DS-I590 – Data Science in Practice Also Known As "Consulting Club"

Spring 2018 Data Science Practice

Proposed Capacity and Schedule outline

> 60 Student Maximum (Semester one) spring 2018
>
> Fall 2018 (Semester two) 120 students Maximum
>
> Divided into 3 cohorts of 20 Students
>
> Class Cohorts will meet once a week for three hours.
>
> Student Project teams will meet up to twice a week.

Prerequisites -- There are no specific prerequisites for this course. It will help if you are ready to learn a collaborative approach to conducting data science in real world situations and working on the process and human side of data science projects.

Course Overview --This course connects interested data science project sponsors with data science students, so that both can accomplish together something that neither one could achieve alone. The overarching goal for the course is for the students to experience the real-world work of Data Science and to complete short consulting/technical projects in small teams.

Who is this course for -- This course is for anyone who applies their expertise to the demands of data-driven decision making and analysis. This is a "learning by doing" course on the practice of delivering data science expertise

This course provides an opportunity for students to learn:

> 1) How **to think** like a data scientist.
>
> 2) How to **apply the knowledge** required to practice data science.
>
> 3) How to continually **develop the skills** required to leverage your data science expertise, and
>
> 4) How to have the biggest impact in providing value and getting your expertise used though applications of the **tools and methods** of a data scientist.

During this course, you will complete several short real-world project segments as part of small teams.

Some students might think that **Data Science into Practice** means "how to run the data science technical machine"; that is of course part of the effort you will put forth in this class and in your career as a data scientist -- what this course will also do is to give you experience in real projects. You will be working in small teams of 3 to 4 students to put **Data Science into Practice**. This means you will

learn by doing and also develop your skills and practice how to work with people and projects. This includes finding customers, working with student colleagues to deliver great data science results and in the process fill out your resume with learning experiences that will help you build your proficiency and track record.

Schedule and how the course works:

When applying your skills as a data scientist, you soon realize that in real life there is a focus on finding what works. One thing everyone knows is that everyone wants to hire Data Scientist candidates that have a **diverse project portfolio** of experiences solving **real world business and decision problems**. So a key question for students becomes how to quickly develop thinking approaches, applied skills and then expert processes and then work on projects where you can get real world experiences. During this course we will work with project sponsors of actual data science problems.

Required Course Materials:

1) Textbook: Pete Block, Flawless Consulting: A guide to getting your expertise used, 2nd ed., Pfeiffer A Wiley Imprint, 3rd edition, 2000. 389 pages.

2) Textbook: Brian Godsey, Think like a Data Scientist, Tackle the data science process step-by-step. Manning Publications Co. 2017. 298 pages

There will be extensive resources and materials posted in the canvas files section of the course and readings will be assigned from these documents as well.

You will need a computer, webcam and a headset/microphone

Contact-Information
Email: kylestir@indiana.edu
Phone: 650-283-5953
Office hours: by appointment via telephone, or via email.

Draft "Consulting Club Participation Expectations / Commitments and Participation Requirements"

(Draft of: September 15, 2017)

I. Consulting Club Participant Level "Grinder" Commitments

Required:

- Commit to participation in the Data Science Practice "Consulting Club" for a minimum of one semester – This means that you will not drop the course except under extreme circumstances or situations – you are committing to participate for the entire semester. This means you are making a commitment to yourself, your project sponsors as well as a commitment to your teammates.

- Attend and participate in weekly Monday class zoom teleconference meetings with your cohort – This means one cohort meeting and two project team meetings per week.
  - The Fall 2017 will have up-to 3 cohorts of up to 20 students each and will meet weekly
  - Each cohort of 20 students will have 2 AI or Coaches (one for every 10 students)
  - Cohort 1 will meet on Mondays from 6:30PM to 9:00PM Eastern
  - Cohort 2 will meet on Mondays from 6:30PM to 9:00PM Pacific
  - Cohort 3 will meet at a time TBD.

