Customer Subscription Prediction Using SVM: Analyzing Bank Marketing Data

Githublink:

https://github.com/31SathyaPriya/ClassficationMLProject

Student Name:Doddipalli Sathyapriya

Student ID:23109239

Table of Contents

List of Figures

Introduction

The supervised learning algorithms are called Support Vector Machines (SVM) they are used for classification and regression. The SVM finds the optimal hyper plane which maximizes the margin between the support vectors and boundary. It is effective in high dimensional space and relatively less vulnerable to overfitting because it possesses a regularization parameter.

SVM can handle linear classification, as well as non-linear classification, through kernel tricks, which are used to transform non-linear separable data in high order. It is useful for large data sets. In this tutorial, we perform SVM on the Bank Marketing dataset of the UCI repository. We classify whether a bank customer subscribes to a term deposit using Linear, Polynomial and RBF kernels. Through these kernels, we show how SVM can be used in finance and marketing for the practical purposes of enhancing customer targeting and decision-making for businesses.

Understanding the Technique: Support Vector Machines (SVM)

Concept of SVM: Decision Boundaries and Support Vectors

The Support Vector Machines (SVM) is a supervised algorithm for classification and regression problems. It is operated by determining the decision boundary which distinguishes the data points of type into other classes (Montesinos, Abelardo and Crossa, 2022). The main idea behind SVM is to find a hyperplane, which separates the closest data points from each class (the so-called support vectors), and which has the largest possible margin (Shetty et al., 2022).

Types of Kernels: Linear, Polynomial, RBF

Different kernel functions in SVM can solve both the linearly separable and non-linearly separable data.

- In the case of linearly separable data, a straight-line hyperplane can separate the classes, this makes it a suitable linear Kernel (Valkenburg et al., 2023).
- Polynomial Kernel takes input features to a higher dimensional space via polynomial functions when the relationship is intricate (Montesinos, Abelardo and Crossa, 2022).
- A Radial Basis Function (RBF) Kernel, allows for an infinite dimensional transformation of the data and is therefore extremely useful for the case of non-linear classification (Valkenborg et al., 2023).

Strengths and Weaknesses of SVM

Strengths:

- Works well in high-dimensional spaces.
- In the case when the number of features is greater than the number of samples.
- Not prone to overfitting, especially better with proper regularization (Shetty et al., 2022).
- Solves non-linearly separable data with kernels.

Weaknesses

- Computationally expensive for large datasets
- Sensitive to feature scaling (Valkenborg et al., 2023).
- Given this, optimal performance requires parameter tuning (C, gamma).

Real-World Applications of SVM

The above areas are just some of the areas in which SVM has been used in finance, healthcare and marketing.

- Spam detection (email classification)
- Fraud detection in banking transactions
- Medical diagnosis (cancer detection)
- Customer behaviour prediction in marketing campaigns.

Research and Literature Review

Key Research Papers and Blogs on SVM

The theoretical and practical use of support vector machines has been studied in much machine learning literature and countless research papers and blogs about support vector machines (Montesinos, Abelardo and Crossa, 2022). The one of the most influential papers on SVM, where the concept of maximizing the margin between the support vectors is used to improve the generalization capability. The first was this foundational work that is the basis for current modern SVM implementations. Valkenborg et al. (2023) contribute another important factor in studying kernel methods and how they can assist in transforming data into higher dimensions for more efficient classification (Valkenborg et al., 2023). In addition, Shetty et al., (2022) give a practical guide to SVM parameter selection, the importance of selecting the right kernel and hyperparameters in achieving the best model performance. Tutorials and simplified explanations of SVM implementation on solving real-world problems are given on several blogs, like as the blogs of Towards Data Science and Analytics Vidhya.

SVM in Finance and Marketing Analytics

This feature has made it possible for SVM to be applied in financial and marketing analytics for they can work with high dimensional epochs of data and can classify ad hoc patterns. In finance, SVM is used in credit scoring, fraud detection, and stock price prediction. Through investigating historical patterns and transaction data, SVM manages to achieve high precision when trying to detect fraudulent activities (Tanveer et al., 2022). Customer segmentation, targeted advertising, and churn prediction are possible using SVM in marketing analytics. For example, in the banking sector, SVM helps classify customers in terms of their probability of subscribing to financial products, making the campaign efficient (Tang and Zhu, 2024). The Bank Marketing dataset used in this tutorial is a real-life application where SVM can be applied to predict how the customer would behave and how marketing strategies can be optimized (Shetty et al., 2022).

