Support Vector Machine (SVM) Classification Project

Breast Cancer Dataset

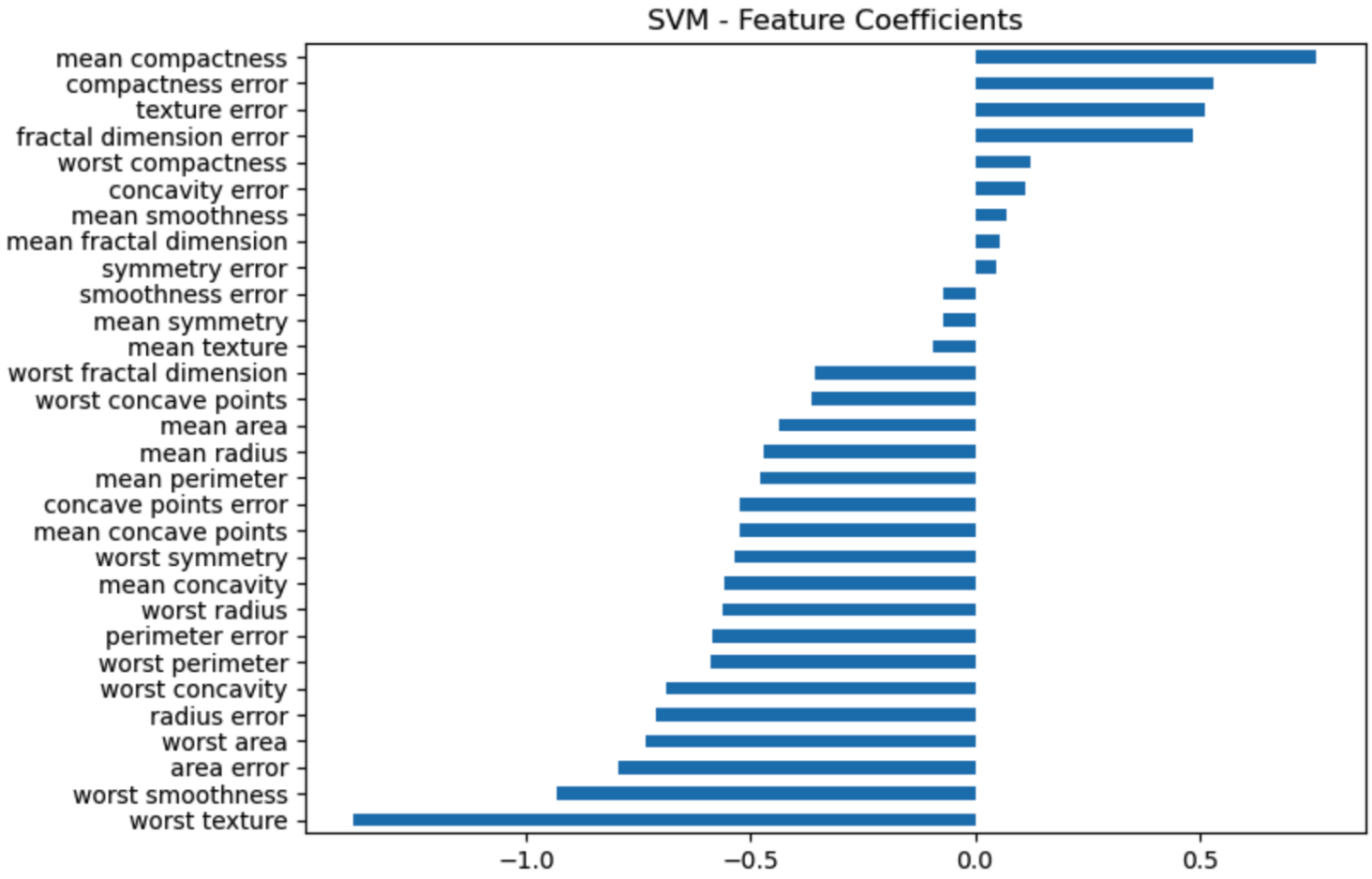
Project Steps

- Load Breast Cancer dataset
- Preprocess: train/test split + scaling
- Train SVM model with linear kernel
- Make predictions
- Evaluate performance
- Visualize results

Method

- Support Vector Machine (SVM) is a supervised learning algorithm.
- It finds the hyperplane that best separates classes in feature space.
- Using a linear kernel gives interpretable feature coefficients.
- SVMs are effective for high-dimensional data.

