PES UNIVERSITY

ML's LAB

Section:- F

NAME :- MOHIT KUMAR

5TH SEM / F-SECTION

SRN :-PES2UG23CS350

Analysis Questions

Moons Dataset Questions (2 questions):-

(1) Inferences about the Linear Kernel's performance.

The Linear Kernel shows limited effectiveness on the Moons dataset due to its inability to model non-linear relationships. Since the data forms two intertwined half-moon clusters, a straight decision boundary cannot capture the curved separation needed. This leads to poor accuracy and numerous misclassifications, especially in the central overlapping region. The model's simplicity becomes a drawback here, confirming that linear separation fails when data lacks linear separability.

(2) Comparison between RBF and Polynomial kernel decision boundaries.

Both RBF and Polynomial kernels handle the non-linear structure of the Moons dataset, but RBF performs more naturally. It creates smooth, adaptive boundaries—often circular or wave-like—that closely follow the data's curvature. The Polynomial kernel also generates non-linear boundaries but depends heavily on the degree parameter; if too low, it underfits, and if too high, it may overfit by creating unnecessarily complex curves. RBF's flexibility makes it more robust without requiring degree tuning.

Banknote Dataset Questions (2 questions):-

(1) Which kernel was most effective for this dataset?

The RBF kernel typically delivers the best performance on the Banknote dataset. It achieves high accuracy and F1-score by effectively capturing subtle, non-linear patterns in the feature space (e.g., variance vs. skewness). Its ability to model local relationships using radial distance makes it well-suited for this real-world classification task, where class boundaries are not perfectly linear but still structured.

(2) Why might the Polynomial kernel have underperformed here?

The Polynomial kernel may underperform due to sensitivity to the degree parameter. If the degree is too low, the model lacks the complexity to separate classes effectively. If too high, it risks overfitting by capturing noise rather than the true pattern. Unlike RBF, which adapts via the gamma parameter, Polynomial requires precise degree selection, making it less reliable without extensive hyperparameter tuning.

Hard vs. Soft Margin Questions (4 questions):-

(1) Which margin (soft or hard) is wider?

The soft margin (C=0.1) produces a wider separation between classes. A smaller C value reduces the penalty for misclassification, allowing the model to prioritize maximizing the margin over perfect classification. This results in a broader decision boundary that generalizes better.

(2) Why does the soft margin model allow "mistakes"?

The soft margin model permits some misclassifications to avoid overfitting and improve robustness. By tolerating errors—especially from outliers or noisy data—it focuses on finding a balanced, generalized solution. This trade-off, controlled by C, enhances performance on unseen data.

(3) Which model is more likely to be overfitting and why?

The hard margin model (C=100) is more prone to overfitting. With high C, the model enforces strict classification of all training points, leading to a narrow, complex boundary that fits noise. This reduces its ability to generalize to new data, especially in real-world scenarios with variability.

(4) Which model would you trust more for new data and why?

The soft margin model (C=0.1) is more reliable for new data. It emphasizes generalization through a wider margin, making it resilient to noise and minor variations. In practical applications, starting with a soft margin is preferred, as it balances accuracy and robustness, reducing overfitting risk.

SCREENSHOTS

1. Training Results:-

(a) Moons dataset:-

```
SVM with LINEAR Kernel <PES2UG23CS350>
              precision    recall  f1-score   support

           0       0.85      0.89      0.87        75
           1       0.89      0.84      0.86        75

    accuracy                           0.87       150
   macro avg       0.87      0.87      0.87       150
weighted avg       0.87      0.87      0.87       150
```

```
SVM with RBF Kernel <PES2UG23CS350>
              precision    recall  f1-score   support

           0       0.95      1.00      0.97        75
           1       1.00      0.95      0.97        75

    accuracy                           0.97       150
   macro avg       0.97      0.97      0.97       150
weighted avg       0.97      0.97      0.97       150
```

```
SVM with POLY Kernel <PES2UG23CS350>
              precision    recall  f1-score   support

           0       0.85      0.95      0.89        75
           1       0.94      0.83      0.88        75

    accuracy                           0.89       150
   macro avg       0.89      0.89      0.89       150
weighted avg       0.89      0.89      0.89       150
```

(b)Banknote Authentication:-

```
SVM with LINEAR Kernel <PES2UG23CS350>
              precision    recall  f1-score   support

      Forged       0.90      0.88      0.89       229
     Genuine       0.86      0.88      0.87       183

    accuracy                           0.88       412
   macro avg       0.88      0.88      0.88       412
weighted avg       0.88      0.88      0.88       412
```

