MIDOCEAN UNIVERSITY

Practical Machine Learning and Data Exploration

Lecture Intro

Prepared by
Dr. Hager Saleh

Who am I?

MIDOCEAN UNIVERSITY

Hager Saleh Mohammed Ahmed

About me

Postdoc

Insight SFI Research Centre for Data Analytics, School of Engineering, University of Galway, Ireland.

Assistant Professor

Faculty of Computers and Artificial Intelligence, South Valley University, Hurghada, Egypt.

Senior Machine Learning Engineer

My expertise includes programming languages such as Python and artificial intelligence, big data, data science, software development, database management, image processing, natural language processing, streaming data analytics, and time series.

Educational Degrees

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

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

M. Sc. From Aswan University

Bachelor from Faculty of Computer and Information, Assiut University.

Objectives

Course Objectives Summary:

1. **Installing the main environment for the practical application**
2. **Understand the role and importance of 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.
3. **Machine Learning (ML):**
   - Understand the machine learning workflow.
   - Train supervised models
   - Evaluate model performance using accuracy, precision, and other metrics.
   - Optimize hyperparameters using cross-validation and grid search.
4. **Apply feature engineering techniques.**
5. **Apply knowledge through hands-on real-world projects.**

Exploratory Data Analysis (EDA)

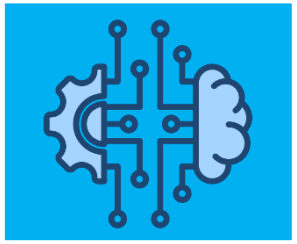
Key aspects of EDA include:

•**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 Leaning 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).

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
Assignment	Week 4	30%
Assignment	Week 7	30%
Project	Week 10	40%

MIDOCEAN UNIVERSITY

Thank You