Classifier Comparison Summary Report

Business Objective:

Predict whether a bank customer will subscribe to a term deposit (y = yes) based on client and bank-related features.

Models Evaluated:

1. Decision Tree
2. Logistic Regression
3. Support Vector Machine (SVM)
4. K-Nearest Neighbors (KNN, on smaller subset)

Best Hyperparameters:

- Decision Tree: criterion='gini', max_depth=3, min_samples_split=2
- Logistic Regression: C=0.01, penalty='l2', solver='lbfgs'
- SVM: C=1, kernel='linear', gamma='scale'
- KNN (1k train): n_neighbors=7, weights='uniform', metric='euclidean'

Test Set Performance:

- Decision Tree: Accuracy=0.8734, Precision=0.0000, Recall=0.0000, F1=0.0000
- Logistic Regression: Accuracy=0.8734, Precision=0.0000, Recall=0.0000, F1=0.0000
- SVM: Accuracy=0.8734, Precision=0.0000, Recall=0.0000, F1=0.0000
- KNN (1k train): Accuracy=0.8642, Precision=0.2358, Recall=0.0324, F1=0.0569

Confusion Matrix Highlights:

- All models (except KNN) predicted only the majority class ("no")
- KNN: 25 true positives, 81 false positives

Challenges:

- Severe class imbalance (~11% of samples are 'yes')

- Models failed to identify positive cases

Recommendations:

1. Apply class balancing (class_weight='balanced', SMOTE)

2. Explore additional features (e.g., campaign, economic)

3. Use metrics like F1-score, recall, ROC-AUC

4. Try ensemble models (Random Forest, XGBoost)