 2</td>
</tr>
<tr>
<td>2. Different types of distribution to understand a data</td>
<td>1, 2</td>
</tr>
<tr>
<td>Different types of tests for data analysis</td>
<td>EQMFP</td>
</tr>
<tr>
<td>3. Supervised and unsupervised learning</td>
<td>1, 2</td>
</tr>
<tr>
<td>4. Project Designing, Writing, Analysis and Presentation</td>
<td>1, 2, 3, 4</td>
</tr>
<tr>
<td>5. Write programs to perform data analytics on large, complex datasets in R</td>
<td>1, 2</td>
</tr>
</tbody>
</table>
EXPECTATIONS, GUIDELINES AND POLICIES

Attendance

• Class attendance is required for classroom-based courses.
• Attendance shall be taken in every class. If you do not sign the attendance sheet while in class, you shall be marked absent.
• Signing the attendance sheet for another student is prohibited.
• The instructor is required to submit to the Registrar a record of student attendance, and action shall be taken if the record conveys a trend of absenteeism.
• Absences must be explained to the satisfaction of the instructor, who will decide whether omitted work may be made up.
• Missing class reduces your grade through the following grade reduction policy:
  • You are allowed ONE excused or unexcused absences.
  • Regardless of the reason, a 2nd absence results in a 25% reduction in your final grade
  • A 3rd absence results in a 50% reduction.
  • Further absences result in an F in the course.
  • Missing class may also reduce your grade by eliminating opportunities for class participation.

Class Preparation

• You are expected to read the chapters and the material given in the class
• Research shows that regular attendance, preparation and active class participation have a positive impact on your final grade for a course.
• Ask whatever questions you have pertaining to the course, while we are face to face.
• When not in class, ask on the class forum and ask your questions and receive answers. In this way, the entire class can benefit from your question. There are no silly questions!!!!

Late Assignments and Submission Of Assignments

• All work (unless otherwise noted) should be submitted via an attachment in the Assigned area.
• Home work will be due by 11:55 PM of the specified day. If your Home Work is late, your respective assignment will be assessed a 25% late penalty. Any assignment that is not turned in by 24 hours after the due date
will not be accepted and you will receive a zero (0) for that particular assignment. Also, if I give out a solution and you have not submitted your work, you will not be able to turn in the late work and you will also receive a zero (0) for that particular assignment.

In class Assignment
- In class assignments need to be completed in the class

Grading Information
- Attendance: 5%
- Home Work: 15%
- In class assignment: 20%
- Quiz: 20%
- MidTerm: 20%
- Project Presentation and Report: 20%

(All assignments/homework must be done using R programming)

Grading Scale
- A+: 100% - 97%
- A: 96.99% - 94%
- A-: 93.99% - 90%
- B+: 89.99% - 87%
- B: 86.99% - 83%
- B-: 82.99% - 80%
- C+: 79.99% - 77%
- C: 76.99% - 73%
- C-: 72.99% - 70%
- D+: 69.99% - 67%
- D: 66.99% - 63%
- D-: 62.99% - 60%
- F: 59.99% - 57%

WEEKLY SCHEDULE
- Every Day Class Schedule
  - 4:30 pm to 5:30 pm - Tuesday’s Lecture
4:30 pm to 5:45 pm  -  In class assignment on Thursday

**Date for each class meeting**

- Specific pre-class readings
- Specific subject matter/topics covered
- Goals and objectives of each class period