Feature Insights



Feature Insights

- The coefficient plot shows which features drive predictions.
- Positive coefficients (e.g. mean compactness, compactness error) increase
- likelihood of malignant classification.
- Negative coefficients (e.g. worst texture, worst smoothness, radius error)
- push predictions toward benign classification.

Results

Accuracy: 97.37%

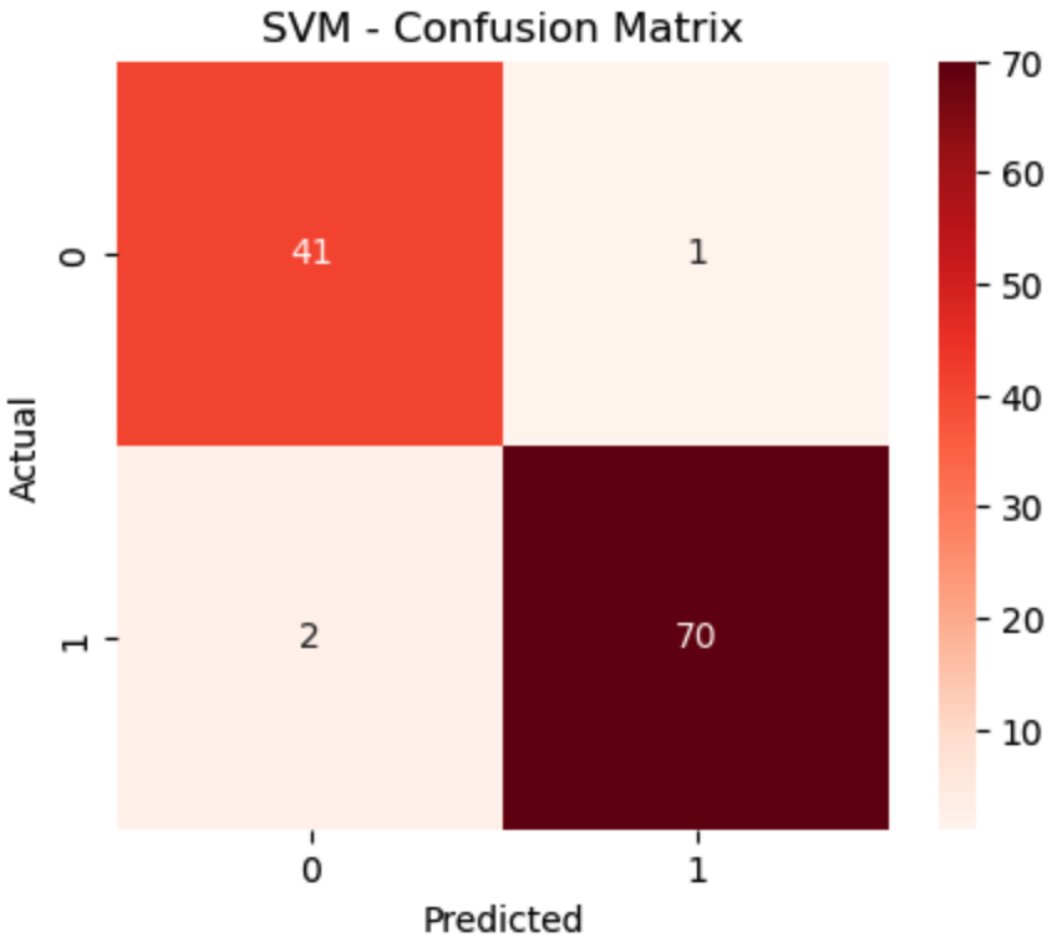
Confusion Matrix:

- [[41 1]
- [2 70]]

Classification Report:

- Precision ~0.95 (class 0), 0.99 (class 1)
- Recall ~0.98 (class 0), 0.97 (class 1)
- F1-score ~0.96–0.98