Theoretical Foundations Behind Kernel Functions

The role of kernel functions in SVM is very important because they allow the SVM to handle the non-linearly separable data. The simplest among all is the linear kernel: it’s a kernel that can be applied to data having a straight-line decision boundary. The use of a polynomial kernel entails the mapping of inputs to a ‘hard higher dimensional’ space to capture complex relationships between the data (Valkenburg et al., 2023). Perhaps the most ubiquitous kernel is the Radial Basis Function (RBF) kernel, which transforms the input features into an infinite-dimensional space where it can perform very nicely at a task requiring intricate classification. These kernels have to be understood in selecting the best technique to use to gain the most out of SVM on the given data set (Tanveer et al., 2022).

Dataset Selection: Bank Marketing Dataset

The UCI Machine Learning Repository is used to source the Bank Marketing dataset which contains data regarding the direct marketing campaigns of a Portuguese bank. The phone calls are involved in these campaigns for promoting term deposits. It has 45,211 instances and 16 input features, of which the classification goal is to estimate whether a client will subscribe to a term deposit (y: yes/no).

Dataset Link: https://archive.ics.uci.edu/dataset/222/bank%2Bmarketing?utm_source

Features and Target Variable

Categorical and numerical features about the client's demographic, financial, and marketing campaign activity (age, job, marital status, education, balance, loan status, contact type, number of contacts, and previous campaign outcome) are present in the dataset. The offspring of the target variable (y) represents whether the client subscribed to a term deposit (yes=1, no=0).

Handling Categorical and Numerical Features

For SVM, categorical features need to be encoded (one hot or label encoding) and numerical features should be scaled (standardization) (Shetty et al., 2022). The feature, since it directly affects the target outcome, creates data leakage, therefore it must be removed (Tanveer et al., 2022).

Ethical and Unbiased Predictions

Fairness should also be evaluated in features like age, job, marital status etc. Marking datasets are unevenly distributed, which means our dataset also has a class imbalance that requires handling (Jadhav et al., 2022). Thus, ethical considerations are related to the avoidance of discriminatory predictions against any group, and the transparency of model decisions (Shetty et al., 2022).

Exploratory Data Analysis (EDA)

Subscription Distribution (Target Variable)

The class imbalance is very severe as most of the customers do not subscribe to the service and are thus represented with a new colour bar. If the imbalance occurs, it will cause the predictions of the machine learning model to be biased. This however requires improving predictive performance and model fairness using, e.g., oversampling, undersampling, or weighted loss functions (Montesinos, Abelardo and Crossa, 2022).

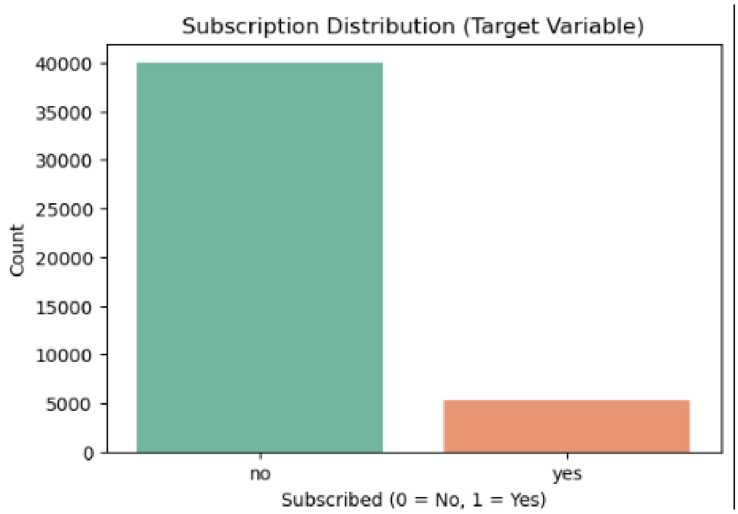

Figure 1 Subscription Distribution

Age Distribution of Clients

As shown in the age distribution histogram, the majority of the clients have grown between 30 to 40 years, having a right-skewed distribution. Because of this, marketing efforts should be aimed at this age group (Shetty et al., 2022). Although subscription rates would be increased with a more personalized marketing strategy for the middle age segment versus other segments of age, some older clients are present as well.

Figure 2 Age distribution

Subscription Rate by Job Type

There is a bar chart that identifies which has the highest subscription rates while the other has the lowest, the highest subscription rates have students and retirees, and low subscription rates have blue-collar workers. This implies that having financial stability and free time do in fact matter in terms of what you decide to subscribe to. Different types of jobs should be targeted with benefits that are of interest to their financial and lifestyle needs.