- Clearly communicate your availability and desired Level of Commitment / Engagement to teammates and Associate Instructors (AIs). Participation and continued engagement will count significantly towards final grade.

- Complete an individual self-evaluation your performance 6 times during the fall 2017 semester and these will usually occur with the completion of each segment of your project and class requirements.

Fall Semester 2017 Self-Assessment Review Due Dates and final project sponsor feedback review.

Initial Review, Friday (Second week) at 11:59 PM Eastern

Second Review, Friday (Fourth Week) at 11:59 PM Eastern

Third Review, Friday (Eight Week) at 11:59 PM Eastern

Fourth Review, Friday (Twelfth Week) at 11:59 PM Eastern

Fifth Review, Friday (First day of finals) at 11:59 PM Eastern

Sixth Review, Friday (Last day of finals) at 11:59 PM Eastern Note: this final review consists of a peer (team mates) and project sponsor review which are not counted in your course grade and is solely used for you to develop your individual personal ongoing data science skill development plan.

- **Choosing your personal learning and development goals for Data Science Practice** -- At the beginning of the term you will create your own personal learning goals for the semester. These will include and prioritize individual goals in each of the 4 areas outlined below:

    1) **Data Science Thinking** (Your existing project approach and mind set) as articulated by Brian Godsey),

    2) **Data Science Knowledge** (Participation in the IU "Data Science Knowledge Base" and "On-Ramps") to round out a broad understanding of

    3) **Data Science Consulting Process Skills**, (Flawless Consulting Approach articulated by Peter Block)

    4) **Data Science Tools and Methods** (Selected by project sponsors, teammates and yourself)

II. Project Level "Minder" Level Commitments

Required:

- Commit to a specific project for the duration of the identified current project phase and semester.
- Attend and participate in project meetings, typically two per week.
- Contribute to team discussions via course tools and/or e-mail.
- Contribute to the project by completing assigned tasks and assessments on time.
- Encourage learning and feedback within the team.
- Evaluate self and team performance at requested intervals and at each project phase.

Optional:

- Assist in direct project leadership tasks, learning opportunities, and project promotion
- Serve as the "Back-Up" for the teammates and project leader (in case of absence, etc.)

III.   Project Leader Level "Finder Role"Commitments

Required:

- Review and commit to a project for the entire duration of the project
- Identify near-term and long-term project goals with project sponsor
- Provide leadership and encourage learning where appropriate
- Schedule regular team meetings (two times weekly)
- Schedule project sponsor meetings regularly (weekly or as needed)
- Lead discussions during team and project sponsor meetings
- Breakout and assign project tasks and time-frames (as needed)
- Facilitate communications with all team members about all relevant project developments
- Evaluate self and team performance at the end of each project phase

Example of Current Course Projects:

As a student you will work on a project similar to those described below. The Data Science Consulting Club started in September 2016. We have successfully collaborated with several companies – big and small, in delivering data science oriented projects. To date, we have completed 8 projects with many more in pipeline. Our course membership is expected to be around 60 students during the spring of 2018 – including both residential and online students. The projects that we undertake are 100% on-line and remotely staffed. These are typically short term, ranging from few weeks to couple of months.

Our projects can be broadly classified into 2 categories – research oriented and delivery focused. Projects that are in collaboration with external companies are mainly delivery focused, while internal projects are often research driven. Research projects tend to be longer running than delivery focused projects. Following is a table of projects that we have undertaken so far

Table 1: List of Projects – current and past

Project Name	Category	Timeline
MEPS – SPEA IU	Delivery Focused	Sep – Oct 2016
Mindshift Solutions	Delivery Focused	Dec 2016 – May 2017
PubMed MEDLINE	Research Oriented	Feb 2017 – ongoing
News Analytics (Yahoo)	Hybrid	Feb 2017 – ongoing
eKryp	Delivery Focused	Feb 2017 – ongoing
High Alpha	Delivery Focused	May – June 2017
Givaudhan	Delivery Focused	May –September 2017
Regenstrief Institute	Research Oriented	Jan 2017 –

