

# Homework – Week 6: Hyperparameter Tuning

Course: Introduction to Machine Learning

## Objective

Explore the impact of hyperparameter tuning using cross-validation and search techniques. Learn how to prevent data leakage and improve model performance through informed tuning.

## Instructions

1. Choose a classification dataset:
  - You may use a built-in dataset from ‘sklearn.datasets’, or one from Kaggle.
  - Make sure your dataset has a clear target variable and enough features.
2. Preprocess the data:
  - Handle missing values, if any.
  - Apply scaling if necessary (e.g., for SVM or Logistic Regression).
  - Use train test split and ensure no data leakage occurs.
3. Select one model to tune, such as:
  - LogisticRegression
  - RandomForestClassifier
  - SVC (Support Vector Classifier)
4. Try at least two different hyperparameter search strategies:
  - Manual Search
  - Grid Search or Randomized Search
5. For each search method:
  - Show the best hyperparameters found
  - Report the cross-validation accuracy
  - Evaluate the final model on the test set

6. (Bonus) Explore a third method using Bayesian optimization (e.g., with ‘optuna’ or ‘skopt’).
7. Submit a Jupyter Notebook or a PDF including:
  - The dataset used (name and link if from Kaggle)
  - All your code and outputs
  - Brief explanations and comparisons of results
  - A reflection: what did you learn about tuning?

## Tips

- Use ‘GridSearchCV’ and ‘RandomizedSearchCV’ from ‘sklearn.model selection’.
- Use pipelines to avoid data leakage when preprocessing.
- You can use cross val score to evaluate intermediate models.