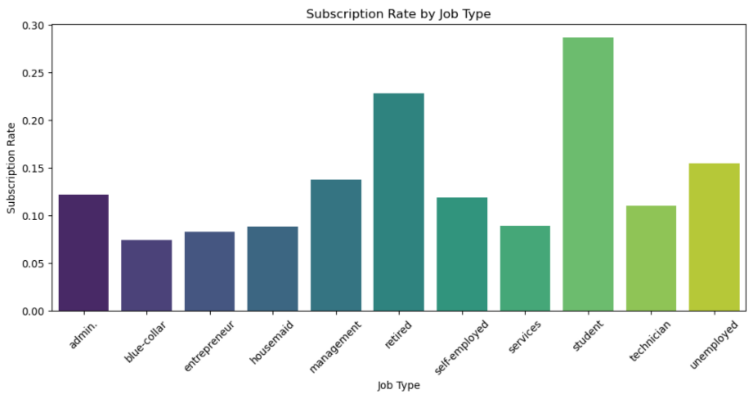

Figure 3 Subscription rate by job type

Effect of Previous Campaigns on Subscription

The boxplot suggests that people who subscribed also had more previously contacted, which is a strong indicator of a positive impact on subscription rate from persistent follow-ups. However, too much contact can also have a negative impression. However, it is necessary to find an optimal balance between marketing interactions to preserve the customer's interest and avoid any possibility of irritation (Jadhav et al., 2022).

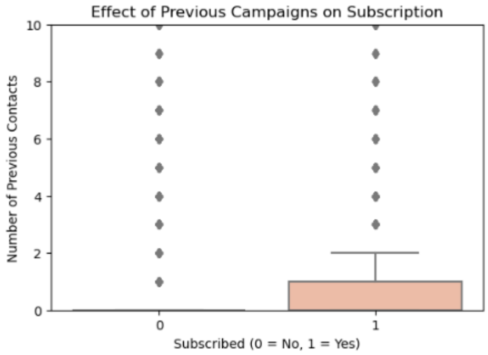

Figure 4 Effect of Previous Campaigns on Subscription

Correlation Heatmap

It is clear from the correlation matrix that call duration has the strongest positive correlation with subscription success—the longer a person spends on the phone, the more likely he or she will be to be persuaded. Most features have weak correlations with the target variable and some of the other variables, such as past campaigns and the number of previous contacts, show moderate.

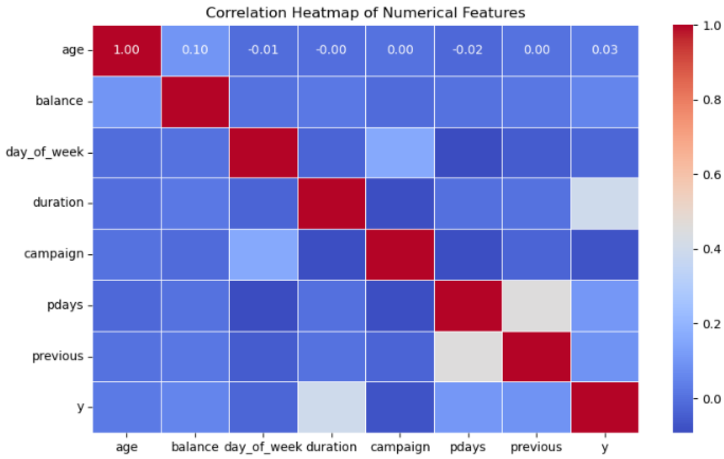

Figure 5 Correlation heat map

Data Preprocessing for SVM

Handling Missing Values

SVM can be affected negatively if missing values are present, so the missing values should be handled before training the model. This means that for numerical features, missing values can be set to the mean or median and for categorical features the mode or a separate 'unknown' category (Tanveer et al., 2022).

Encoding Categorical Features

Categorical features must be encoded for use by SVM since it requires numerical input. When a category has an order, ordinal variables are best suited for label encoding, which assigns numerically to the categories (Shetty et al., 2022). By one hot encoding, we prevent the algorithm from assuming that some column list is an ordinal relationship. If categorical variables are not ordinal, it is often the case that one-hot encoding is preferred to avoid introducing unwanted hierarchy.

Normalization and Feature Scaling

SVM is sensitive to feature magnitudes, depending most of all on RBF and polynomial kernels. Zero means, unit variance or MinMax scaling up to 1 makes sure all features have a proportional contribution to the decision boundary. If there is not enough scaling, features with higher values may collapse and overrule the model's decision when making decisions.

Splitting Data into Training and Validation data set

The dataset will be split into training and testing subsets to evaluate the evaluation of the SVM model, usually with an 80-20 or 70-30 ratio. It prevents such a model from overfitting and thus generalizing well to previously unseen data. If the dataset is imbalanced, stratified sampling can be used so that class distribution is similar within splits.

