



Intermediate Data Science with Machine Learning

Technical Component - Morning

Sana Jabbar

sana.jabbar@lums.edu.pk

Rehman Sadaqat

zaynali0987@gmail.com



Warning: All the information in the said document relates only to the technical session.

Learning Objectives

This AI course is intended for students to learn the essential foundations of AI and gain fundamental data science skills through hands-on exercises.

- Understand the foundational math behind data science and machine learning: linear algebra, probability, and statistics.
- Be able to do data preprocessing with the Python libraries (NumPy and Pandas) for the execution of optimal machine learning models and data visualization
- Explore supervised and unsupervised learning and be able to apply the most suitable machine-learning algorithm.
- Learn to process textual data to derive high-quality information from text and apply new insights to real-world business (NLP).
- Build and train deep neural networks, use deep learning libraries such as TensorFlow and Keras to gain proficiency, as well as handle various deep learning techniques.

Course Modules

Numpy	Statistical Analysis Using Numpy
Pandas	Exploratory Data Analysis, Data Pre-processing, Data Cleaning, filing (json, csv, txt), regex
Matplotlib	Visualizations using matplotlib
Seaborn	Visualizations using seaborn
Plotly	Visualizations using Plotly
SQL	Basic commands of SQL and practice webs
Machine Learning	Mathematics behind ML; regression, classification and clustering techniques, PCA, GD
Deep Learning	Neural Networks, Partial Derivatives, YOLO, CNN, RNN, LSTMs, RCNN
Power BI	Visualizations, EDA, Reports, Interactive Dashboards

Week 1

1. What is data science and its applications?
2. Introduction to Python for data science
3. Basic data analysis using Numpy
4. Introduction to NumPy and its functions for mathematical operations on arrays
5. **Mini Project 1:** Hands-on practice with 4 Jupyter Notebooks

Week 2 - 3

6. Overview of statistical concepts like mean, median, mode, and standard deviation
7. Loc and iloc practices using boolean condition
8. Probability distributions and hypothesis testing
9. Overview of data wrangling and cleaning
10. Cleaning and transforming data using Python libraries like Pandas
11. Handling missing values and outliers
12. Merging and reshaping data
13. Regex
14. Loading data with different file formats and performing EDA (txt, JSON)
15. Introduction to Python libraries such as Scipy and Statsmodels
16. Installation and setup of the required libraries
17. Importing and loading data
18. Performing statistical analysis using Python libraries
19. Visualization of results using Matplotlib or other visualization libraries
20. **Mini Project 2:** Hands-on practice implementing statistical analysis using Python libraries like Numpy and Pandas
21. **Mini Project 3:** Hands-on project using the real-life dataset implementing Regex and JSON formats

Week 4 - 5

22. Why visualization is important in data science
23. Introduction to Power BI for data visualization
24. Connecting data sources within Power BI
25. Creating basic charts and graphs in Power BI
26. Understanding advanced visualization techniques in Power BI
27. Creating interactive visualizations in Power BI
28. Customizing colors, labels, and titles
29. Adding filters and sorting data

- 30. Working with calculated fields and parameters in Power BI ; DAX
- 31. Best practices for designing effective dashboards in Power BI
- 32. Hands-on practice creating visualizations using real-world datasets
- 33. **Mini Project 4:** A full-fledged advanced Power BI dashboard for portfolio and practice

Week 6

- 34. What is machine learning and its applications
- 35. Supervised and unsupervised learning
- 36. Types of machine learning algorithms: linear regression, classifications, and clustering
- 37. Introduction to sci-kit-learn library for machine learning
- 38. Hands-on practice implementing basic machine learning algorithms in Python
- 39. **Mini Project 5:** Implementation of Machine Learning Models using real-time datasets.

Week 7 - 8

- 40. What is SQL
- 41. SQL and data science
- 42. Basic Queries Structure
- 43. SQL Commands: DDL, DQL, DML, DCL, and TCL With real-life datasets.
- 44. **Mini Project 6:** SQL Project on a real-life dataset to implement data retrieval processes.

