**Walchand College of Engineering, Sangli**

## **Machine Learning Lab (6CS372)**

**TY BTech | AY 2023-2024 | Even Sem**

**Assignment 4**

1. **Linear regression: Single feature vs multiple features**
2. Download dataset as per your batch.
3. Preprocessing: Null value handling, standardization, replace categorical values with numeric values (e.g. 0, 1, 2 etc.)
4. Data splitting: Split data as 70% train and 30% test using train\_test\_split function.
5. Fit model using fit function taking a single feature at a time and all independent features at a time.
6. Report parameter values, training error and test error and model accuracy for Linear regression with single feature and multiple features.
7. **Answer following questions (include question and answer as markdown cell in your notebook)**
   1. Provide a general multiple linear regression equation and explain all the terms.
   2. Explain the concept of a dummy variable and how such variables are calculated. Why is it necessary to convert nominal variables to dummy variables when performing linear regression?
   3. Explore and mention assumptions in linear regression with suitable explanation.

**Datasets:**

**Batch 1 -** <https://www.kaggle.com/datasets/spittman1248/cdc-data-nutrition-physical-activity-obesity>

**Batch 2-** <https://archive.ics.uci.edu/ml/datasets/Air+Quality>

**Batch 3-** <https://archive.ics.uci.edu/ml/datasets/Appliances+energy+prediction>

**Batch 4-** <https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset>

**Batch 5-** <https://archive.ics.uci.edu/dataset/186/wine+quality>

**Batch 6-** <https://archive.ics.uci.edu/dataset/29/computer+hardware>

**Batch 7-** <https://archive.ics.uci.edu/dataset/477/real+estate+valuation+data+set>

**Batch 8-** <https://archive.ics.uci.edu/dataset/162/forest+fires>