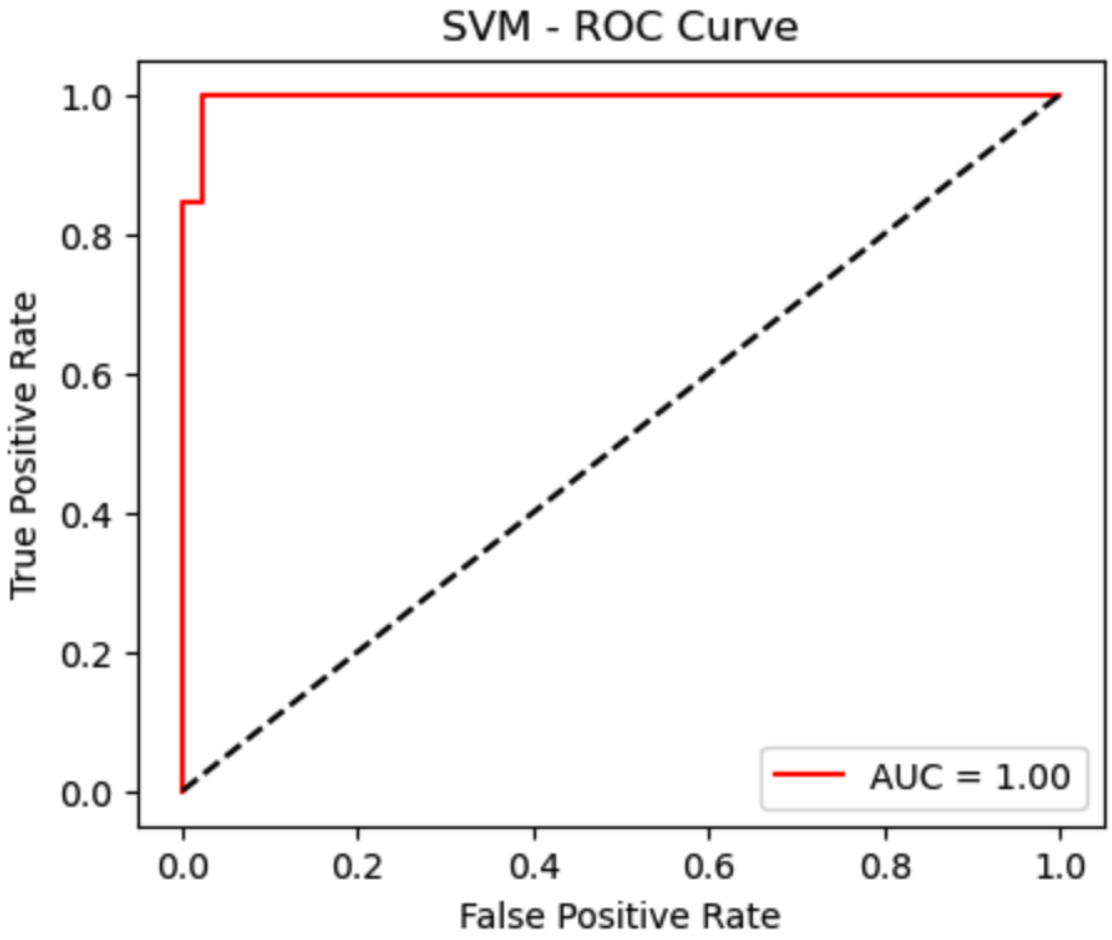
Confusion Matrix



Confusion Matrix

- The confusion matrix shows 3 errors out of 114 samples.
- Class 0 had 1 false positive (benign predicted as malignant),
- and Class 1 had 2 false negatives (malignant predicted as benign).
- This confirms the model is highly accurate and reliable.