SVM with RBF Kernel <PES2UG23CS350>

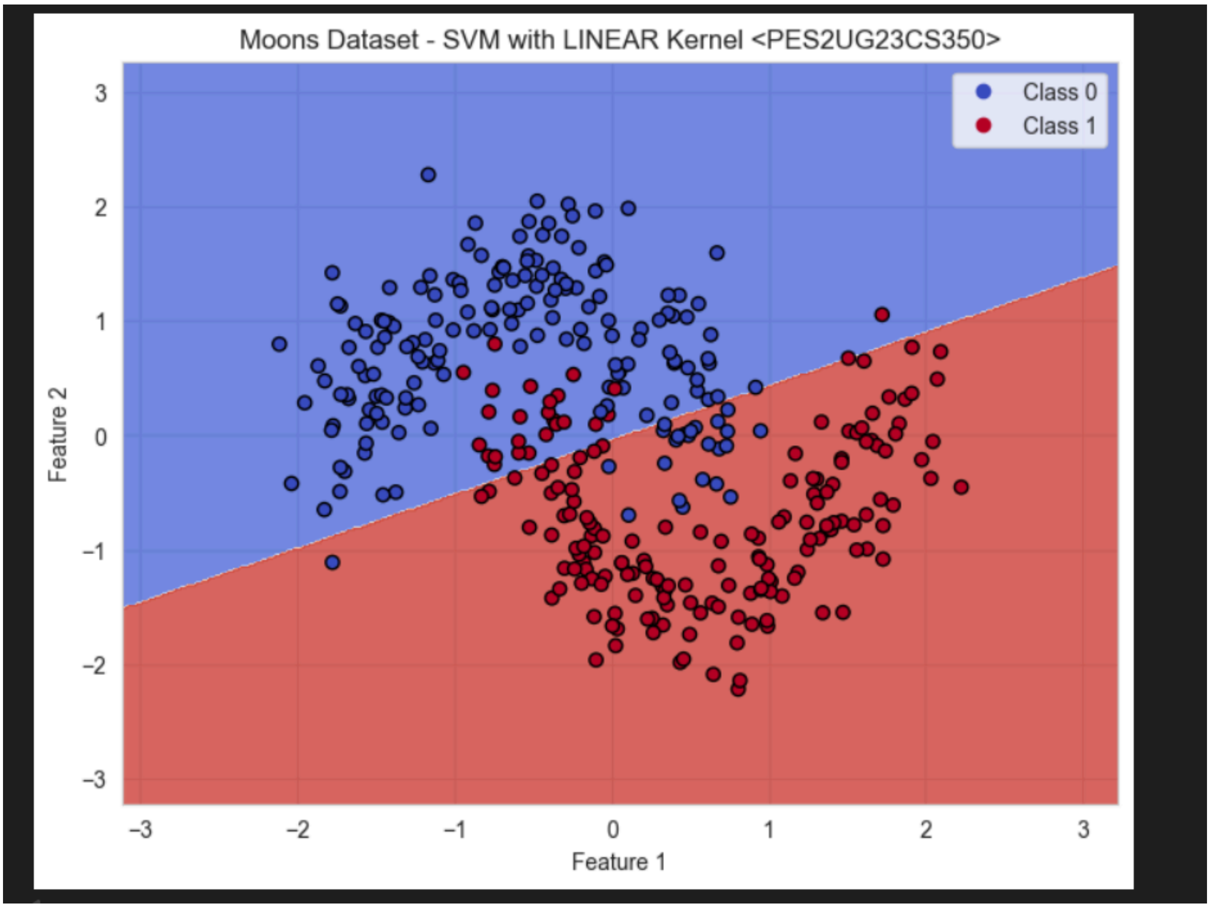
	precision	recall	f1-score	support
Forged	0.96	0.91	0.94	229
Genuine	0.90	0.96	0.93	183
accuracy			0.93	412
macro avg	0.93	0.93	0.93	412
weighted avg	0.93	0.93	0.93	412

