AIR FORCE INSTITUTE OF TECHNOLOGY

SCHOOL OF SYSTEMS AND LOGISTICS

ADVANCED DATA SCIENCE IN PYTHON

WKDSS420

SYLLABUS

COURSE TITLE

SYLLABUS

Contents

COURSE TITLE

Course Description

The overall goal for this course is to introduce students to various advanced level concepts of data science, especially in regard to application. This workshop will train students on a range of topics including Recommender Systems, Reinforcement Learning, Natural Language Processing, Transformers, Generative AI, and AI Bias and Ethics.

Course Learning Objectives

At the end of this course each student will be able to:

- Identify the 2 main types of Recommender Systems
- Understand how cosine similarity is used to identify similar items
- Develop a recommender system through 2 different methods
- Understand when to use Reinforcement Learning over Supervised Learning
- Identify the major components of a Reinforcement Learning model
- Identify the core components of Natural Language Processing (NLP)
- Build a NLP model that can perform binary classification
- Understand concepts of Attention, Encoder-Decoder model, and Transformers
- Build and use a Transformer Model, and edit it for a specific use case
- Understand different forms of Generative AI and applications in industry
- Identify types of Bias in AI
- Identify the Department of Defense Ethical AI Principals

Student Evaluation and Standards

The student is evaluated with Python language exercises throughout each lesson and either instructor led or online knowledge checks throughout.

Instructional Methods

This course is designed to prepare Department of the Air Force (DAF) members who may not already be data scientists but have a STEM background and seek to apply data science methods to their missions. This is achieved using recorded virtual lessons. Online presentations are made by faculty members of the School of Systems and Logistics.

Academic Freedom, Non-Attribution, Academic Integrity, and Student Rights

All students must be familiar with and adhere to the standards of academic freedom, non-attribution and academic integrity. Refer to the Student Handbook for these standards.

Student Preparation

The student is expected to either have taken Workshop DSS 220 and DSS 320, or be familiar with the concepts, including basic knowledge of Python programming, Pandas and Numpy libraries and the basics

of cleaning, analyzing and data visualization of a data science project, and fundamentals of Neural Networks. It is desired, but not required, that the student have rudimentary knowledge of calculus and linear algebra, to include partial derivatives, matrix multiplication and dot products, and matrix inversion.

Students should either have Anaconda and Jupyter Notebooks installed on their government or personal computers. They must have administrative rights to install various libraries.

Recommended Textbooks

Although not required, following textbooks are helpful resources for this course and beyond:

- Practical Recommender Systems by Kim Falk
- Getting Started with Natural Language Processing by Ekaterina Kochmar
- Natural Language Processing in Action by Hannes Hapke, Hobson Lane, Cole Howard
- Reinforcement Learning by Phil Winder

Student Evaluation of the Course

There are two principal means of providing student evaluation of the course content and conduct:
1. Providing direct feedback to the course instructor at any time.
2. Completing an end-of-course evaluation online.

COURSE TITLE

Course Schedule

Lesson No.	Estimated Time
Module 0 Introduction to the Workshop	6 Minutes
Module 1 Recommender Systems	84 Minutes (total)
Lesson 1: Recommender Systems	84 Minutes
Module 2: Reinforcement Learning	33 Minutes (total)
Lesson 1: Reinforcement Learning	33 Minutes
Module 3: Natural Language Processing	195 Minutes (total)
Lesson 1: Natural Language Processing Theory	94 Minutes
Lesson 2: Transformers	101 Minutes
Module 4: AI Ethics and Bias	30 Minutes (total)
Lesson 1: AI Ethics and Bias	30 Minutes

Total Course Duration   **348 Minutes (5 hours 48 Minutes)**