THE DATA MINE:
ACADEMIC PARTNER WORKSHOPS

Workshop #2: February 8, 2023

PURDUE
UNIVERSITY®

AGENDA

January 25	Intro | Stakeholders | Students
February 8	Students | Seminar | Technical Support & Research Computing
February 22	Corporate Partners
March 8	Business Partnership & Recruiting Industry Sponsored Projects
March 22	Legal | Operations

Please RENAME yourself in ZOOM:

University | Name   *Purdue | Jessica Jud*

STUDENTS

Recruiting
- *How will you source students?*
- *Where will you source from?*
- *Which departments or student communities?*

Registration
- *How will you set up your course in the Course Catalog?*
- *Does your institution have an established process for setting up new courses?*
- *What information will you need in preparation?*

Learning Communities
- *Does your institution have Learning Communities?*
- *Are there financial/staff/student resources available?*

Application
- *How will you gain an understanding of your students?*
- *Current knowledge/skills? Interests?*

Experience
- *How will you commit to the student experience?*
- *Will you host student events?*
- *Mentorship?*
- *Professional development opportunities?*

SEMINAR

- ## Course Tempo

  - 1x per week 50 mins (asynchronous)
    - 1 project per week = 1–3 hours per week
    - R in fall, Python in spring
    - 1 credit course
    - Minimum requirement for all data mine students

SEMINAR

- ## Course Tempo

  - 1x per week 50 mins (asynchronous)
    - 1 project per week = 1–3 hours per week
    - R in fall, Python in spring
    - 1 credit course
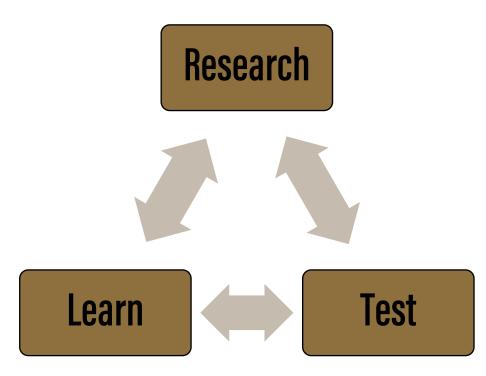   - Minimum requirement for all data mine students

- ## Building Projects

  - It can take a lot of talent and time to design technical seminar projects for 5 different course levels.

  - **Team Structure is important!**

    - *Do we have a project specialist?*

    - *Do we adapt existing coursework?*

    - *Does it depend on what department or group the program is in?*

    - *Other accommodations to consider?
      Ex: Add closed captioning*

TEACHING DATA SCIENCE – TEAM GOALS

- How do we ensure that we are teaching data science "the right way"

- Working with world-class faculty at Purdue

- Helping to teach students the iterative research method

```
        ┌──────────────┐
        │   Research   │
        └──────────────┘

┌──────────┐        ┌──────────┐
│  Learn   │ ◄────► │   Test   │
└──────────┘        └──────────┘
```

DESIGNING DATA SCIENCE CONTENT

MAKING DATA SCIENCE ACCESSIBLE

- Accessibility in terms of:

  - skill

  - communication style

  - learning styles

  - technical resources, and many other factors

How do we design something that isn't already available?

How would these resources be most helpful to students?

How is content appealing to participants?

DATA SCIENCE TECHNIQUES

- We are the first to admit that we are not experts at everything we teach

- The most important aspects are understanding the fundamentals and assumptions and team communication/collaboration

- Frequent topics include:

  - Time series analysis

  - Geospatial information systems (GIS)

  - Neural networks (all sorts)

  - Natural language processing
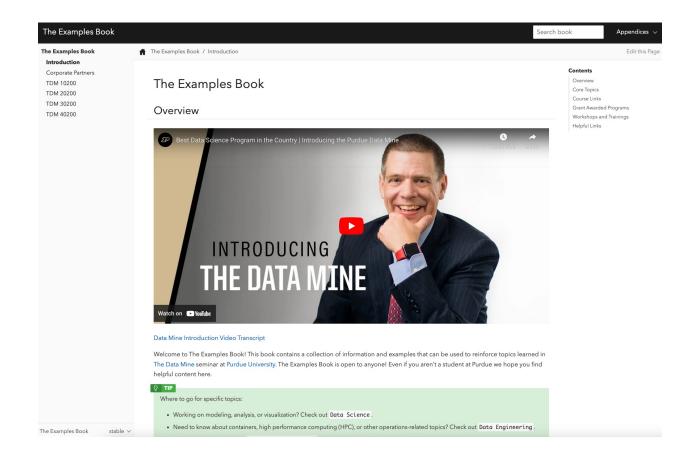
  - Web app development

DATA SCIENCE GOALS

- **What are the key components of analytics research that you want students to learn?**
  - The Data Science team within TDM focuses on building student's familiarity with research and testing
  - Our goal is to work with the world-class faculty to teach algorithms and more technical processes

- **How do we attempt to build those core skills?**
  - Labs have been vital in helping students identify and test different solutions
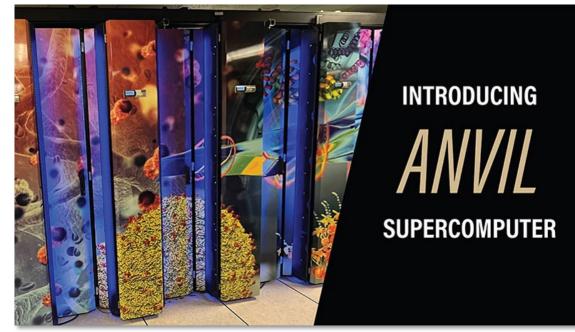 - Help to build the confidence to dive in and try things

THE EXAMPLES BOOK

- Open-source introduction to all things data

  - Hosts:
    - seminar projects
    - team support
    - data science
    - data engineering

- Built using The Data Mine's GitHub repository:

  - This is also where we host our kernel containers!

TECHNICAL ENVIRONMENT - HPCs

- The Data Mine runs almost everything on Purdue's Anvil environment

  - https://www.rcac.purdue.edu/compute/anvil

  - All of our Python and R kernels run here

- The Python and R kernels are containerized!

  - Much easier for support

  - Better version control

  - Do have slow build times

- The Data Mine also supports Windows-only applications like Power BI or Tableau

INTRODUCING

ANVIL

SUPERCOMPUTER

RESEARCH COMPUTING TEAM

GETTING TO KNOW YOUR RESEARCH COMPUTING TEAM IS ESSENTIAL!

- The Research Computing (RCAC) team at Purdue is key to enabling everything that we do.

Tip: Be prepared to push your University in unexpected ways

- We use servers in ways that the University doesn't typically see
- Python isn't a language that's widely used in Purdue's environment

DO YOU HAVE A GOOD POINT OF CONTACT FOR TECHNICAL RESOURCES?

TECHNICAL LESSONS LEARNED

Consolidated environments are much easier to support, but they do create bottlenecks

Technological flexibility is a challenge, but is necessary

Computational efficiency is core to many data science challenges

Student collaboration and iterative testing make projects excel

Be sure to thank your Research Computing team! :)

LET'S

TALK

PURDUE
UNIVERSITY.

WHAT'S NEXT?

January 25	Intro | Stakeholders | Students
February 8	Seminar | Technical Support & Research Computing
February 22	**Corporate Partners**
March 8	Business Partnership & Recruiting Industry Sponsored Projects
March 22	Legal | Operations

- All post-meeting materials will be posted on The Examples Book

- Questions? Email Jessica Jud at jljud@purdue.edu

THANK YOU