<table>
<thead>
<tr>
<th>Week #</th>
<th>Day- Date</th>
<th>Lecture #</th>
<th>Topic</th>
<th>Assignment IN</th>
<th>Assignment DUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tues. Aug-23</td>
<td>1</td>
<td>Class Outline</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Introduction to Data and Big Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thurs. Aug-25</td>
<td>2</td>
<td>Introduction to Data Types</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Tues. Aug-30</td>
<td>3</td>
<td>Sample comparison, population, Introduction to supervised and unsupervised learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thurs. Sept-1</td>
<td>4</td>
<td>Introduction to Data Mining Tools - R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Tues. Sept-6</td>
<td>5</td>
<td>Introduction to Data Mining Tools - R and WEKA</td>
<td>Home Work I</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thurs. Sept-8</td>
<td>6</td>
<td>In class Assignment I</td>
<td>Project Report I</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Tues. Sept-13</td>
<td>7</td>
<td>Introduction to R</td>
<td>Home Work I</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thurs. Sept-15</td>
<td>8</td>
<td>Introduction to R</td>
<td></td>
<td>Project Report I</td>
</tr>
<tr>
<td>5</td>
<td>Tues. Sept-20</td>
<td>9</td>
<td>Introduction to Decision Trees</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Project Data Resources</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thurs. Sept-22</td>
<td>10</td>
<td>Introduction to Decision Trees</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Tues. Sept-27</td>
<td>11</td>
<td>Introduction to Decision Trees</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>R- Decision Tree/ WEKA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thurs. Sept-29</td>
<td>12</td>
<td>In class Assignment II A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Tues. Oct-4</td>
<td>13</td>
<td>Quiz : Lecture 1, Lecture 2, Lecture 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>Day</td>
<td>Time</td>
<td>Activity</td>
<td>Group Meetings</td>
<td>Project Report</td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>-------</td>
<td>-----------------------------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Thurs. Oct:6</td>
<td>14</td>
<td>14</td>
<td>In class Assignment II B_ WEKA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Tues. Oct:11</td>
<td>15</td>
<td>Rule Based Classifier</td>
<td></td>
<td>HomeWork II</td>
</tr>
<tr>
<td></td>
<td>Thurs. Oct:13</td>
<td>16</td>
<td>Rule Base Classifier / Revision</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Tues. Oct:18</td>
<td>17</td>
<td>FALL BREAK – NO CLASS</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thurs. Oct:20</td>
<td>18</td>
<td>Mid Term</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Tues. Oct:25</td>
<td>19</td>
<td>Clustering Analysis</td>
<td>Schedule Group Meetings with Dr. Pradhan</td>
<td>HomeWork II</td>
</tr>
<tr>
<td></td>
<td>Thurs. Oct:27</td>
<td>20</td>
<td>Clustering Analysis</td>
<td></td>
<td>Project Report II</td>
</tr>
<tr>
<td>11</td>
<td>Tues. Nov:2</td>
<td>21</td>
<td>Cluster Analysis_ In class Assignment IIIA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thurs. Nov:3</td>
<td>22</td>
<td>Cluster Analysis _ In class Assignment III B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Tues. Nov:8</td>
<td>23</td>
<td>Project Presentation</td>
<td></td>
<td>Home Work III</td>
</tr>
<tr>
<td>13</td>
<td>Tues. Nov:15</td>
<td>25</td>
<td>Linear Regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thurs. Nov:17</td>
<td>26</td>
<td>Linear Regression Analysis</td>
<td></td>
<td>Home Work III</td>
</tr>
<tr>
<td>14</td>
<td>Tues. Nov:22</td>
<td>27</td>
<td>In class Assignment IV</td>
<td></td>
<td>Project Report III</td>
</tr>
<tr>
<td></td>
<td>Thurs. Nov:24</td>
<td>28</td>
<td>THANKS GIVING BREAK-NO CLASS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Tues. Nov:29</td>
<td>29</td>
<td>Quiz</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thurs. Dec:1</td>
<td>30</td>
<td>Project Discussion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Tues. Dec:6</td>
<td>31</td>
<td>Project Presentation</td>
<td></td>
<td>Complete Project Report due</td>
</tr>
<tr>
<td></td>
<td>Thurs. Dec:8</td>
<td>32</td>
<td>Project Presentation</td>
<td></td>
<td>Complete Project Report due</td>
</tr>
</tbody>
</table>

**Grading Scale:**

- **A+** 97–100% Professional level work, showing highest level of achievement
- **A** 93–96% Extra ordinarily high achievement, quality of work; shows command of the subject matter
- **A–** 90–92% Excellent and thorough knowledge of the subject matter
B+  87–89%  Above average understanding of material and quality of work
B    83–86%  Mastery and fulfillment of all course requirements;
            good, acceptable work
B–  80–82%  Satisfactory quality of work
C+  77–79%  Minimally acceptable performance and quality of work
C    73–76%  Unacceptable work, does not demonstrate mastery
C–  70–72%  Unacceptable work
D+  67–69%  Unacceptable work
D    63–66%  Unacceptable work
D–  60–62%  Unacceptable work
F    Below 60 Failure

MISSION STATEMENT

The Mission of IUPUI is to provide for its constituents excellence in

- Teaching and Learning;
- Research, Scholarship, and Creative Activity; and
- Civic Engagement.