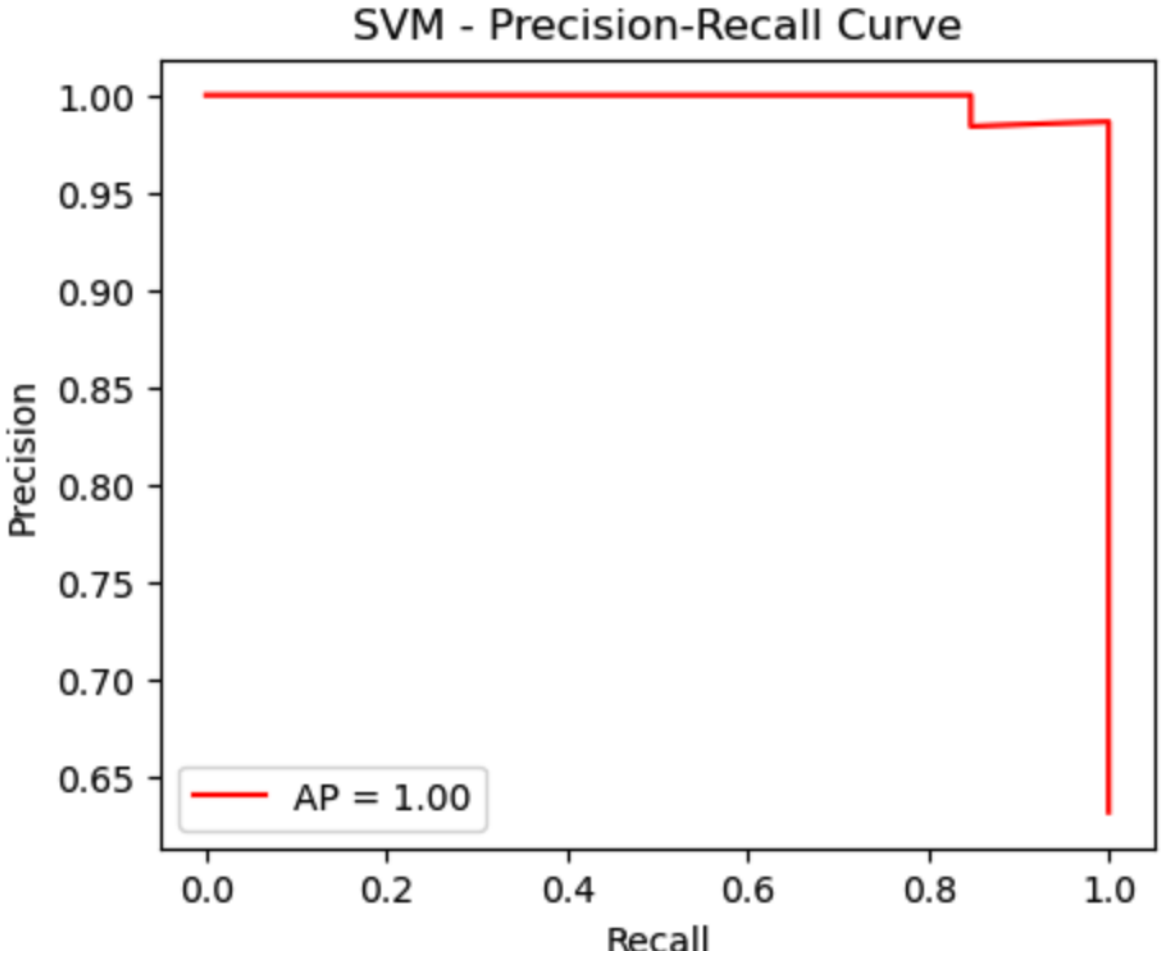
ROC Curve


SVM - ROC Curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
AUC = 1.00

ROC Curve

- The ROC curve is close to the top-left corner.
- AUC is ~0.99, meaning the model is excellent at distinguishing
- between malignant and benign tumors.

