INFO 1223
Data Fluency

Department of Human Centered Computing, Informatics Program
Indiana University School of Informatics and Computing, Indianapolis
Fall 2017

Section No.: Credit Hours: 3 credit hours
First Class: August 2017
Instructor: Ran Chang, Ph.D.
Office Hours: By Appointment (via Adobe Connect)
Office: 535 West Michigan St., Indianapolis, IN 46202 [map]
Phone: Email:

Prerequisites: None (Not an extension of any undergraduate or graduate course)

COURSE DESCRIPTION

Pervasive, vast, and growing describe data in today’s environment. This course introduces fundamental skills for extracting from data actionable knowledge. Students create, access, munge, analyze, and visualize data to draw inferences and make predictions. The course uses real datasets from a variety of disciplines including healthcare, business, and the humanities.

Note: All aspects of data storage, munging, analysis, and visualization are implemented in Python or use Python libraries.

Required Texts:
https://www.learnpython.org

Required Software: (open source, no cost)
 Anaconda – a Python distribution for data analytics and scientific computing, which includes the Jupyter Notebook, NumPy, SciPy, Matplotlib, and Pandas.

Optional Software:
R – a programming language and software environment for data analysis and visualization
RStudio – an integrated development environment (IDE) for R
Rcommander – a graphical user interface (GUI) for commonly used R statistical functions

**STUDENT LEARNING OUTCOMES**

Upon completion of this course, students will

<table>
<thead>
<tr>
<th>RBT</th>
<th>PUL</th>
<th>SWC</th>
<th>PLO</th>
<th>Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. store, structure, and access data of different types using simple relational models and tables</td>
<td>3</td>
<td>1B, 3</td>
<td>3.1</td>
<td>1</td>
</tr>
<tr>
<td>2. munge data to prepare raw data for further analysis</td>
<td>3</td>
<td>1B, 2</td>
<td>3.3</td>
<td>1</td>
</tr>
<tr>
<td>3. analyze large, complex datasets with supervised learning methods, including linear regression and k-nearest neighbors for functional approximation and naïve Bayes and decision trees for classification and predictive modeling</td>
<td>4</td>
<td>1B</td>
<td>3.4–3.8</td>
<td>1</td>
</tr>
<tr>
<td>4. analyze large, complex datasets with unsupervised learning methods, including k-means clustering</td>
<td>4</td>
<td>1B</td>
<td>3.4–3.8</td>
<td>1</td>
</tr>
<tr>
<td>5. calculate probabilities by applying additive and multiplicative laws, permutations and combinations, and conditional probability</td>
<td>3</td>
<td>1B</td>
<td>3.3</td>
<td>1</td>
</tr>
<tr>
<td>6. calculate expectation and variance from the probability distribution of a random variable</td>
<td>3</td>
<td>1B</td>
<td>3.3</td>
<td>1</td>
</tr>
<tr>
<td>7. assess model fit (e.g., overfitting or underfitting)</td>
<td>5</td>
<td>1B</td>
<td>3.6, 3.7</td>
<td>1</td>
</tr>
<tr>
<td>8. create visualizations of data to communicate and persuade</td>
<td>6</td>
<td>1B, 2</td>
<td>3.2, 3.8</td>
<td>1, 4</td>
</tr>
<tr>
<td>9. derive information from data and support conclusions or recommendations based on evidence existing in the data</td>
<td>5</td>
<td>1B, 2</td>
<td>3.4, 3.6, 3.8</td>
<td>1, 4</td>
</tr>
</tbody>
</table>

**Revised (by Anderson) Bloom’s Taxonomy (RBT)**


**Principles of Undergraduate Learning (PUL)**
Learning outcomes are assessed in the following areas:

1A. Core communication: written, oral and visual skills

1B. Core communication: quantitative skills

1C. Core communication: information resources skills

2. Critical thinking

3. Integration and application of knowledge

4. Intellectual depth, breadth, and adaptiveness

5. Understanding society and culture

6. Values and ethics

Statewide Competencies (STC)

3. Quantitative Reasoning

Upon completion of the General Education Transfer Core, students will

3.1. Interpret information that has been presented in mathematical form (e.g. with functions, equations, graphs, diagrams, tables, words, geometric figures).

3.2. Represent information/data in mathematical form as appropriate (e.g. with functions, equations, graphs, diagrams, tables, words, geometric figures).

3.3. Demonstrate skill in carrying out mathematical (e.g. algebraic, geometric, logical, statistical) procedures flexibly, accurately, and efficiently to solve problems.