The nature of work done is pre-dominantly technical with significant portion of programming using Python. Each of the projects have a GitHub repository with access limited to project members and clients. Delivery focused projects involve code deliverables at pre-arranged intervals with a review project sponsor

feedback. All projects have at least 2 -3 web meetings per week – one with the project sponsor and at least one internal engineering meeting. Long running projects are broken down into multiple iterations, with each iteration focused on delivering one or more project objectives. Some of the projects and their nature of work is shown in the table below

Table 2: Nature of Projects – Tools and techniques

Project Name	Description	Techniques Involved
PubMed MEDLINE	Create a clustered search on PubMed corpus of millions of medical/bio-medical journals using some High-Performance Computing Machinery at IU	Unsupervised learning – document clustering; indexing and search; word embeddings; feature extraction
Givaudhan	Extract and annotate ingredient information from online recipe data, which can then be used to predict current flavour trends	Supervised learning – Conditional Random Fields; ontologies; classification
High Alpha	Build a one-vs-rest classifier to classify documents into 20 categories where there is a severe case of class imbalance	Supervised learning – classification; Tf-IDf bag of words
Mindshift Solutions	Extract contextual information from short unstructured text documents, which can be used to extract meta-information about the corpus	Natural Language Processing – Part of Speech tagging; word embeddings; context extraction

Future Course Projects:

As a student in this class it is possible to participate in the recruitment of and establishing new partnerships with project sponsors:

Current Target Potential Project Sponsors Include: Those outlined in Appendix.

Course Learning Outcomes:

As a student in this class -- upon completing this course you will be able to:

1) Complete several data science consulting projects so that you can include these on your resume and build references.

2) Have a command of the Thinking Skills, Data Science Knowledge, Data Science Skills and Data Science Tools you require to apply your skills in Data Science during every stage of the consulting process.

3) Describe and give examples of how to conduct data science consulting behavior that is effectively used in engagements or projects.

4) Identify your current data science approach, knowledge, skills and skill gaps through reading, discussion and reflection.

5) appreciate the best practices and lessons-learned in developing, enhancing and applying your skills and data science technologies in a real world short project setting.

ASSESSMENT AND GRADING

Class Participation is required in weekly class and weekly small group meetings.

Deliverables – code and/or project research, as required by the project

Definitions of Letter Grades

Letter grades have been defined as follows and will be used as an aid in evaluation of academic performance and to assist students by giving them an understanding of the grading standards

A	4.0	Outstanding achievement. Student performance demonstrates full command of the course materials and shows a high level of originality and/or creativity that far surpasses course expectations.
A-	3.7	Excellent achievement. Student performance demonstrates thorough knowledge of the course materials and exceeds course expectations by completing all requirements in a superior manner.
B+	3.3	Very good work. Student performance demonstrates above-average comprehension of the course materials and exceeds course expectations on all tasks as defined in the course syllabus.
B	3.0	Student performance meets designated course expectations and demonstrates understanding of the course materials at an acceptable level.
B-	2.7	Marginal work. Student performance demonstrates incomplete understanding of course materials.
C+	2.3	Unsatisfactory work. Student performance demonstrates incomplete and inadequate understanding of course materials.
C	2.0	
C-	1.7	
D+	1.3	Unacceptable work.
D	1.0	
D-	0.7	
F	0.0	Failing.

Grades are assigned based on a combination of student performance measures developed for the semester. Student achievement of course objectives is usually assessed through the use of multiple performance measures. For example, a combination of several of the following assessment methods are common including graded discussions or participation, self-assessment, examinations, participation, written assignments and exercises, or work generated from small team consulting projects. Other methods, depending on course content and objectives, may also include other small-group exercises, oral presentations, and a review of applied data science projects.