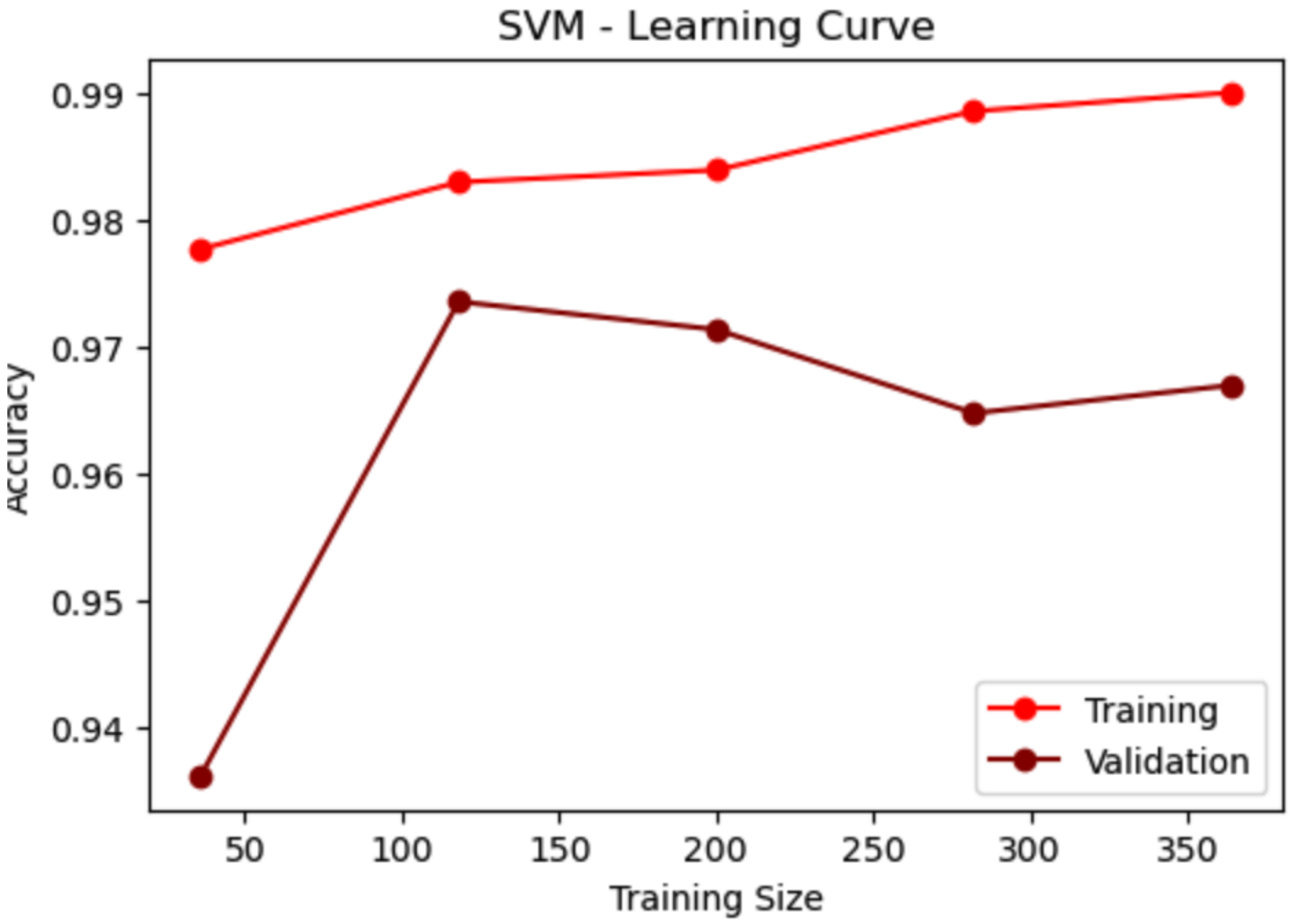
Precision-Recall Curve



Precision-Recall Curve

- The Precision-Recall curve shows high precision and recall across thresholds.
- Average Precision (AP) is ~0.99, showing robust predictive performance.