3.4. Analyze mathematical arguments, determining whether stated conclusions can be inferred.

3.5. Communicate which assumptions have been made in the solution process.

3.6. Analyze mathematical results to determine the reasonableness of the solution.

3.7. Cite the limitations of the process where applicable.

3.8. Clearly explain the representation, solution, and interpretation of the math problem.

Program-level Learning Outcomes (PLO)

Graduates of the Informatics undergraduate program will demonstrate expertise in the following core competencies essential to success as an informatics, computing and information technology professional:

1. Technical Knowledge:
   a. Demonstrate knowledge and skills in the mathematical and logical foundations of informatics, data representation, models, structures and informatics-centric management
   b. Define terms and explain basic principles essential to the design and development of IT and computing systems
   c. Acquire fundamental concepts and skills in software architectures and the development of information systems

2. Social Dynamics of Informatics and Information Technology:
   a. Understand and apply major societal trends affecting the development and deployment of modern day IT, such as access, privacy, intellectual property, security and others
b. Critically analyze the impact of IT on individuals, groups and organizations at local and global levels

c. Apply a user-centered approach to interaction design and product usability, including techniques for quantitative and qualitative testing of interface and interaction design

d. Utilize digital tools to communicate with a range of audiences

e. Analyze the social, cultural and organizational settings in which IT solutions will be deployed to achieve successful implementation

3. Domain-specific Critical Thinking and Problem Solving Skills:
   a. Define terms and explain basic principles, concepts and theories from another domain or discipline in which IT skills will be applied
   b. Deploy IT resources in the context of another domain and/or discipline
   c. Synthesize, analyze and conceptualize information and ideas from multiple sources and perspectives
   d. Evaluate data, arrive at reasoned conclusions and solve challenging problems

4. Collaborative Teamwork:
   a. Select and effectively utilize oral, written, visual and quantitative communication skills within the context of an interdisciplinary team
   b. Identify and demonstrate the skills, behaviors and attitudes necessary to function as an effective team member, including working cooperatively with diverse group members
   c. Acquire the skills to initiate, manage and execute an IT project
   d. Articulate legal and ethical issues when using the creative work of others; respect the intellectual property of others

5. Professional Ethics and Development:
   a. Create a personal code of ethics; articulate principles for resolving ethical conflicts
   b. Commit to a regular program of continuing education and lifelong learning that is independent of employer sponsorship
   c. Participate in professional organizations that promote responsible computing and service to society

Teaching and Learning Methods (TLM)

1. Active learning: Prelecture student discovery on topic
2. Lecture by instructor: Use of slides and audiovisual aids
3. Self-directed learning: Assignments on the textbook and lectures
4. Lab-based learning: Gain practical experience of concepts covered in lectures
5. Peer-assisted learning: Project on data analytics

Topics Covered

1. What is Data Science?
   ◦ Big Data and Data Science
   ◦ What can data analysis discover? A case study with adverse drug events
The current landscape
Skills of data scientists – learning the application domain, communicating with uses, seeing the picture of a complex system, know how data can be represented, transformed, visualized, analyzed, and reasoned about ethically.

2 What is Data?
- Data definition – bits to bytes to larger structures
- Data collection
- Data storage – Writing a Python program to convert text to rows and columns in a .csv file to a relational database using sqlite3
- Accessing and using data in the relational database – a project with rows and columns in a single table (Case Study and Homework)
- Data visualization of the table in iPython’s Jupyter Notebook

3 Exploratory Data Analysis in Python using the Jupyter Notebook
- Statistical inference
- Populations and samples
- Analysis of histograms
- Introduction to probability as fractional counts
- Exploratory data analysis
- Case Study: RealDirect – How do they make money?

4 Mining the Data
- Case study and application – spam filter
- Introduction to learning - What can be learned?
- Algorithm – linear regression
- Algorithm – decision tree
- Algorithm – k-nearest neighbor

5 Big data
- Populations and samples of big data
- Big data = big assumptions?
- What are the technologies? Parallel computations – an example
- Case study - Watson project

6 Successful Applications
- Financial modeling
- Recommendation engines
- Fraud detection
- Social networks
- Medical applications

7 Into the future
- Crowdsourcing
- What are next generation data scientists – problem solvers, soft skills, asking the right questions
- Being an ethical data scientist
# COURSE SCHEDULE

