# Course Logistics

**Title:** Agriculture Data Science (4 credits)

**Course number and section:** APSC-5984

**Term:** Spring 2023

**Meeting time and location:**

**Lecture**:Monday andWednesday 4:00 – 5:15 pm, SAUND (Saunders Hall) 408

**Lab**: Friday 2:30 – 4:30 pm, JCH (Cheatham Hall) 317A

**Format:**

Two lectures (75 minutes each), for each lecture:

* 30-minute theory instruction
* 30-minute live coding session
* 15-minute discussion

One lab course (2 hours)

# Instructor Information

**Name:** Dr. James Chen

**Office Location:** Litton-Reaves 3090

**Office Hours:** By appointment or GitHub Classroom

**Email Address:** [niche@vt.edu](mailto:niche@vt.edu)

# Course Information

**Course Description:**

An unprecedentedly significant amount of data is produced in the modern agriculture industry every day. As a discipline to organize, analyze and visualize large data sets, data science has become essential knowledge for agriculture students. The course will cover many important topics in data science, including data preprocessing, database construction, supervised and unsupervised learning models, data visualization, and web app development. Students will work with real agriculture production data and implement each core topic using the programming language, Python. This course also requires students to attend the laboratory section, where further hands-on experience in data analysis can be obtained. The students will have a chance to establish a programming environment and derive a Python script to solve real-world problems using example datasets.

**Course Prerequisites:**

Students should have a basic understanding of analyzing research data with Microsoft Excel. Familiarity with Python would be very helpful in this class. However, students are assumed to have no background in any programming language.

**Course Objectives:**

After completing the course, the students are anticipated to

* Be capable of performing data analysis on a large scale (> 10,000 records).
* Use programming language to help understand their research data.
* Be able to develop a computation tool to answer a scientific question on their own.

**Textbooks:**

* Daumé, H. *A Course in Machine Learning.* 2017 (<http://ciml.info/>)
* Goodfellow, I., Y. Bengio, and A. Courville. 2016. *Deep Learning*. MIT Press. 2016. (<https://www.deeplearningbook.org/lecture_slides.html>)
* McKinney, W. *Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter*, 3rd edition. O’Reilly Media. 2022. (<https://wesmckinney.com/book/>)
* Szeliski, R. *Computer Vision: Algorithms and Applications*, 2nd edition. Springer. 2022. (<https://szeliski.org/Book/>)

**Required Materials:**

* A laptop
* A Google account (for Google Colab)

# Grading

**Grading Methods:**

Assignment 60%

Final project 15%

Lab final exam 25%

**Grading Scale:**

|  |  |
| --- | --- |
| **Grade** | **Aggregate Score Range** |
| **A** | 94 – 100 |
| **A-** | 87 – 93 |
| **B+** | 80 – 86 |
| **B** | 75 – 79 |
| **B-** | 70 - 74 |
| **C+** | 65 – 69 |
| **C** | 60 – 64 |
| **F** | < 60 |

# Course Schedule (Tentative)

|  |  |  |  |
| --- | --- | --- | --- |
| **Weeks** | **Lecture 1** | **Lecture 2** | **Lab** |
| Week 1 | Martin Luther King Jr. Day | Why Data Science? | Environment setup |
| Week 2 | Coding environment | If statements and for-loops | Basic Python implementation I |
| Week 3 | Data frames I | Data frames II | Pandas library |
| Week 4 | Data preprocessing I | Data preprocessing II | PCDART database |
| Week 5 | Data quality control | Introduction to database I | Futures market database |
| Week 6 | Presidents’ Day | Introduction to database II | Database application |
| Week 7 | Design of database I | Design of database II | SQLite3 |
| Week 8 | Spring break | | |
| Week 9 | Supervised learning: regression model I | Supervised learning: regression model II | Python Scikit-learn |
| Week 10 | Feature selection | Model validation | Feature selection |
| Week 11 | Principal Component Analysis | Unsupervised learning:  K-means clustering | Python Scikit-learn |
| Week 12 | Intro to computer vision | Computer vision: convolution | Python OpenCV |
| Week 13 | Data visualization | Web app deployment | Python Bokeh |
| Week 14 | What is object-oriented programming | Encapsulation | OOP implementation |
| Week 15 | Object and Class in Python | Inheritance and polymorphism | OOP implementation |
| Week 16 | Project presentation | Project presentation | Final exam |