With each of these core activities characterized by

- Collaboration within and across disciplines and with the community;
- A commitment to ensuring diversity; and
- Pursuit of best practices.

IUPUI’s mission is derived from and aligned with the principal components—Communities of Learning, Responsibilities of Excellence, Accountability and Best Practices—of Indiana University’s Strategic Directions Charter.

STATEMENT OF VALUES

IUPUI values the commitment of students to learning; of faculty to the highest standards of teaching, scholarship, and service; and of staff to the highest standards of service. IUPUI recognizes students as partners in learning. IUPUI values the opportunities afforded by its location in Indiana’s capital city and is committed to serving the needs of its community. Thus, IUPUI students, faculty, and staff are involved in the community, both to provide educational programs and patient care and to apply learning to community needs through service. As a leader in fostering collaborative relationships, IUPUI values collegiality, cooperation, creativity, innovation, and entrepreneurship as well as honesty, integrity, and support for open inquiry and dissemination of findings. IUPUI is committed to the personal and professional development of its students, faculty, and staff and to continuous improvement of its programs and services.

CODE OF CONDUCT

10
All students should aspire to the highest standards of academic integrity. Using another student’s work on an assignment, cheating on a test, not quoting or citing references correctly, or any other form of dishonesty or plagiarism shall result in a grade of zero on the item and possibly an F in the course. Incidences of academic misconduct shall be referred to the Department Chair and repeated violations shall result in dismissal from the program.

All students are responsible for reading, understanding, and applying the Code of Student Rights, Responsibilities and Conduct and in particular the section on academic misconduct. Refer to The Code > Responsibilities > Academic Misconduct at http://www.indiana.edu/~code/. All students must also successfully complete the Indiana University Department of Education “How to Recognize Plagiarism” Tutorial and Test. https://www.indiana.edu/~istd You must document the difference between your writing and that of others. Use quotation marks in addition to a citation, page number, and reference whenever writing someone else’s words (e.g., following the Publication Manual of the American Psychological Association). To detect plagiarism instructors apply a range of methods, including Turnitin.com. http://www.ulib.iupui.edu/libinfo/turnitin

ACADEMIC MISCONDUCT:

1. **Cheating:** Cheating is considered to be an attempt to use or provide unauthorized assistance, materials, information, or study aids in any form and in any academic exercise or environment.
   a. A student must not use external assistance on any “in-class” or “take-home” examination, unless the instructor specifically has authorized external assistance. This prohibition includes, but is not limited to, the use of tutors, books, notes, calculators, computers, and wireless communication devices.
   b. A student must not use another person as a substitute in the taking of an examination or quiz, nor allow other persons to conduct research or to prepare work, without advanced authorization from the instructor to whom the work is being submitted.
   c. A student must not use materials from a commercial term paper company, files of papers prepared by other persons, or submit documents found on the Internet.
   d. A student must not collaborate with other persons on a particular project and submit a copy of a written report that is represented explicitly or implicitly as the student’s individual work.
   e. A student must not use any unauthorized assistance in a laboratory, at a computer terminal, or on fieldwork.
   f. A student must not steal examinations or other course materials, including but not limited to, physical copies and photographic or electronic images.
   g. A student must not submit substantial portions of the same academic work for credit or honors more than once without permission of the instructor or program to whom the work is being submitted.
h. A student must not, without authorization, alter a grade or score in any way, nor alter answers on a returned exam or assignment for credit.

2. **Fabrication:** A student must not falsify or invent any information or data in an academic exercise including, but not limited to, records or reports, laboratory results, and citation to the sources of information.