<table>
<thead>
<tr>
<th>Week</th>
<th>Topic</th>
<th>Assessment</th>
</tr>
</thead>
</table>
| 1A   | Introduction to Data Fluency  
What is data? Definition and collection |  |
| 1B   | Installing Anaconda, Python programming 1 |  |
| 2A   | Python programming 2 |  |
| 2B   | Python programming 3 | Homework 1 |
| 3A   | Entity–relationship models and functions | Quiz 1 |
| 3B   | Storing, accessing, and using data with sqlite3 database |  |
| 4A   | Sampling | Quiz 2, Homework 2 |
| 4B   | Data visualization with Matplotlib |  |
| 5A   | Probability theory 1 | Homework 3 |
| 5B   | Probability theory 2 |  |
| 6A   | Machine learning 1 |  |
| 6B   | Machine learning 2 |  |
| 7A   | Pandas for data analysis | Homework 4 |
| 7B   | Naïve Bayes classifier 1 |  |
| 8A   | Midterm Review |  |
| 8B   | | Midterm 1 |
| 9A   | Naïve Bayes 2  
Decision trees 1 |  |
| 9B   | Decision trees 2 | Homework 5 |
| 10A  | Decision trees 3, Evaluation 1 |  |
| 10B  | Evaluation 2 (and overfitting) |  |
| 11A  | Supervised Learning review  
Midterm review | Homework 6 |
| 11B  | | Midterm 2 |
| 12A  | $k$-nearest neighbors |  |
| 12B  | Clustering |  |
| 13A  | Applications in Python |  |
| 13B  | Machine Learning review | Quiz 3 |
| 14A  | Big data | Project |
| 14B  | In-class student presentations |  |
| 15A  | In-class student presentations | Homework 7 |
| 15B  | Wrap up and review | Final Exam |

## Assessment

- **Midterm 1**: 20%
- **Midterm 2**: 20%
- **Final Exam**: 20%
- **Quizzes**: 5%
- **Mini-project**: 10%
Homework 25%

Grading Scale:

<table>
<thead>
<tr>
<th>Grade</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A+</td>
<td>100%</td>
</tr>
<tr>
<td>A</td>
<td>93–99%</td>
</tr>
<tr>
<td>A−</td>
<td>90–92%</td>
</tr>
<tr>
<td>B+</td>
<td>87–89%</td>
</tr>
<tr>
<td>B</td>
<td>83–86%</td>
</tr>
<tr>
<td>B−</td>
<td>80–82%</td>
</tr>
<tr>
<td>C+</td>
<td>77–79%</td>
</tr>
<tr>
<td>C</td>
<td>73–76%</td>
</tr>
<tr>
<td>C−</td>
<td>70–72%</td>
</tr>
<tr>
<td>D+</td>
<td>67–69%</td>
</tr>
<tr>
<td>D</td>
<td>63–66%</td>
</tr>
<tr>
<td>D−</td>
<td>60–62%</td>
</tr>
<tr>
<td>F</td>
<td>Below 60</td>
</tr>
</tbody>
</table>

http://registrar.iupui.edu/gradecover.html

EXPECTATIONS, GUIDELINES, AND POLICIES

Attendance:

A basic requirement of this course is that you will participate in all class meetings, whether online or face-to-face, and conscientiously complete all required course activities and assignments. Class attendance is required for classroom-based courses. It entails being present and attentive for the entire class period. 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.

Only the following are acceptable excuses for absences: death in the immediate family (e.g. mother, father, spouse, child, or sibling), hospitalization or serious illness; jury duty; court ordered summons; religious holiday; university/school coordinated athletic or scholastic activities; an unanticipated event that would cause attendance to result in substantial hardship to one’s self or immediate family. Absences must be explained with the submission of appropriate documentation to the satisfaction of the instructor, who will decide whether missed work may be made up. Absences that do not satisfy the above criteria are considered unexcused. To protect your privacy, doctor’s excuses should exclude the nature of the condition and focus instead on how the condition impacts your attendance and academic performance.

Missing class reduces your grade through the following grade reduction policy: You are allowed two excused or unexcused absences. Each additional absence, unless excused, results in a 5% reduction in your final course grade. More than six absences result in an F in the course. Missing class may also reduce your grade by eliminating opportunities for class participation. For all absences, the student is responsible for all covered materials and assignments.

