

Term deposit subscription classification using MLP and SVM

Abstract

This comprehensive study investigates the efficacy of phone-based direct marketing campaigns for a Portuguese banking institution, with the objective of predicting client subscription to a term deposit. To achieve this goal, two machine learning algorithms, Multi-Layer Perceptron (MLP) and Support Vector Machines (SVM), are employed. To optimize the models, a grid search is employed, which involves systematically varying hyperparameters through an iterative process. This approach enables the identification of optimal parameter values that maximize predictive accuracy. Following model training on 80% of the dataset, performance evaluation is conducted on unseen data comprising 20% of the total dataset. This approach ensures the development of robust models that generalize well to new data.

1. Introduction

Marketing campaigns frequently rely on direct marketing strategies to enhance business performance. Companies utilize targeted outreach to specific customer segments to achieve specific goals. By centralizing customer interactions within a contact center, companies can efficiently manage campaigns. These centers facilitate communication through various channels, with phone calls (landline or mobile) being a popular choice. Telemarketing, a marketing approach that leverages remote interactions, is a key strategy in contact centers [3]. Contacts can be categorized as inbound or outbound, depending on who initiated the contact. Each type presents unique challenges, such as the perception of outbound calls as intrusive. Technology enables marketers to rethink their approach by focusing on maximizing customer lifetime value [7]. By analyzing available data and customer metrics, marketers can build stronger, longer-lasting relationships that align with business demands [1]. Furthermore, selecting the most promising clients who are likely to subscribe to a product is a complex task, considered NP-hard in complexity [2].

This study introduces a novel intelligent decision-making system that can predict phone calls intended to encourage long-term deposits on its own. This approach, which aims to make managerial decision-making easier, gives marketing managers the ability to target the most promising customers and optimise their campaigns. By incorporating analysis, the system evaluates the likelihood of success and gives the manager the last say over how many contacts to make. This strategy results in a more effective use of resources by drastically cutting the time and expenses related to marketing initiatives. Additionally, the system reduces client stress and intrusiveness by focusing on the most likely customers, which results in a more efficient and customer-focused strategy.

1.1 MLP:

A neural network architecture called the Multilayer Perceptron (MLP) imitates the structure of the human brain. It is similarly comparable to an artificial neural network that is feed-forward (ANN). Between the input and output layers of the MLP, there are a number of hidden layers, the number of which varies according on the particular data mining task as shown in Fig.1. A complex network is formed by the connections between each neuron in the hidden layer and the neurons in the following layer. During the learning phase, the weights—which are the connections between neurons—are modified. This procedure is carried out repeatedly until

the error value is less than a predefined cutoff. While the output layer makes predictions about classifications based on the data received from the input layer, the input layer aggregates feature values. The error is computed by comparing the categorised output with the observed output. Based on the error, the network weights are changed from the output layer to the input layer via intermediary layers. Combining the activation functions, node values, and connecting weights allows one to compute the sent information [6].

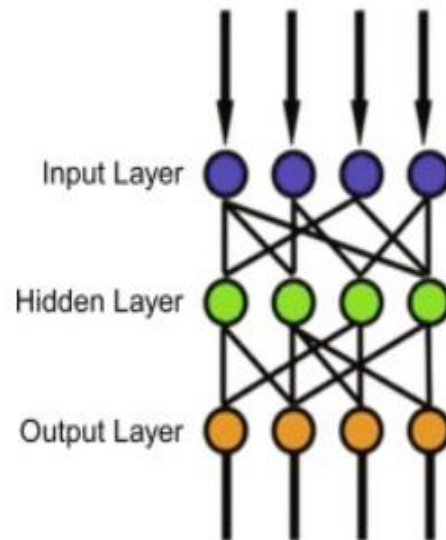


Fig. 1: MLP base architecture

1.2 SVM

Support Vector Machines (SVMs) are a type of supervised machine learning algorithm that can be applied to both classification and regression tasks. The concept of SVMs was first introduced by [4], building on the foundations of statistical learning theory. The core principle of SVMs is to identify the optimal hyperplane that separates classes, utilizing support vectors - the points on the margin that represent each class. This is illustrated in Fig. 2, which depicts the various components of SVMs. When SVMs are applied to regression tasks, they are often referred to as Support Vector Regression (SVR).

