**Support Vector Machines (SVM) - Report**

**Objective :**Use Support Vector Machines (SVMs) for binary classification with both linear and non-linear (RBF) kernels.

**Tools & Libraries Used**

* Python
* Scikit-learn
* Pandas
* NumPy
* Matplotlib
* PCA (for 2D visualization)

**Dataset Description**

The dataset used is breast-cancer.csv, which contains 30 numerical features extracted from digitized images of a breast mass. The target variable is diagnosis where:

* M = Malignant (1)
* B = Benign (0)

The dataset was preprocessed by:

* Dropping the id column
* Encoding the diagnosis into binary values (M → 1, B → 0)
* Standardizing feature values

**Tasks Performed**

1. **Data Preprocessing**
2. **Dimensionality Reduction** using PCA for visualization
3. **Model Training** using:
   * Linear Kernel SVM
   * RBF Kernel SVM
4. **Decision Boundary Visualization**
5. **Hyperparameter Tuning** (GridSearchCV on C and gamma)
6. **Evaluation** using:
   * Classification report
   * 5-fold cross-validation

**Results**

* The **RBF Kernel SVM** outperformed the Linear Kernel due to better handling of non-linear decision boundaries.
* GridSearchCV revealed the optimal parameters:
  + C (Regularization strength)
  + gamma (Influence of single training examples)
* Final model achieved **~98-99% accuracy** in cross-validation.

**What I Learned**

* **Margin Maximization**: SVM aims to find the optimal hyperplane that maximizes the margin between two classes.
* **Kernel Trick**: The RBF kernel allows SVM to handle non-linearly separable data by mapping to a higher-dimensional space.
* **Hyperparameter Tuning**: Learned how C and gamma impact model complexity and generalization.
* **Cross-validation**: Used to verify model robustness and reduce overfitting.
* **Visualization**: PCA helped interpret model behavior visually in 2D space.