Incomplete:

The instructor may assign an Incomplete (I) grade only if at least 75% of the required coursework has been completed at passing quality and holding you to previously established
time limits would result in unjust hardship to you. All unfinished work must be completed by the date set by the instructor. Left unchanged, an Incomplete automatically becomes an F after one year. http://registrar.iupui.edu/incomp.html

**Deliverables:**

You are responsible for completing each deliverable (e.g., assignment, quiz) by its deadline and submitting it by the specified method. Deadlines are outlined in the syllabus or in supplementary documents accessible through OnCourse. Should you miss a class, you are still responsible for completing the deliverable and for finding out what was covered in class, including any new or modified deliverable. In fairness to the instructor and students who completed their work on time, a grade on a deliverable shall be reduced 10%, if it is submitted late and a further 10% for each 24-hour period it is submitted after the deadline.

**CODE OF CONDUCT**

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/turmitin

**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

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. **Administrative withdrawal**: Students must 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, the student must 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 occurs after the full refund period, and a student who has been administratively withdrawn is ineligible for a tuition refund.

2. **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. Cell phones, media players, or any noisy devices should be turned off during a class. Texting, web surfing, and posting to social media are generally not permitted. Laptop use may be permitted if it is used for taking notes or conducting class activities. Students should check with the instructor about permissible devices in class. 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. **Communication**: For classroom-based courses, the instructor or teaching assistant should respond to emails by the end of the next class or, for online courses, within two Indiana University working days, which excludes weekends and holidays. The instructor should provide weekly office hours or accept appointments for face-to-face, telephone, or teleconferenced meetings, and announce periods of extended absence in advance.

4. **Counseling and Psychological Services (CAPS)**: Students seeking counseling or other psychological services should contact the CAPS office at 274-2548 or capsindy@iupui.edu. For more information visit http://life.iupui.edu/caps/.

5. **Course evaluations**: Course evaluations provide vital information for improving the quality of courses and programs. Students are urged to complete one course and instructor evaluation for each section in which they are enrolled at the School of Informatics and Computing with the following exceptions: (a) The student has
withdrawn from the course; (b) fewer than five students are enrolled in the section (in which case maintaining anonymity is difficult); and (c) the section is a laboratory that must be taken with a course having a different section number. Course evaluations are completed at https://soic.iupui.edu/app/course-eval/. Course evaluations are typically open from the eleventh week. Course evaluations are anonymous, which means that no one can view the name of the student completing the evaluation. In addition, no one can view the evaluation itself until after the instructor has submitted the final grades. In small sections, demographic information should be left blank, if it could be used to identify the student.

6. **Disabilities policy:** All qualified students enrolled in this course are entitled to reasonable accommodations for a disability. Notify the instructor during the first week of class of accommodations needed. Students requiring accommodations 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). For more information visit [http://aes.iupui.edu](http://aes.iupui.edu).

7. **Email:** Indiana University uses your IU email account as an official means of communication, and students should check it daily. Although you may have your IU email forwarded to an outside email account, please email faculty and staff from your IU email account.

8. **Emergency preparedness:** Know what to do in an emergency so that you can protect yourself and others. For more information, visit the emergency management website at [http://protect.iu.edu/emergency](http://protect.iu.edu/emergency).

9. **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](http://registrar.iupui.edu/course_policies.html)

10. **No class attendance without enrollment.** Only those who are officially enrolled in this course may attend class unless enrolled as an auditor or making up an Incomplete by prior arrangement with the instructor. This policy does not apply to those assisting a student with a documented disability, serving in an instructional role, or administrative personnel. [http://registrar.iupui.edu/official-enrollment-class-attendance.html](http://registrar.iupui.edu/official-enrollment-class-attendance.html) Children may not attend class with their parents, guardians, or childcare providers.

11. **Religious holidays:** Students seeking accommodation for religious observances must submit a request form to the course instructor by the end of the second week of the semester. For information visit [http://registrar.iupui.edu/religious.html](http://registrar.iupui.edu/religious.html).

12. **Right to revise:** The instructor reserves the right to make changes to this syllabus as necessary and, in such an event, will notify students of the changes immediately.

13. **Sexual misconduct:** IU does not tolerate sexual harassment or violence. For more information and resources, visit [http://stopsexualviolence.iu.edu/](http://stopsexualviolence.iu.edu/).

14. **Student advocate:** The Student Advocate assists students with personal, financial, and academic issues. The Student Advocate is in the Campus Center, Suite 350, and may also be contacted at 317 274-4431 or studvoc@iupui.edu. For more information visit
http://studentaffairs.iupui.edu/advocate.

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.