

Practical Machine Learning and Data Exploration

Lecture Intro

Prepared by
Dr. Hager Saleh



Who am I?

Hager Saleh Mohammed Ahmed

About me

Assistant Professor at the Artificial Intelligence Department at Midocean University

Educational Degrees

Ph.D. from Faculty of Computer and Information , Minia University

Ph.D's thesis title: Efficient Predictive Analytics using Big Data Platforms



Objectives

Course Objectives Summary:

1. Installing the main environment for the practical application.
2. Understanding the main knowledge before applying machine learning.
3. Understand the role and importance of Exploratory Data Analysis (EDA).
 - Detecting outliers.
 - Handle missing or corrupted data.
 - Visualize data using various techniques (e.g., histograms, scatter plots).
 - Use Python libraries (Pandas, Matplotlib, Seaborn) for EDA.
4. Machine Learning (ML):
 - Understand the machine learning workflow.
 - Train supervised models
 - Evaluate model performance using accuracy, precision, and other metrics.
 - Studying different types of splitting dataset.
 - Optimize hyperparameters of models.
5. Key aspects of ensemble Learning
6. Apply feature engineering techniques.
7. Apply knowledge through hands-on real-world projects.

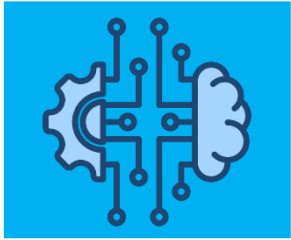


Exploratory Data Analysis (EDA)

Key aspects of EDA include:

- **Handling Missing Values:** Detecting and deciding how to address missing data points.
- **Summary Statistics:** Calculating key statistics that provide insight into data trends and nuances.
- **Outlier Detection:** Identifying unusual values that deviate from other data points, and how handling these.

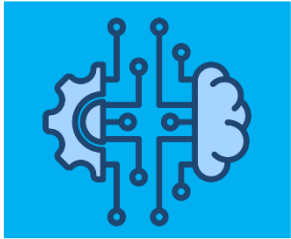
Graphical Representations: Utilizing charts such as histograms, box plots, scatter plots, and bar charts to visualize relationships within the data and distributions of variables.



Machine Learning (ML)

Key aspects of Machine Learning include:

- Supervised learning involves training a model on labeled data
- Studying two types of Supervised learning:
 - Regression: Predicting continuous values (e.g., house prices).
 - Classification: Predicting categorical outcomes (e.g., spam vs. non-spam emails).
- Studying different types of splitting dataset.
- Exploring different types of evaluation models.



Ensemble Learning

Key aspects of ensemble Learning include:

- Studying different models of ensemble learning such as bagging, boosting, stacking, or voting.
- Studying different types of splitting dataset.
- Exploring different types of evaluation models.



Feature Selection (FS)

Key aspects of Feature Selection include:

- Studying the main objectives of applying feature selection methods.
- Exploring different types of feature selection methods.
- Practically applying feature selection methods.



Assignments

Assignments	Due Week	Degree
Assignment1	Week 4	50%
Assignment2	Week 10	50%



Thank You