



DIGITAL[®]
EMPOWERMENT
NETWORK

Digital Empowerment Network

Artificial Intelligence

Week 04

**Feature Selection and Classification with and
without Optimization**

Mentor: Hussain Shoaib

Feature Selection and Classification with and without Optimization

Objective:

To train a classification model using **selected features only**, and compare performance **before and after feature selection**, explaining the difference in results.

Step 1: Dataset Selection

Choose one of the following CSV type datasets:

- Parkinson's Disease Classification Dataset
- Breast Cancer Wisconsin Dataset
- Heart Disease Dataset
- Any dataset with at least 15+ features and 500+ rows

Step 2: Data Preparation

- Load dataset and inspect
- Handle missing values
- Encode categorical variables
- Scale numerical features
- Split into train/test sets (80/20)

Step 3: Feature Selection

Apply at least **one** of the following techniques, then select the best set of features for model training:

- **Filter method** — SelectKBest with Chi-Square or ANOVA
- **Wrapper method** — Recursive Feature Elimination (RFE)
- **Embedded method** — Feature importance from Random Forest or Lasso Regression

Step 4: Model Training

Part 1 — Without Feature selection

Train at least two classifiers **using all the features in data**:

Part 2 — With Feature selection

Train the classifier using selected features In feature selection step.

- Logistic Regression
- Random Forest Classifier

- Use default parameters (no tuning)

Step 5: Model Evaluation and Comparison

- Evaluate both models (before and after optimization) using:
- Accuracy, Precision, Recall, F1-score
- Confusion Matrix

Write a **short explanation** of how optimization changed the model performance and why

Step 6: (Optional) Web App Deployment Using Streamlit

If you want to deploy your model:

- Create a Streamlit app where users can input feature values
- Predict class based on **selected features**
- Toggle between **Without Optimization** and **With Optimization** models
- Show probability scores for each class

Deliverables:

- Jupyter Notebook with feature selection, model training, optimization, and evaluation
- Two saved trained model files (.pkl)
- *(Optional)* Streamlit app file (app.py)
- Screenshots of predictions from both models
- Written comparison of results with and without optimization

Deadline:

Submit within **7 days**

Tools to Use:

- Python
- Pandas, NumPy, scikit-learn
- Matplotlib, Seaborn
- *(Optional)* Streamlit