Support Vector Machine (SVM) Results on Bank Marketing Dataset

# Dataset & Preprocessing

- Target variable y mapped to binary (1 = subscribed, 0 = not subscribed).  
- One-hot encoding for categorical variables.  
- Standardization of numerical features.  
- Train-test split (80/20).

# Models Evaluated

1. Linear SVM (unweighted)  
2. RBF SVM  
3. Polynomial SVM  
4. Linear SVM with class weights (class\_weight='balanced')

# Results

## Linear SVM (Unweighted)

* Accuracy: 90.5%
* Recall (class 1): 0.34
* Observation: Strong majority-class performance, but poor recall for minority class (subscribers).

## RBF SVM

* Accuracy: 89.5%
* Recall (class 1): 0.21
* Observation: Lower recall than Linear SVM, weaker minority class detection.

## Polynomial SVM

* Accuracy: 89.5%
* Recall (class 1): 0.22
* Observation: Similar to RBF, still struggles with minority class.

## Linear SVM with Class Weights

* Accuracy: 85.8%
* Recall (class 1): 0.88 (huge improvement)
* Precision (class 1): 0.44
* F1-score (class 1): 0.59
* Observation: Accuracy dropped compared to unweighted models, but recall for subscribers increased dramatically (0.34 → 0.88). The weighted model catches almost all potential subscribers, though at the cost of more false positives.

# Key Insights

* Unweighted SVMs prioritize accuracy and majority class (non-subscribers), but fail to capture minority class effectively.
* Class weighting trades accuracy for recall, making the model much better at detecting the minority class (which is more valuable in marketing).
* This shift is often desirable: in marketing, it’s more acceptable to target some non-subscribers (false positives) than to miss actual subscribers (false negatives).

# Recommendations

* Linear SVM with class weights is preferable for this problem if the business objective is maximizing subscriber detection.
* For further improvement:
* - Try SMOTE (oversampling) or undersampling along with class weighting.
* - Explore tree-based models (Random Forest, XGBoost, LightGBM) which often handle imbalanced data more effectively.
* - Evaluate using precision-recall AUC instead of plain accuracy, since imbalance makes accuracy misleading.