Assignment 5

Stroke Prediction using SVM Classifier

Enhance your previous **Stroke Prediction Dataset notebook** from **Assignment 2** by adding further analysis, preprocessing improvements, and applying **Support Vector Machine (SVM) classifier** with different configurations to determine the best performing model.

Dataset

Use the **same cleaned dataset you prepared in Assignment 2.**

Requirements:

1. Additional Data Analysis & Preprocessing

- Add at least two new analyses (e.g., correlation heatmaps, outlier detection, feature importance analysis).
- Apply any further preprocessing steps that could improve SVM results (feature scaling, transformations, class balancing, etc.).
- Justify why you made each preprocessing choice.

2. Modeling and Evaluation

- Implement **multiple types of SVM**:
    - **Hard Margin SVM**
    - **Soft Margin SVM** (tuning C parameter).
    - **Kernel SVMs**:
        - Linear kernel.
        - RBF kernel (at different values of gamma).
        - Polynomial kernel (at different values of degree of polynomial).
        - Any other kernel you want to try (e.g., Sigmoid).

3. Decision Boundary Visualization

- For at least **two kernels**, reduce the feature space to **2D using PCA** (or select two best features) and **plot decision boundaries**.
- Interpret the plots:
    - How does the boundary shape differ for each kernel?

- Does it align well with your data distribution?

4. Model Comparison and Interpretation

- Compare all tested SVM models in a **summary table** (metrics + best parameters).
- State **which SVM configuration performed best** and explain **why** based on:
  - The data characteristics.
  - The kernel's mathematical properties.
  - Overfitting/underfitting observations.
- Discuss **the limitations** of SVM for this dataset and possible improvements.

Instructions:

- All codes should be written **independently without using AI tools**.
- Submit your notebook by **uploading it to Kaggle** and sharing the public notebook link.
- Inside your GitHub repository named **IEEE-ML-2025**, add a folder named **Stroke_Prediction** which must contain:
  - Your **.ipynb** notebook.
  - A **README.md** file specific to the project.
  - A **plots/ folder** to store relevant visualizations and metrics.

Expected Deliverables:

- **Your GitHub repository link**
- **Your public Kaggle notebook link** demonstrates your full implementation.

Submission Deadline: Friday 22/8/2025 before 11:59 PM