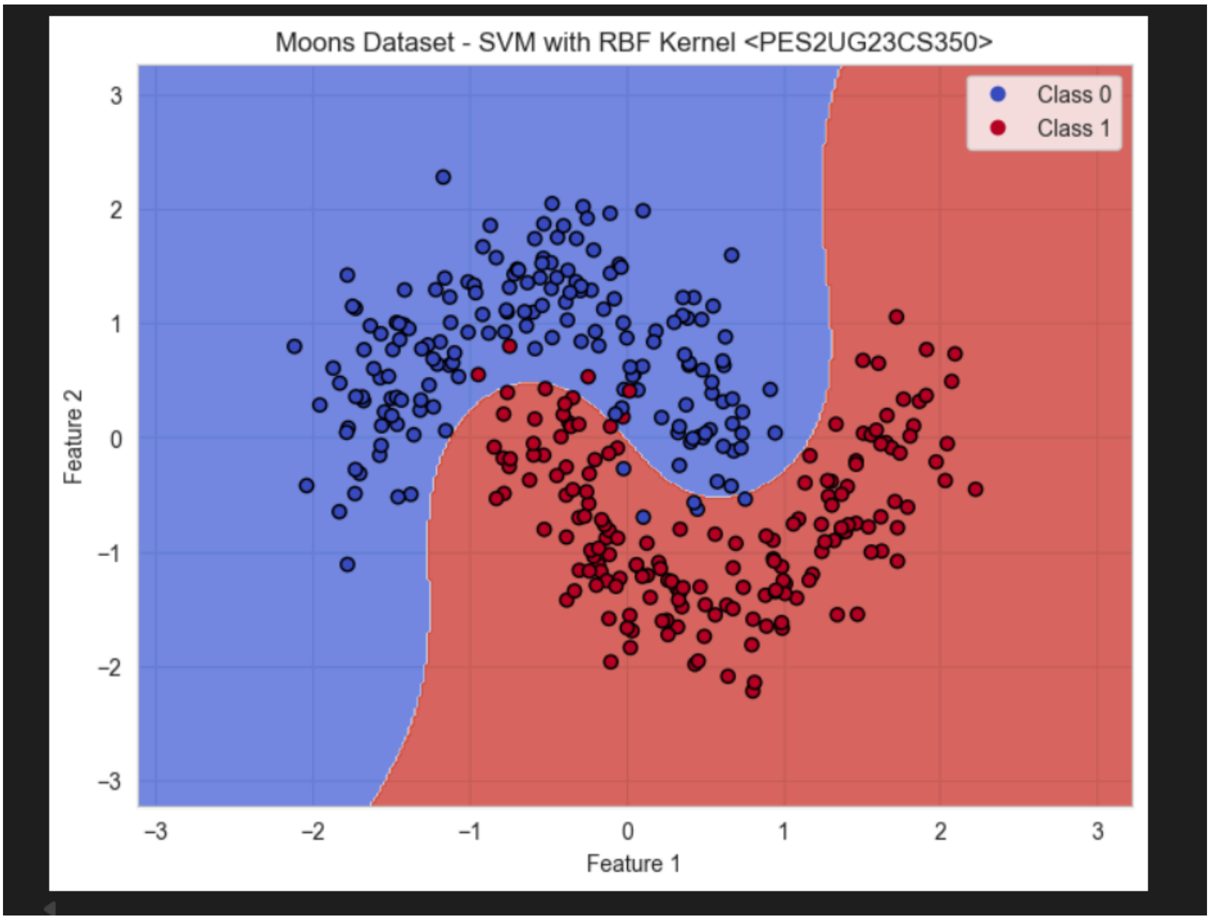
SVM with POLY Kernel <PES2UG23CS350>

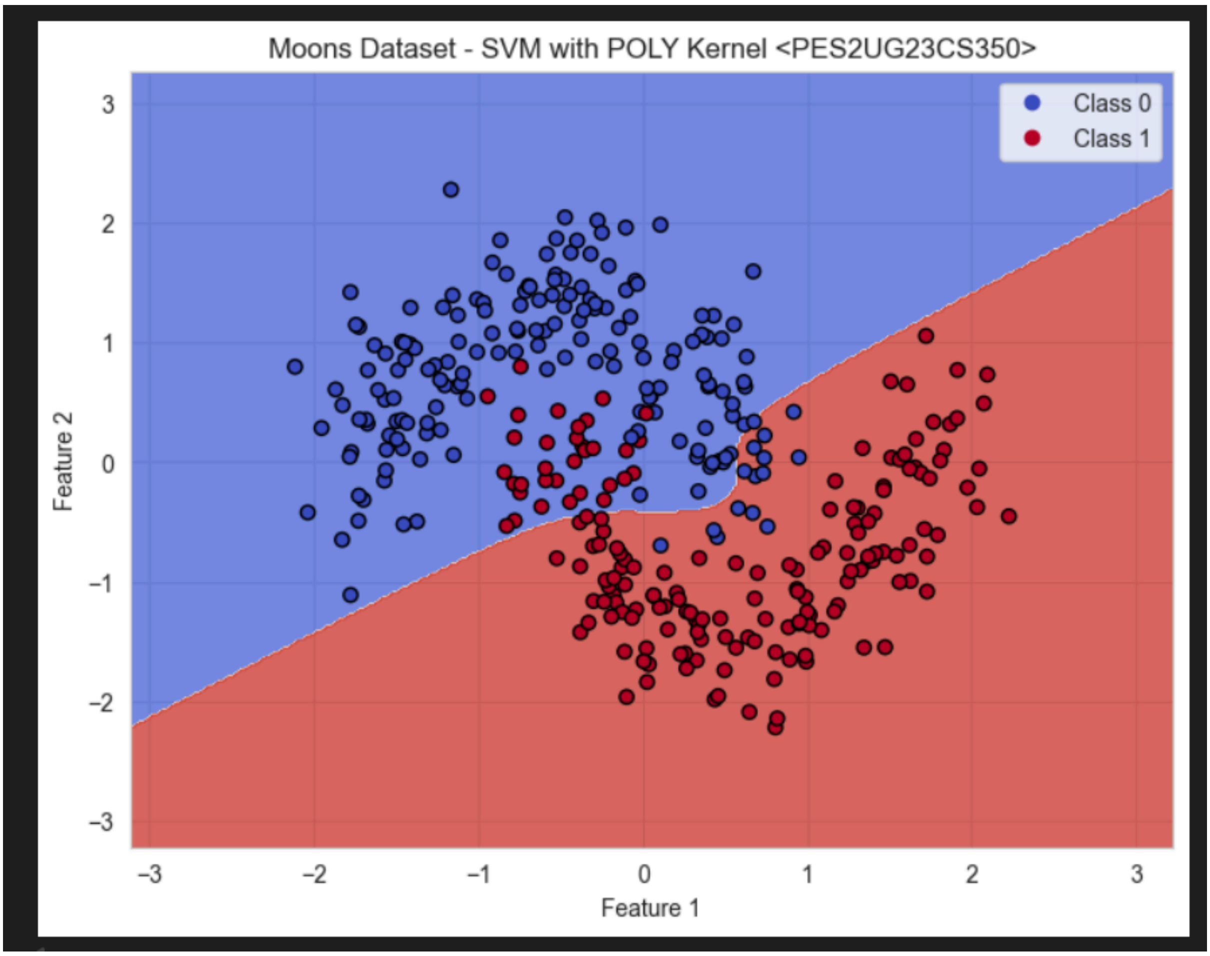
	precision	recall	f1-score	support
Forged	0.58	0.98	0.73	229
Genuine	0.84	0.11	0.20	183
accuracy			0.60	412
macro avg	0.71	0.55	0.47	412
weighted avg	0.70	0.60	0.50	412