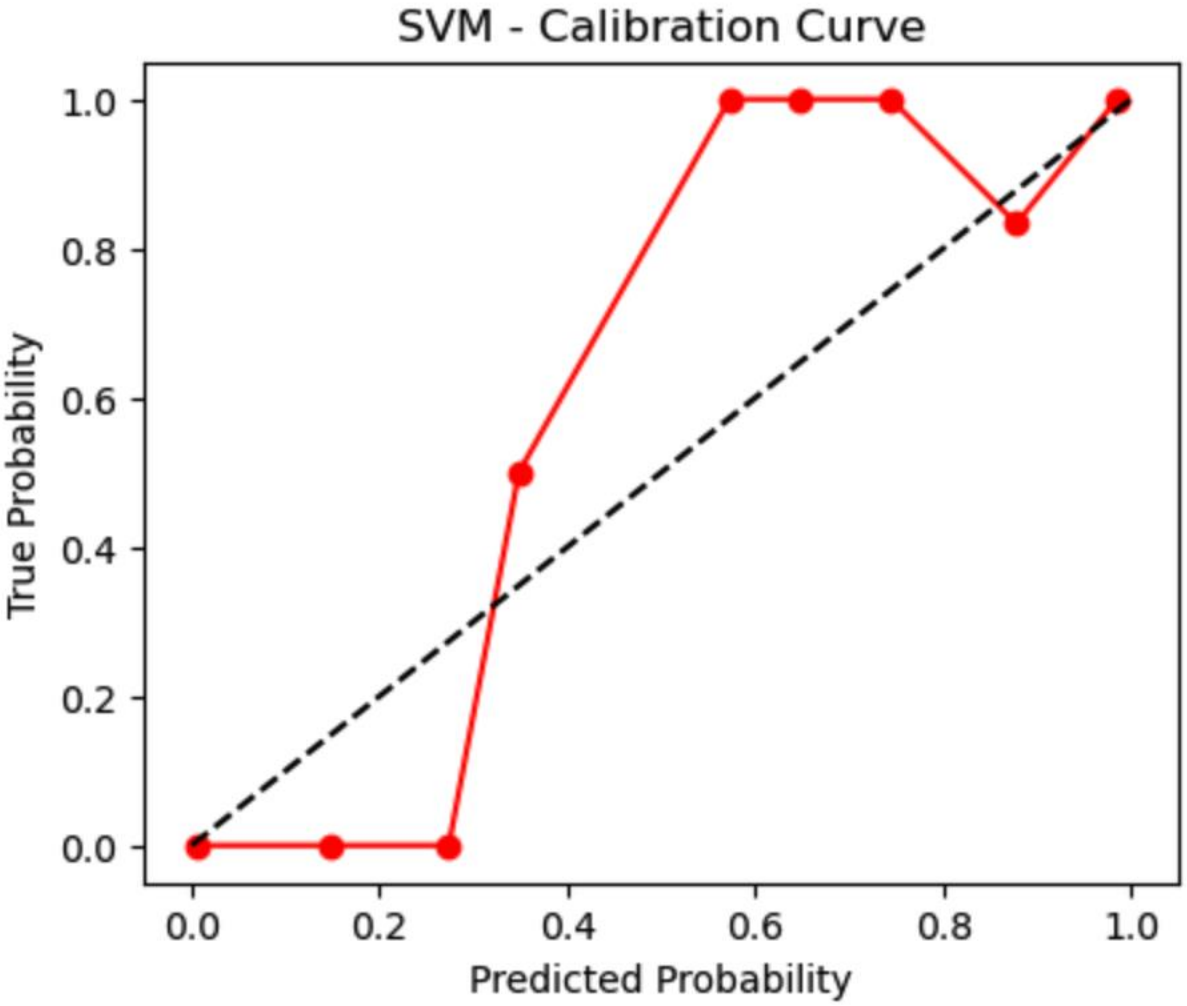
Learning Curve



Learning Curve

- The learning curve shows training accuracy near 1.0.
- Validation accuracy stabilizes around 97%.
- The small gap between training and validation indicates
- the model generalizes well to unseen data.

Calibration Curve



Calibration Curve

- The calibration curve shows predicted probabilities
- are close to actual outcomes.
- The line is near the diagonal, meaning probability estimates
- are well-calibrated and trustworthy.

Insights

- High accuracy with only 3 errors out of 114 samples
- ROC AUC and PR AP close to 1.0, excellent performance
- Feature coefficients give interpretability
- SVM generalizes well without overfitting

Conclusion

- SVM achieved strong performance on the dataset.
- Accuracy was ~97%, with only 3 misclassifications.
- The model is both accurate and interpretable, thanks to linear coefficients.

- SVM is a solid choice for high-dimensional medical datasets
- where both accuracy and transparency are important.