CSE443 – CE471 AI For Intelligent Built Environment Systems
Fall 2025 - Assignment#1
(Due date: 01/12/2025 Monday, at 23:59)

Urban Building Energy: Preprocessing + ANN for Simple_Building_Energy_Rating

The goal of this assignment is to develop a building energy rating prediction system using machine learning on the urban building dataset. The task involves preprocessing a unified building energy dataset, applying appropriate feature engineering techniques, and implementing an Artificial Neural Network (ANN) model to predict Simple_Building_Energy_Rating. Additional preprocessing steps such as categorical encoding, feature scaling, and handling missing values will be required to prepare the dataset for ANN training.

- Prepare a detailed report.
- Examine the manuscript in [1], prepare a flow similar to the illustration in Figure 2.
- Draw a table of accuracy, precision, recall, F1 results of each class.
- Represent the confusion matrix, training/validation loss and training/validation accuracy graphs as in [2].
- Show the validation and test results by commenting on them as in [2].
- Calculate Cohen's Kappa score and print as output as in [2].
- Draw precision recall and AUC-ROC curve as in [2].

Dataset: *Urban building energy performance prediction and retrofit analysis using data-driven machine learning approach* - Mendeley Data [3]. The only output for the model is Simple_Building_Energy_Rating, ignore other outputs.

Students are allowed to use high-level libraries. Recommended list:
- Data handling & numeric:
  - pandas, numpy
- Visualization & EDA:
  - matplotlib, seaborn
- Preprocessing & utilities:
  - scikit-learn (sklearn) — SimpleImputer, StandardScaler/MinMaxScaler, OneHotEncoder, ColumnTransformer, Pipeline, train_test_split, StandardScaler
  - category_encoders (optional) for target/ordinal encoding
  - imbalanced-learn (if classes are imbalanced and classification is used)
- Metrics:
  - sklearn.metrics (mean_absolute_error, mean_squared_error, r2_score, accuracy_score, classification_report if classification)

Submission Details
- ❖ Submit the source code. Submit the report.
- ❖ This assignment effects overall 40% of your final grade.
- ❖ Source codes and the report must be compressed as a **single zip** file and must be submitted via **Google Classroom Group**.
- ❖ Zip file must be named with all group members' student numbers (example→ 2007900011_2007900012.zip).
- ❖ All group members' names must exist in the report file.
- ❖ One of the group members may submit the assignment.
- ❖ In any forms of copying and cheating all parties will get zero grade.

Good Luck!

[1] https://www.sciencedirect.com/science/article/pii/S0378778823009982

[2] https://github.com/naomifridman/Neural-Network-Churn-Prediction

[3] https://data.mendeley.com/datasets/m6vv9k9gcd/5