# COMPLEX ENGINEERING PROBLEM- MACHINE LEARNING

INSTRUCTOR: Dr. Hammad Afzal
COURSE CODE: CS-471
SEMESTER: 7th (Fall, 2023)
Credit Hours: 03 (Theory)+ 1 (Lab)

Semester project is designed in a way to able students to solve the complex engineering problem. Following characteristics of complex engineering problem are targeted in this semester project of Machine Learning:

| **WP-1**<br>**Depth of Knowledge Required** | | **WP3 Depth of Analysis Required** | **WP-4 Familiarity of Issues** |
|---|---|---|---|
| CLO | WA/PLO[1] | ***Definition: Have no obvious solution and require abstract thinking, originality in analysis to formulate suitable models.*** | ***Definition: Involve infrequently encountered issues.*** |
| CLO-2<br>CLO-3 | PLO-4<br>PLO-5 | Exploratory Data Analysis<br>Techniques to show the initial insights into different variables | What types of attributes are more suitable to be used in ML models |

**Note:**

- **Total Number of WPs are 7. For any particular CEP, WP-1 is mandatory, and Some or All of WP-2 to WP-7 can be used.**

## Objective:

The objective of this assignment is to develop a machine learning model to predict employee attrition in a company based on a provided dataset. Students will work on a binary classification task, and the assignment includes preprocessing, exploratory data analysis (EDA), model selection, and evaluation.

## Dataset:

The dataset contains various features related to employees in a software engineering company, and the target variable is "Attrition," indicating whether an employee has left the company (1) or is still with the company (0).
Dataset is available at:
https://github.com/nelson-wu/employee-attrition-ml/blob/master/WA_Fn-UseC_-HR-Employee-Attrition.csv

## Tasks:

### Data Preprocessing:

- Handle missing data by either imputation or removal of incomplete records.
- Encode categorical variables appropriately.
- Scale numerical features to ensure uniformity.

[1] Mapping of CLO to WA/PLO from First Day Handout

- Explore and handle any outliers or anomalies in the data.

## Exploratory Data Analysis (EDA):

- Conduct a thorough analysis of the dataset to understand its characteristics.
- Visualize the distribution of employees with respect to different features.
- Analyze the correlation between various features and attrition.
- Identify patterns or trends that may influence employee attrition.

## Attrition Prediction:

- Split the dataset into training and testing sets.
- Choose an appropriate machine learning algorithm for binary classification (e.g., Logistic Regression, Random Forest, or Decision Tree).
- Train the model on the training set.
- Evaluate the model on the testing set and report performance metrics such as accuracy, precision, recall, and F1 score.

## Model Improvement (Optional):

- Experiment with different hyperparameters or feature selection techniques to improve model performance.
- Consider using techniques such as cross-validation to fine-tune the model.

## Documentation:

- Create a single presentation slide (like a poster) (i) Document your approach, including the preprocessing steps, choice of model, and any adjustments made. (ii) summarizing your findings, challenges faced, and the overall performance of your machine learning model.

## Note:

- Feel free to use any popular machine learning libraries (e.g., scikit-learn, TensorFlow, PyTorch) and visualization tools (e.g., Matplotlib, Seaborn) in your implementation. Ensure that your code is well-commented and your documentation is clear and organized.

## Deliverables:

| **Items** | **CLOs** | **Marks** |
|---|---|---|
| Jupyter Notebook or Python script containing the code. | CLO2, CLO3 | 10 |
| Preprocessing, EDA, Visualization | CLO2, CLO3 | 40 |
| Prediction, Cross Validation | CLO2, CLO1 | 30 |
| A presentation slide deck summarizing key findings and insights. | CLO3 | 10 |
| Creativity and innovation in addressing challenges. | CLO1, CLO2 | 10 |