Training the SVM Model with Three Kernels

Implementing SVM using sklearn.svm.SVC ()

The SVM model is trained using the SVC class from the sklearn. The crucial hyperparameters are C (for Regularization), Kernel (Linear, RBF or Polynomial, in the case of Polynomial the Value of Gamma parameter is needed) and Gamma (for RBF and Polynomial kernels). If the C is higher, the margin is stricter since a high C will allow misclassification to avoid overfitting.

Training on Linear, RBF and Polynomial Kernels

- Best if linearly separable data, as this creates a straight-line decision boundary.
- Complex non-linear patterns can be captured by data mapped to a higher dimensional space by RBF Kernel (Razaque et al., 2021).
- A polynomial kernel captures non-linear relations from polynomial transformations that are expensive computationally for high degree values (Tanveer et al., 2022).

Decision boundaries for different kernels.

It results in very different decision boundaries across kernels. Depending on the kernel, straight boundaries are generated by the linear kernel, flexible, curvilinear ones by the RBF and more complex ones by the polynomial kernel. Accuracy, precision, recall, and F1 score are used to evaluate model performance to choose the kernel for a given dataset that is the best (Tanveer et al., 2022).

Evaluating Model Performance

Accuracy, precision, recall, F1 score and ROC AUC are used to assess the SVM performance. Given that, accuracy measures how correct is the whole thing, and precision and recall provide a way of assessing class-specific performance. Instead of merely misclassification patterns, the confusion matrix indicates information about the patterns.

```
Training SVM with linear kernel...

Accuracy: 0.8830

Classification Report:
              precision    recall  f1-score   support

           0       0.88      1.00      0.94      7985
           1       0.00      0.00      0.00      1058

    accuracy                           0.88      9043
   macro avg       0.44      0.50      0.47      9043
weighted avg       0.78      0.88      0.83      9043

[[7985    0]
 [1058    0]]
```

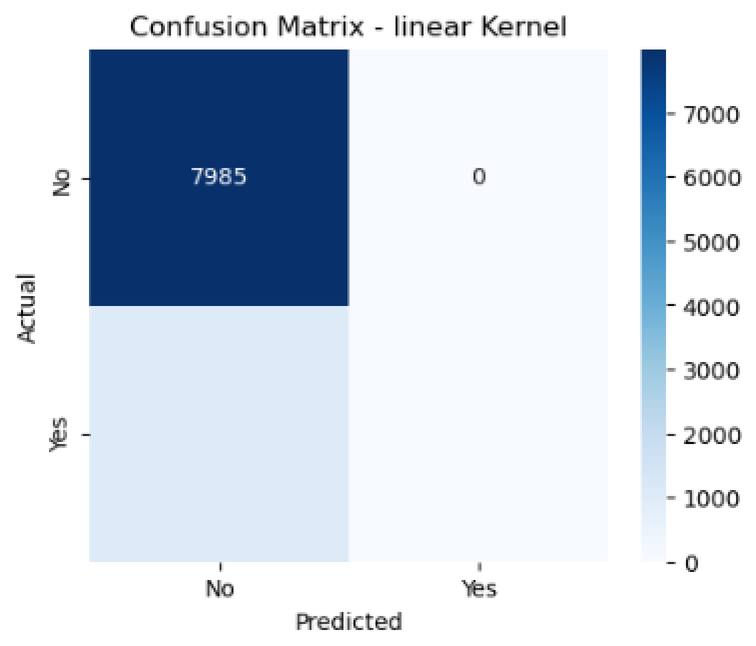

Figure 6 SVM result with linear kernel

The linear kernel is only able to achieve 88.3% accuracy but has zero positive classification accuracy. RBF kernel improves the accuracy to 89.3% and recall to 19% while the Polynomial kernel achieves accuracy to 88.8% and recall to 12%. All the kernels perform well in the classification of the majority class, however, they fail due to the minority class.

```
Training SVM with rbf kernel...

Accuracy: 0.8927

Classification Report:
              precision    recall  f1-score   support

           0       0.90      0.99      0.94      7985
           1       0.64      0.19      0.29      1058

    accuracy                           0.89      9043
   macro avg       0.77      0.59      0.62      9043
weighted avg       0.87      0.89      0.87      9043

[[7876  109]
 [ 861  197]]
```

