Introduction to Scientific Computing II

Amir Farbin

A Data Science BS Degree

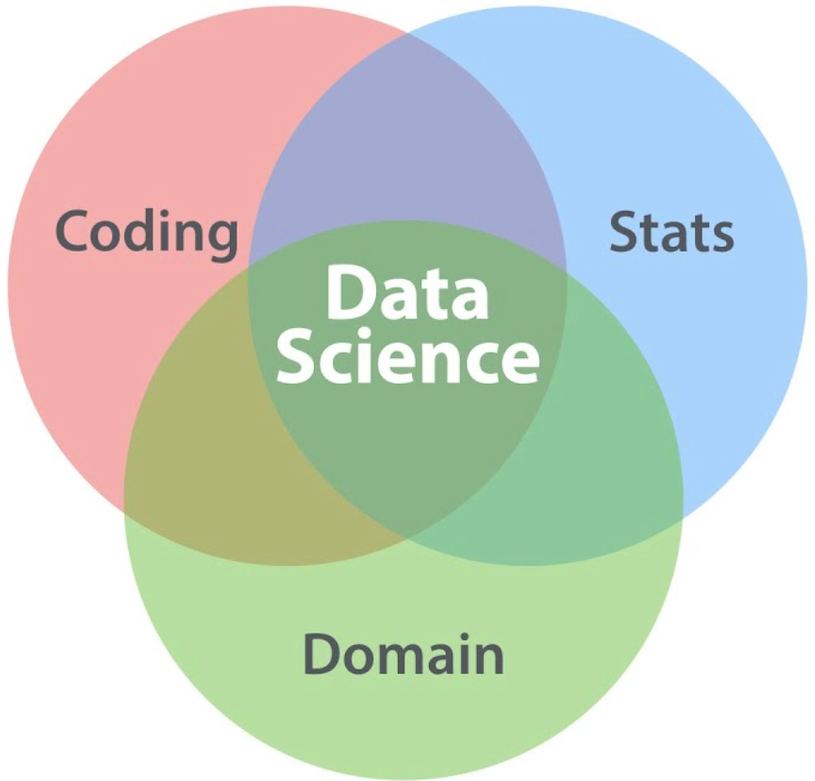
- The Degree Proposal has been in progress for 3 years.
- **Approved** by *Texas Higher Education Coordinating Board* (October 22)
    - Full program launch Fall 2021.
- Courses available since Fall 2018.
- *Minor* define Fall 2020.
- Unique Degree
    - Undergraduate
        - Most programs are professional masters or PhD.
    - Within *College of Science*
        - Most are in Computer Science or Business
        - Requires concentration and Capstone Project
- Aim to prepare students:
    - Entry-level Data Science jobs
    - Better Science Research
        - Undergraduate
        - Better positioned for Graduate School
            - Stronger Application
            - Start on research quicker

The UTA Data Science faculty

https://www.uta.edu/science/data-science/

The Challenge

- Practicing Data Science requires
    - Coding
    - Math (e.g. Statistics)
    - Domain expertise (e.g. physics, biology, …)
- Each of these areas can be a degree onto itself…
    - Usually people come into data science from one of these areas.
- *Challenge*: Start with a University Freshman with no preparation.



The Courses

- Note courses are being renumbered/renamed for Fall 2021. Pending UCC approval:
    - ***DATA 1301*** – Introduction to Data Science
    - ***DATA 3401*** (formerly 1401) -- Python for Data Science 1
    - ***DATA 3402*** (formerly 1402) -- Python for Data Science 2
    - ***DATA 3421*** (formerly 3401) -- Data Mining, Management, and Curation
    - ***DATA 3441*** (formerly 3402) -- Statistical Methods for Data Science 1
    - ***DATA 3442*** (formerly 3403) -- Statistical Methods for Data Science 2
    - ***DATA 3461*** (formerly 3404) -- Machine Learning
    - ***DATA 4380*** (formerly 4301) -- Data Problems
    - ***DATA 4381*** (formerly 4302) -- Data Capstone Project 1
    - ***DATA 4382*** (formerly 4303) -- Data Capstone Project 2
- Major: core + these courses + 2 math courses (and calculus I)
- Minor: ~ 5 courses

Logistics

- ***Lectures***: Tuesday/Thursday 2-3:20
- **Lab** (50%) Friday
- 2 Stages:
    - Pre-spring break: Like Data 1401
        - ~4 Labs, each taking ~2 weeks
        - 50% of grade
- Post-spring break: Projects/Presentations
    - Project in 4 steps: proposal, feasibility, prototype, production
    - One presentation at each step.
    - Details TBA.
    - 50% of grade

Logistics (2)

- Homework Policy
    - You can work with others, but do not copy/paste code from another student.
    - Submitted via git (Version Control System).
- Help
    - ***Clinic***: We usually run a clinic where students can ask for help… wasn't effective virtually.
        - Will run a poll to determine optimal time for an additional help session.
    - ***Office Hours*****:** I'll generally be available after every classes session.
- ***Laptop*** (with a physical keyboard).
    - Highly recommended to have a laptop for this class
        - Doesn't matter what OS you run… all you need is a browser.
    - If you don't have a laptop, you can rent one. Details …
- First lab will be this Friday (1/22)

Course Plan

- Before Spring Break (~ 8 Weeks)
  - Setting up
    - Platforms: Linux, Windows, Mac
    - Unix Review
    - PyPi
    - Virtual Environment
    - Containers
  - Review of 1401
    - Pandas
    - numpy
    - Matplotlib
  - Advanced Python
    - Decorators
    - Computation
      - Mutli-treading/multi-processing
    - TensorFlow/PyTorch as computation engines
- Packages/Projects
  - Sympy
  - SciPy
  - SciKit
  - SciKit-learn
- Statistical Inference
  - Statistical Modeling
  - Parameter Estimation
    - Regression
    - Maximum Likelihood
    - Confidence Intervals
  - Monte Carlo
  - Machine Learning
    - Classification
      - Performance Measures
    - Deep Learning
      - Problem Formulation
- Map Reduce

- After Spring Break
  - Kaggle
  - Project
    - Proposal
    - Feasibility
    - Prototype
    - Production
  - Software Development
    - Waterfall vs Agile

Introduce Yourself

- Your UTA Degree
    - What is your major?
    - What year?
    - When will you graduate?
- Interests
    - Is there a specific scientific or professional field?
    - Have you done any research?
    - Any hobbies, etc, that you can apply DS to?
- Your goals
    - What’s next (job, grad school)?
    - How can this course help?
- Your setup
    - What kind of computer? (Windows, Mac, Linux)
- Anything else…