COURSE Schedule

Module 1: Introduction (week 1)

Introduction to the course and to Canvas. Includes introductions discussion and class information/syllabus quiz and personal goal setting for the course.

Module 2: The Roles Data Science consultant serve in practice (week 2)

Overview of Module 2

Definitions

Thinking like a data scientist

Skills and Techniques

Actual Project Work

Module 3: Agreeing on a Scope of work and contracting (week 3)

Overview of Module 3

Contracting Overview

Conducting a Contracting meeting

Module 4: Putting it all together (weeks 4 and 5)

Overview of Module 4

Diagnosis and Discovery

Dealing with Process Problems

Moving from Data to Decisions (weeks 6 and 7)

Mid Term Meetings (Weeks 8 and 9)

Spring Break (Week 10)

Module 5 Moving from Data to Decisions (weeks 10)

Overview of Module 5

Tools to use with Clients

Module 6: Using the right Techniques at the right Time (weeks 11 and 12)

Overview of Module 6

Creating Trust

Addressing and Avoiding the Blockers

Module 7: Communicating Results Making an Impact (weeks 13 and 14)

Overview of Module 7

Review of checklist and process that work

Module 8 Reviewing the results and planning for future skill development (weeks 15 and 16)

Overview of Module 8

Module 9 (Week 17).

Small group review and feedback sessions.

OTHER INFORMATION

Accommodations for Students with Disabilities

Every attempt will be made to accommodate qualified students with disabilities (e.g. mental health, learning, chronic health, physical, hearing, vision, neurological, etc.) You must have established your eligibility for support services through the appropriate office that services students with disabilities. Note that services are confidential, may take time to put into place, and are not retroactive. Captions and alternate media for print materials may take three or more weeks to get produced. Please contact your campus office (Links to an external site.) as soon as possible if accommodations are needed.

Technology Accessibility Information

For accessibility information for persons using adaptive technology with Canvas, please visit Canvas Product Accessibility (Links to an external site.)

Technology Support

For Canvas questions please see the "Help" link at the top right of the page and check the UITS Knowledge Base for more information (type "Canvas" in the search box for a full list of Canvas-related topics).

If you have any other questions about or issues with any of the technology used in this course please contact the University Information Technology Services (UITS) support team. At IUPUI you can contact the support team by in the following ways:

- o   Call 812-855-6789
- o   Email ithelp@iu.edu

- o   Live Chat at http://ithelplive.iu.edu/
- o   Walk in at Learning Commons, Herman B Wells Library

Students Support Services

The Office of the Associate Vice President for Academic Support and Diversity provides resources and services to students in a variety of areas. For more information, resources, and contact information please see the Office website (Links to an external site.).
IUB policies regarding attendance, withdrawals, conduct, and religious holidays (Links to an external site.)

School Policies

There are a number of campus-wide policies governing the conduct of courses at the School of Informatics and Computing: review these policies online (Links to an external site.).

Academic Misconduct

Students are expected to abide by the rules and regulations of Indiana University in accordance with the Code of Student Rights, Responsibilities, and Conduct (Links to an external site.). Cheating or plagiarism are unacceptable and may result in a failing grade for the course.

Netiquette

One thing to always keep in mind when taking any online course is that the others with which you interact throughout the semester - including me, your instructor - are human beings. The first rule of netiquette (Links to an external site.) is to "remember the human" when you are communicating with me or with your peers. The second rule is to "adhere to the same standards of behavior online that you follow in real life." It's not likely that you would yell at, mock, or belittle another student in a face to face class so don't do it here. The feeling of anonymity that some people have when they are online can lead to those sorts of behaviors but they are **not acceptable** here or in any other online class.

Please take a few minutes and review all the Core Rules of Netiquette (Links to an external site.) and **make sure you have a profile picture added to Canvas** (instructions in the Getting Started with Canvas page) before beginning the class. A picture helps your fellow students "see the human" as we work together through the semester.