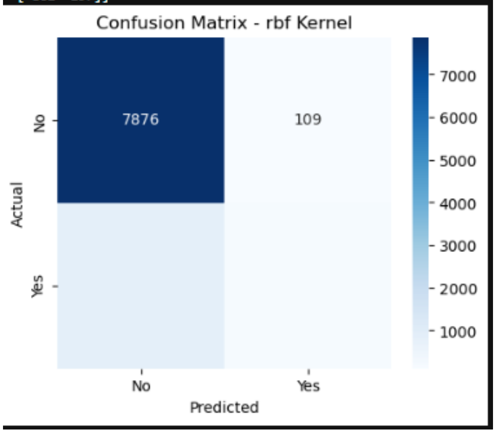

Figure 7 SVM results with RBF kernel

```
Training SVM with poly kernel...

Accuracy: 0.8883

Classification Report:
              precision    recall  f1-score   support

           0       0.89      0.99      0.94      7985
           1       0.62      0.12      0.20      1058

    accuracy                           0.89      9043
   macro avg       0.76      0.56      0.57      9043
weighted avg       0.86      0.89      0.85      9043

[[7905   80]
 [ 930  128]]
```

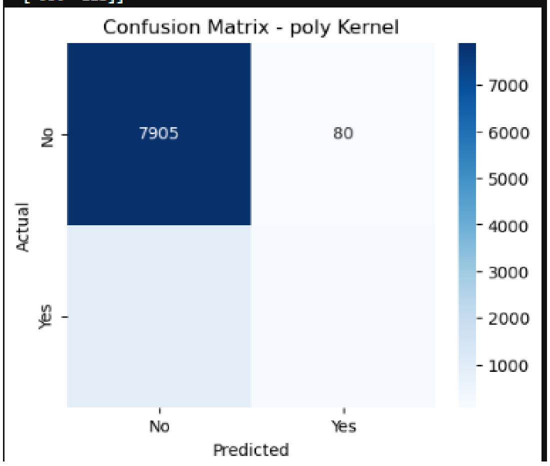

Figure 8 SVM result with Poly kernel

Discussion and Insights

It is useful on its own in the field of bank for fraud detection and customer classification. The linear kernel is computationally efficient but lacking in the case of complex data. RBF kernel is capable of handling non-linear patterns better but is not always overcome with imbalanced data. Even though the Polynomial kernel is flexible it can overfit (Shetty et al., 2022).

To improve performance, techniques such as class balancing procedures (e.g., SMOTE), feature engineering and hyperparameter tuning should be employed. Other alternatives such as Random Forest or Gradient Boosting could be experimented with (Jadhav et al., 2022).

Conclusion

The different SVM kernels have a very significant effect on performance and the Linear kernel is quite fast but is ineffective for imbalanced data. The RBF kernel has a higher flexibility, which causes polynomial kernel risks to overfit and the real-world applications first determine such preprocessing techniques and kernels that help in improving predictive accuracy.

References

Jadhav, A., M. Mostafa, S.M., Elmannai, H. and Karim, F.K. (2022) 'An Empirical Assessment of Performance of Data Balancing Techniques in Classification Task', *Applied Sciences*, 12(8), p.3928. https://doi.org/10.3390/app12083928.

Montesinos, A., Abelardo M. L and Crossa, J. (2022) 'Support Vector Machines and Support Vector Regression', *Springer eBooks*, pp.337–378. https://doi.org/10.1007/978-3-030-89010-0_9.

Razaque, A., Ben Haj Frej, M., Almi'ani, M., Alotaibi, M. and Alotaibi, B. (2021) 'Improved Support Vector Machine Enabled Radial Basis Function and Linear Variants for Remote Sensing Image Classification', *Sensors*, 21(13), p.4431. https://doi.org/10.3390/s21134431.

Shetty, S.H., Shetty, S., Singh, C. and Rao, A. (2022) 'Supervised Machine Learning: Algorithms and Applications', pp.1–16. https://doi.org/10.1002/9781119821908.ch1.

Tang, X. and Zhu, Y. (2024) 'Enhancing bank marketing strategies with ensemble learning: Empirical analysis', *PLOS ONE*, 19(1), p.e0294759. https://doi.org/10.1371/journal.pone.0294759.

Tanveer, M., Rajani, T., Rastogi, R., Shao, Y.H. and Ganaie, M.A. (2022) 'Comprehensive review on twin support vector machines', *Annals of Operations Research*, 339(3), pp.1223–1268. https://doi.org/10.1007/s10479-022-04575-w.

Valkenborg, D., Rousseau, A.-J., Geubbelmans, M. and Burzykowski, T. (2023) 'Support vector machines', *American Journal of Orthodontics and Dentofacial Orthopedics*, 164(5), pp.754–757. https://doi.org/10.1016/j.ajodo.2023.08.003.