# AST4930: Special Topics: Machine Learning and Astronomy

**Fall 2025 | 3 credits**

---

**NOTE:** This course complies with all UF academic policies. For information on those policies and for resources for students, please see UF's "Academic Policies and Resources" web page.

## I. General Information

**Meeting days and times:** Tuesday 10:40 am – 11:30 am, Thursday 10:40 am – 12:35 pm

**Class location:** TUR 2319

**Instructor:**

Name: Jaehan Bae
Office Building/Number: BRT 316
Phone: 352-294-1890
Email: jbae@ufl.edu
Office Hours: Wednesday 1:00 pm – 3:00 pm

**Teaching Assistant:**

Name: Alejandro Camazon
Office Building/Number: BRT 309
Email: acamazon@ufl.edu
Office Hours (in-class worksheet, homework): Tuesday 12:00 pm – 2:00 pm

**Course Description**

Modern science is increasingly enabled by efficient, careful, and detailed analysis of big data. This course will introduce students to essential AI/ML techniques, emphasizing practical application to research-quality datasets through hands-on programing experience. The course will feature coding-based in-class worksheets, homework assignments which will employ real-world datasets, and a term project. The course will be structured with 7 modules (see Section IV. Calendar). Each module will span 1-2 weeks and will include lectures on core AI/ML methods/concepts and hands-on coding exercises which are designed to help students become proficient in applying these methods to real-world problems. The course will be Python-based and will make use of open-source Python packages including Scikit-learn, TensorFlow, and Keras.

**Prerequisites**

AST 2730: Introduction to Python, equivalent Python courses, or instructor permission.

**General Education Designation:** none.

## Course Materials

The course will be taught based on course notes and combination of the following books:

- *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, O'Reilly Media, Géron, ISBN: 978109812
    - This will be the main textbook of the course.
    - It provides an in-depth introduction to Machine Learning with Python including heavy use of Scikit-learn, Keras, and TensorFlow.
    - The full text of this book is available to read online at UF libraries.

- *Introduction to Machine Learning with Python*, O'Reilly Media, Müller & Guido, ISBN: 9781449369415
    - It provides an in-depth introduction to Machine Learning with Python including heavy use of Scikit-learn, but with a more general (i.e. less physical) focus.
    - Accompanied with a series of Jupyter notebooks that give good examples that are significantly broader in nature than those provided with the *Python Data Science Handbook* (see below).
    - The full text of this book is available to read online at UF libraries.

- *Python Data Science Handbook,* O'Reilly Media, VanderPlas, ISBN: 9781491912058
    - If you are not familiar with Python this book offers a great place to start.
    - The full text of this book is freely available online at https://jakevdp.github.io/PythonDataScienceHandbook/

- *Statistics, Data Mining & Machine Learning in Astronomy*, Princeton University Press, Ivezić, Connolly, VanderPlas, & Gray, ISBN: 9780691198309
    - Part of the book (and lots of useful resources) is available online at http://www.astroml.org/
    - This book is probably (on average) a bit above the level of this course. However, the text is very useful in that it is both Python-based and it provides abundant examples about how a wide range of techniques can be applied to astrophysical survey (i.e. very large datasets).
    - The full text of this book is available to read online at UF libraries.

- *Machine Learning techniques for Physics and Astronomy*, Princeton University Press, Acquaviva, ISBN: 9780691203928
    - This book offers probably the most relevant examples to this course as the author is an Astrophysicist. Some astrophysical examples from this book will be introduced during the course.

**Materials Fee:** There are no additional material and supply fees for this course.

## II. Course Goals

**Course Objectives**
This course will:

- introduce foundational AI/ML techniques
- provide hands-on experience with research-quality datasets that reflect real-world scientific challenges
- integrate theoretical knowledge with practical application through coding exercises and project-based learning

**Student Learning Outcomes**
A student who successfully completes this course will be able to:

- apply core AI/ML algorithms to analyze complex, large-scale datasets
- program confidently in Python using essential ML libraries
- evaluate and select appropriate AI/ML algorithms based on data and research questions to obtain meaningful insights from real-world data
- complete an independent research project and communicate findings effectively through term project presentations