3. **Plagiarism:** Plagiarism is defined as presenting someone else’s work, including the work of other students, as one’s own. Any ideas or materials taken from another source for either written or oral use must be fully acknowledged, unless the information is common knowledge. What is considered “common knowledge” may differ from course to course.
   a. A student must not adopt or reproduce ideas, opinions, theories, formulas, graphics, or pictures of another person without acknowledgment.
   b. A student must give credit to the originality of others and acknowledge indebtedness whenever:
      1. directly quoting another person’s actual words, whether oral or written;
      2. using another person’s ideas, opinions, or theories;
      3. paraphrasing the words, ideas, opinions, or theories of others, whether oral or written;
      4. borrowing facts, statistics, or illustrative material; or
      5. offering materials assembled or collected by others in the form of projects or collections without acknowledgment
         [http://life.iupui.edu/dos/code.htm](http://life.iupui.edu/dos/code.htm)
         [https://www.indiana.edu/~istd/test.html](https://www.indiana.edu/~istd/test.html)

4. **Interference:** A student must not steal, change, destroy, or impede another student’s work, nor should the student unjustly attempt, through a bribe, a promise of favors or threats, to affect any student’s grade or the evaluation of academic performance. Impeding another student’s work includes, but is not limited to, the theft, defacement, or mutilation of resources so as to deprive others of the information they contain.

5. **Violation of Course Rules:** A student must not violate course rules established by a department, the course syllabus, verbal or written instructions, or the course materials that are rationally related to the content of the course or to the enhancement of the learning process in the course.

6. **Facilitating Academic Dishonesty:** A student must not intentionally or knowingly help or attempt to help another student to commit an act of academic misconduct, nor allow another student to use his or her work or resources to commit an act of misconduct.

**OTHER POLICIES**

1. **IUPUI course policies:** A number of campus policies governing IUPUI courses may be
found at the following link:  http://registrar.iupui.edu/course_policies.html

2. **Classroom civility**: To maintain an effective and inclusive learning environment, it is important to be an attentive and respectful participant in lectures, discussions, group work, and other classroom exercises. Thus, unnecessary disruptions should be avoided, such as ringing cell phones engagement in private conversations and other unrelated activities. Texting, surfing the Internet, and posting to Facebook or Twitter during class are generally not permitted. IUPUI nurtures and promotes “a campus climate that seeks, values, and cultivates diversity in all of its forms and that provides conditions necessary for all campus community members to feel welcomed, supported, included, and valued” (IUPUI Strategic Initiative 9). IUPUI prohibits “discrimination against anyone for reasons of race, color, religion, national origin, sex, sexual orientation, marital status, age, disability, or [veteran] status” (Office of Equal Opportunity). Profanity or derogatory comments about the instructor, fellow students, invited speakers or other classroom visitors, or any members of the campus community shall not be tolerated. A violation of this rule shall result in a warning and, if the offense continues, possible disciplinary action.

3. **Disabilities Policy**: In compliance with the Americans with Disabilities Act (ADA), all qualified students enrolled in this course are entitled to reasonable accommodations. Please notify the instructor during the first week of class of accommodations needed for the course. Students requiring accommodations because of a disability must register with Adaptive Educational Services (AES) and complete the appropriate AES-issued before receiving accommodations. The AES office is located at UC 100, Taylor Hall (Email: aes@iupui.edu, Tel. 317 274-3241). Visit http://aes.iupui.edu for more information.

4. **Administrative Withdrawal**: A basic requirement of this course is that students participate in all class discussions and conscientiously complete all required course activities and/or assignments. If a student is unable to attend, participate in, or complete an assignment on time, it is the student’s responsibility to inform the instructor. If a student misses more than half of the required activities within the first 25% of the course without contacting the instructor, the student may be administratively withdrawn from this course. Administrative withdrawal may have academic, financial, and financial aid implications. Administrative withdrawal will take place after the full refund period, and a student who has been administratively withdrawn from a course is ineligible for a tuition refund. Contact the instructor with questions concerning administrative withdrawal.