2. Decision Boundary Visualizations:-

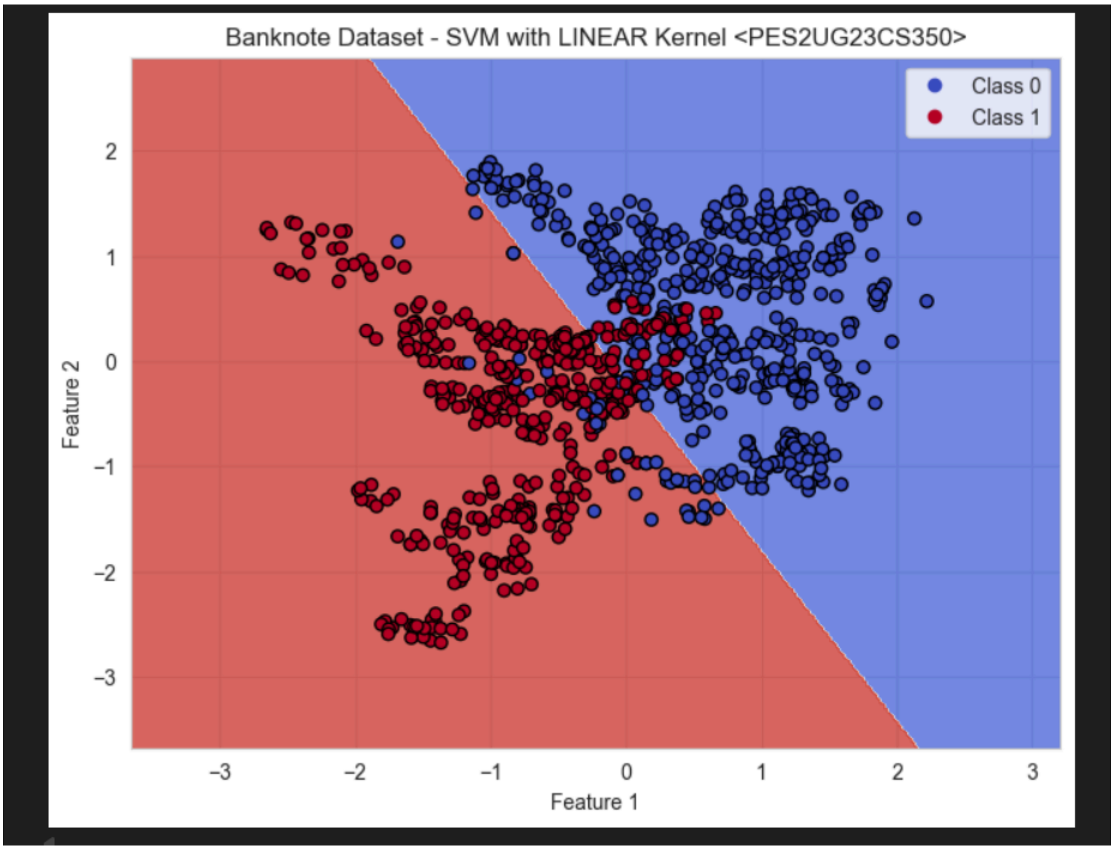
(a) Moons Dataset:-

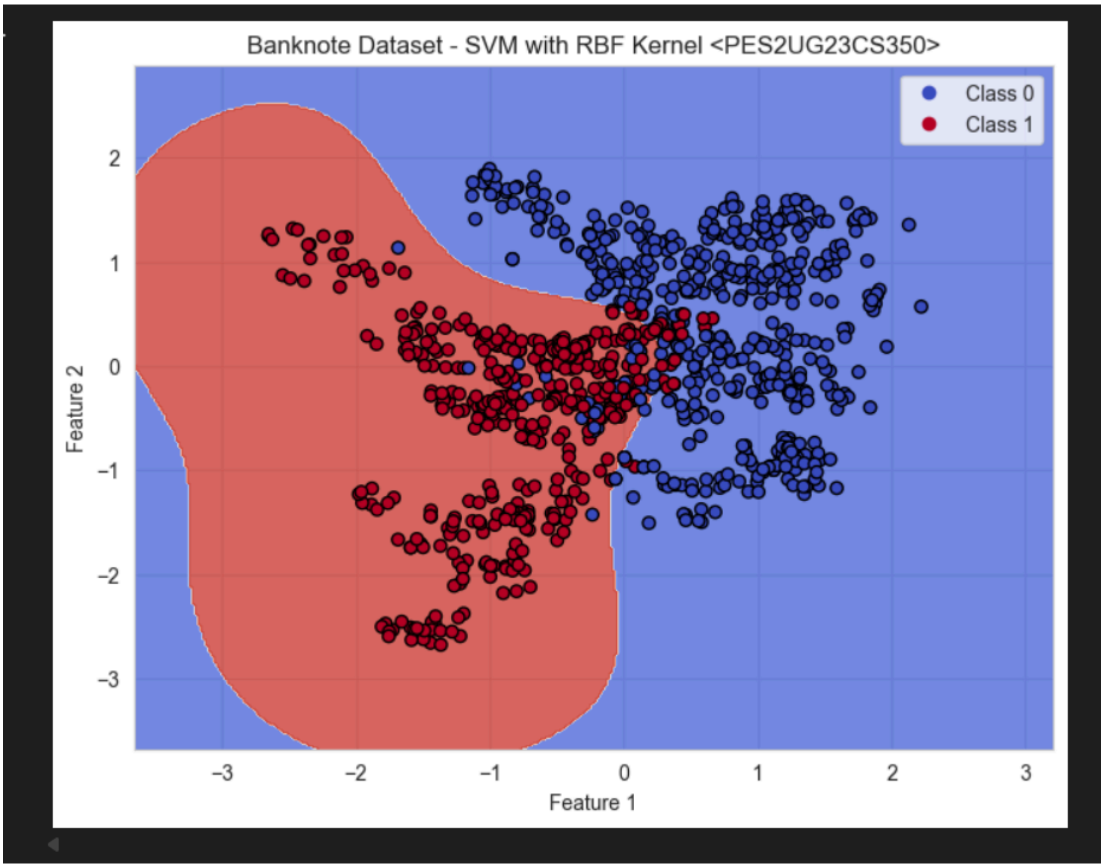


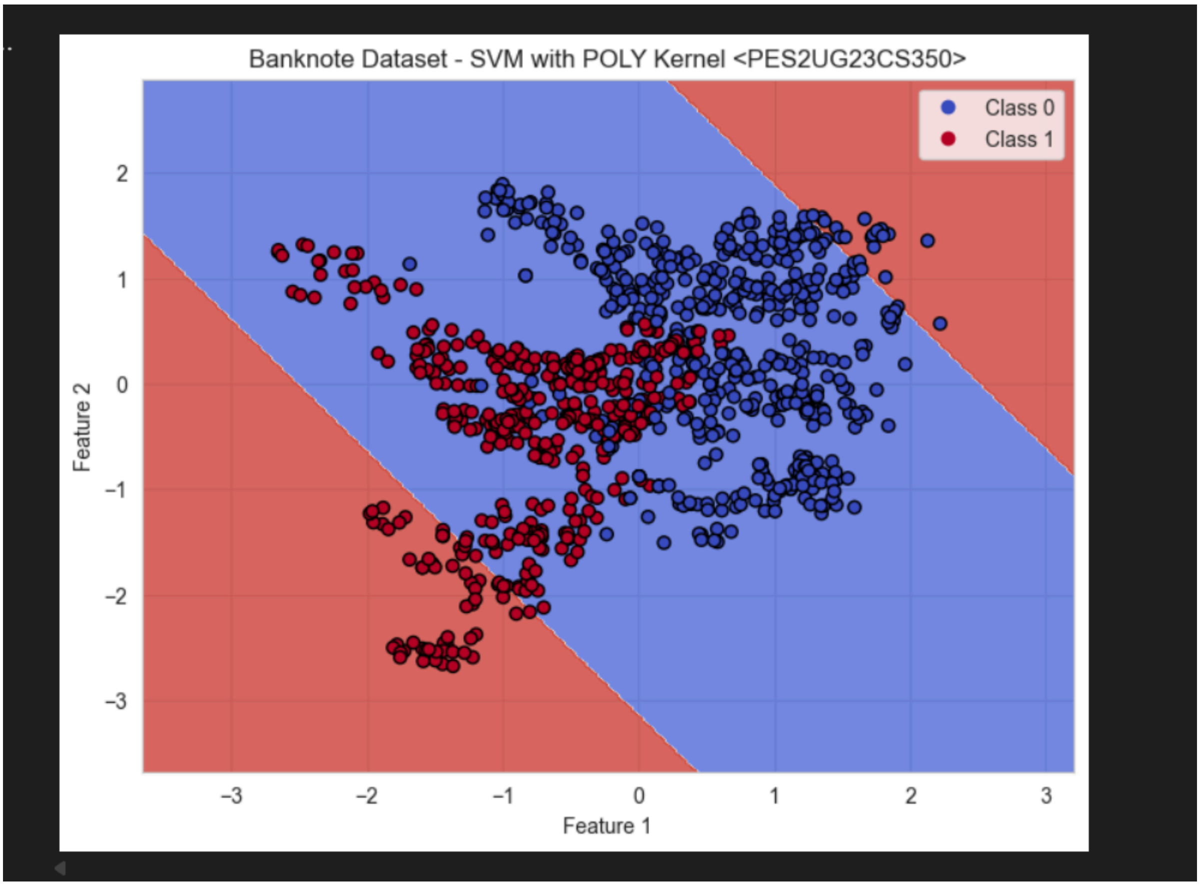

Moons Dataset - SVM with RBF Kernel <PES2UG23CS350>
Class 0
Class 1
Feature 1
Feature 2
−3
−2
−1
0
1
2
3


Moons Dataset - SVM with POLY Kernel <PES2UG23CS350>
Class 0
Class 1
Feature 1
Feature 2
−3
−2
−1
0
1
2
3

(b) Banknote Authentication:-




Banknote Dataset - SVM with RBF Kernel <PES2UG23CS350>
Class 0
Class 1
Feature 2
Feature 1
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3


Banknote Dataset - SVM with POLY Kernel <PES2UG23CS350>
Class 0
Class 1
Feature 2
Feature 1
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

(c) Hard vs Soft Margin:-