Week 9 - 10

- 45. Deep learning and neural networks
- 46. RNN, CNN, and LSTM architecture
- 47. Activation functions: Sigmoid, ReLU, and Tanh.
- 48. Backpropagation algorithm: adjusting weights and biases for error minimization.
- 49. Popular deep learning frameworks: TensorFlow, Keras, and PyTorch.
- 50. YOLO interpretation for Computer Vision
- 51. Object Detection, Image Segmentation, and Image Analysis
- 52. Natural language processing (NLP) and text analytics; Large Language Models
- 53. **Mini Project 7:** Implement Computer Vision/NLP solutions on real-life problem

Week 11 - 12

- 54. **Mega Project** Proposal Submission
- 55. **Mega Project** Evaluation 1
- 56. **Mega Project** Evaluation 2
- 57. **Mega Project** Submission
- 58. **Mega Project** Poster Design
- 59. Viva & Presentation

DATA SCIENCE COURSE CAPSTONE PROJECT

A trainee will have around 2 weeks of uninterrupted timeslots to work on the Capstone Project.

The capstone project evaluation criteria for the trainees will be as follows:

Assessment	Evaluation Criteria	Points
Proposal	<ul style="list-style-type: none">• Creativity and novelty.• Differentiation from the existing known cases.• Impact on the public interest.• Method and technique.	10
Idea Presentation	<ul style="list-style-type: none">• Ability to explain your idea clearly, concisely, and audience engagement	05
Evaluation 1	<ul style="list-style-type: none">• 30% of your project completed.• A prototype in function (if applicable)• Intermediate understanding of your project architecture.	10
Evaluation 2	<ul style="list-style-type: none">• 100% of your project is completed.• Working model/output and/or front-end• Complete understanding of your project architecture.	10
Viva	<ul style="list-style-type: none">• Complete Oral Understanding of your project architecture, workflow explanation, and alternatives explored	05

The final decision regarding your Mega Project rests with the instructors.

The weightage of all modules for evaluation for the trainees will be as follows:

The information regarding the percentage are as follows:					
Workbooks	15%	Projects	20%	Mega Project	30%
Quiz	20%	Assignments	10%	Class Participation	5%

The following assessment modules are compulsory for you to attend. **Missing out on any of them will result in ‘terminating’ status for graduation**

1. Mega Project Submission(s)
2. Knowledge Streams Overall Scheduled Assessment(s)
3. Mock Interviews

Attendance Guidelines

1. At most 5 approved leaves are allowed for a trainee. These approved leaves are only in case of health/family emergencies. Work and other prior commitments that could have easily been scheduled on weekends are not approved.
2. 0 unapproved leaves are allowed to a trainee. If a trainee has unapproved leaves, his/her name will be forwarded to the administration accordingly.
3. You are to write an email with the reason for leave and days required off in an email which is to be sent to zaynali0987@gmail.com and cc'ed to sana.jabbar@lums.edu.pk and maliha.sheikh@knowledge.tech
4. An email MUST be sent to the instructors before the class for an approved leave. If an email is received more than 30 minutes after the class is in session, the leave will automatically be considered unapproved; unless specified otherwise.
5. If a student misses a workbook due to leaves, he/she will have his/her marks obtained in quizzes reflected on the workbooks (as a percentage)

For example, if trainee A has taken off for one day and has missed Pandas Workbook 04 out of 11 total workbooks, his/her marks obtained in the Pandas Quiz will be substituted into the marks of Pandas Workbook 04 rather than allowing him to resubmit the work (to minimize the chances of plagiarism)

6. Quizzes will not be substituted for the students who are on unapproved leave. Students who are on approved leave shall have an interview with the instructors regarding the modules which then can be substituted instead of quiz marks.
7. Instructors have the authority to declare an announcement and/or change in marking, classes, and relative stuff on the spot.
8. Unless a trainee provides satisfactory results on Machine Learning, he/she will not be allowed to move on to the DEEP LEARNING part of the course.
9. DEEP LEARNING part of the cohort is entirely optional and the decision rests with the instructors.
10. You may reach out to the instructors during business hours on WhatsApp and/or Email. However, Whatsapp DM's are strictly prohibited after business hours. You may reach out to the instructor via email 24/7. You can also schedule an appointment or do a walk-in (

as long as your instructors are on campus) one-on-one session with your instructors to discuss your progress.

11. Instructors also have the authority to change the weightage of each module, as desired.

DATA SCIENCE COURSE CERTIFICATION

Trainees will be eligible for graduation certificates if and only if they meet the following criteria.

Assessment	Use of Assessment	To Qualify
Grade	Minimum assessment score required to qualify	60%
Attendance	Minimum attendance required to qualify	80%

You may reach out to your instructors to discuss specific parts of this document in detail.