## III. Graded Work

**Graded Components**

| Work | Description | Points |
|---|---|---|
| In-class Worksheets | In-class worksheets will be assigned during classes to give students opportunities to review the course material and give the instructor the opportunity to check your comprehension of the material. **Worksheets will be due at the end of the class they are assigned and are not accepted late.** Discussing with other students and working in groups are strongly encouraged but write your own answers without looking at the other students' notebook. All in-class worksheets must be submitted through the canvas website. | 30 |

| | | |
|---|---|---|
| Homework Assignments | Homework assignments consist of comprehension questions and coding problems. Homework assignments will ask students to apply the concepts and techniques from the lecture and readings, with a goal of assessing student comprehension. Homework assignment sets will be graded for accuracy. All homework assignments must be submitted through the canvas website. | 30 |
| Exams | There will be one exam in this course. The exam will be closed book and closed notes. The exam is designed to assess student comprehension of the concepts covered in the course and will feature topics from lecture and homework assignments. The exam will be graded for accuracy. | 20 |
| Term Project & Presentation | The term project will be used for students to apply the concepts and techniques from the course to real world astronomical datasets in a more in-depth fashion. Students are welcome to consult with the instructor on the topic of their term project, ideally early during the semester. The term project will consist of submitted code, a write-up, and two oral presentations to the class (proposal + final). The term project will be graded based on the accuracy and efficiency of the submitted code, the submitted written project description, and the quality of oral presentations. | 20 |

The course canvas site will make clear all assignment dates and deadlines. Any questions about deadlines should be directed to the instructor.

**Grading Scale**

For information on how UF assigns grade points, visit: https://catalog.ufl.edu/UGRD/academic-regulations/grades-grading-policies/

| | | | | |
|---|---|---|---|---|
| A | 90 – 100% | | C | 70 – 73.99% |
| A- | 87 – 89.99% | | C- | 67 – 69.99% |
| B+ | 84 – 86.99% | | D+ | 64 – 66.99% |
| B | 80 – 83.99% | | D | 60 – 63.99% |
| B- | 77 – 79.99% | | D- | 57 – 59.99% |
| C+ | 74 – 76.99% | | E | <56.99% |

*Note: A minimum grade of C is required to earn General Education credit.*

## IV. Calendar

| Week | Date | Topic |
|---|---|---|
| Week 1 | 08/21 | Introduction to machine learning |
| Week 2 | 08/26<br>08/28 | Module 1: Supervised learning for classification – k-nearest neighbors and decision trees |
| Week 3 | 09/02<br>09/04 | Module 2: Evaluation metrics, cross validation, parameter optimization, imbalanced data |
| Week 4 | 09/09<br>09/11 | Module 3: Supervised learning for classification – Support vector machines |
| Week 5 | 09/16<br>09/18 | Module 4: Regression methods |
| Week 6 | 09/23<br>09/25 | Module 5: Ensemble methods for classification and regression |
| Week 7 | 09/30<br>10/02 | Module 6: Unsupervised learning – Clustering |
| Week 8 | 10/07<br>10/09 | Module 6: Unsupervised learning – Dimensionality reduction |
| Week 9 | 10/14<br>10/16 | Term project proposal presentations |
| Week 10 | 10/21<br>10/23 | Module 7: Deep learning |
| Week 11 | 10/28<br>10/30 | Module 7: Deep learning |
| Week 12 | 11/04<br>11/06 | Review & Exam |
| Week 13 | 11/13 | Term project presentations |
| Week 14 | 11/18<br>11/20 | Term project presentations |
| Week 15 | | Thanksgiving (no class) |
| Week 16 | 12/02 | Discussion on using AI/ML in higher education |

## V. Procedure for Conflict Resolution

Any classroom issues, disagreements or grade disputes should be discussed first between the instructor and the student. If the problem cannot be resolved, please contact the Undergraduate Coordinater Prof. Desika Narayanan (desika.narayanan@ufl.edu, 352-294-1865) or the Department Chair Prof. Elizabeth Lada (elada@ufl.edu, 352-294-1862). Be prepared to provide documentation of the problem, as well as all graded materials for the semester. Issues that cannot be resolved departmentally will be referred to the University Ombuds Office (http://www.ombuds.ufl.edu; 352-392-1308) or the Dean of Students Office (http://www.dso.ufl.edu; 352-392-1261).

## VI. Using HiPerGator Resources

Students may use HiPerGator resources for homework and term projects. Students that need to access HiPerGator for course-related activities should ensure that they are properly registered in their course. Those who already have HiPerGator accounts will be added to the class group; the rest will have a temporary HiPerGator account created for them to use for the class. Students who join the class late should remind their instructor to send Research Computing a request to add them to the class. Note that class accounts will expire (and any associated data deleted) two weeks after the "Classes End" date listed for the semester: UF Catalog: Dates & Deadlines.