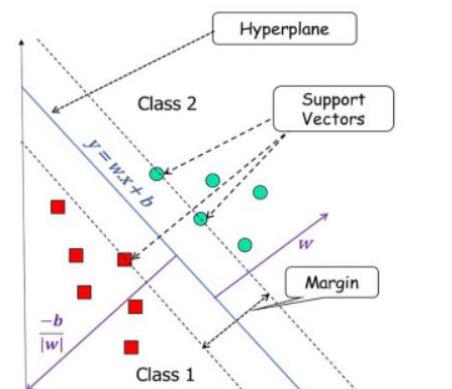


Fig.2. Component of SVM

Because they may function in both linear and non-linear modes, support vector machines (SVMs) provide versatility in solving a range of classification issues. Hard margin and soft margin are the two additional subclasses of linear SVMs. Soft margin SVMs add slack

variables to account for misclassified data points, enabling a more robust classification technique than hard margin SVMs, which assume that the data is totally linearly separable by a hyperplane. The slack variable's value rises as one gets farther away from the margin border. Non-linear SVMs are used when data is not linearly separable. Using kernel functions on the training set, this entails translating the original input space to a higher-dimensional feature space. Sigmoidal, polynomial, and radial basis kernels are frequently used [5]. It has been demonstrated that SVMs are more resilient to noise and outliers, which makes them capable of producing predictions with higher accuracy. The technique provides a variety of model parameter and kernel possibilities, so finding the right combination to produce the best-performing SVM will require careful testing. Moreover, SVMs have proven to perform exceptionally well in high-dimensional spaces, which has led to their widespread use for complicated data sets. SVMs, being a black box method, can be hard to read and comprehend, making it hard to understand the relationships and underlying mechanics of the model.

2. Dataset

The dataset is the result of a large-scale direct marketing campaign that a well-known Portuguese bank. The campaign's goal was to determine the probability that customers would sign up for a bank term deposit by focusing on client interaction through targeted phone calls. Every encounter resulted in a "yes" or "no" response, which offered insightful information about the preferences and behaviour of the clients. We have taken dataset from (<https://archive.ics.uci.edu/dataset/222/bank+marketing>) and is having 45211 instances with 15 features.

2.1 Preprocessing

We are converting categorical attributes to one hot encoding and performing standard scaling to numerical attributes so as to make it fit for model training. Using sklearn pipeline and ColumnTransformer we have performed all this. Also the data was not balanced i.e class 'no' wa having more than 88% of the data as shown in Fig.3. Therefore we applied SMOTE [8] up and down sampling to balance the data to train the model robustly without being biased towards a specific class.

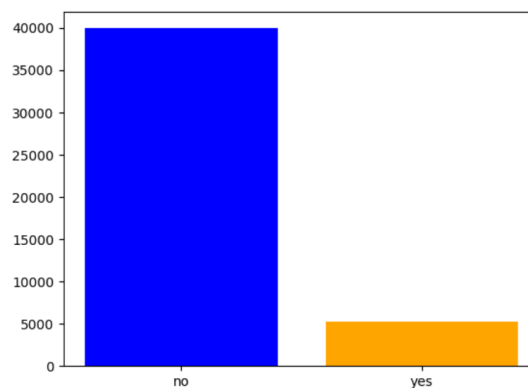


Fig.3: Imbalanced dataset

3. Method

This section covers detailed steps of model training with hyperparameter tuning for MLP and SVM.

3.1 Methodology

We have divided the data into 80:20 ratio for train and test set. Since it was imbalanced, therefore only training set were balanced keeping the test set untouched. Once we have this data we applied grid search on SVM with their various kernel namely *rbf*, *linear*, *poly* and MLP with various *momentum* and *weight_decay* keeping other settings as default . The final best estimator is used to perform prediction on test set. Model is evaluated using accuracy as well as f1-score and precision. We also displayed confusion matrix to see the percentage of correct prediction for both the class to see the reliability of the respective model.

3.2 Architecture and parameters for MLP

MLP with 2 hidden layers with 64 and 32 neurons were used in this study. Following parameters were used to train the mlp.

criterion	CrossEntropyLoss
optimizer	SGD
optimizer__lr	0.01
dropout	0.5
optimizer__weight_decay	[1e-4, 1e-5]
max_epochs	50
batch_size	128
optimizer__momentum	[0.5, 0.9,.999]

3.3 Architecture and parameters for SVM

SVM was used with three different kernels namely liner,ploy and rbf keeping the remaining parameters as default provided by sklearn

4. Result, Findings and Evaluations

4.1 Model selection

Based on results in Table 1 and Table 2 ,we found that SVM with *rbf* kernel on upsampled data outperformed other SVM models and MLP on upsampled data using {'optimizer__momentum': 0.9, 'optimizer__weight_decay': 1e-05} and keeping other parameters as default outperformed MLP variants.

Table 1. Experiment results for MLP

Data	Best parameters	Accuracy	F1-score	precision
imbalanced	{'optimizer__momentum': 0.9,'optimizer__weight_decay': 1e-05}	0.9023	0.543	0.623
Up sampled	{'optimizer__momentum': 0.9, 'optimizer__weight_decay': 1e-05}	0.856	0.587	0.449

downsampled	{'optimizer__momentum': 0.9, 'optimizer__weight_decay': 1e-05}	0.806	0.529	0.374

Table 2: **Experiment results for SVM**

Data	kernal	Accuracy	F1-score	precision
imbalanced	rbf	0.898	0.430	0.664
Up sampled		0.858	0.587	0.453
downsampled		0.819	0.543	0.390
imbalanced	linear	0.882	0.158	0.578
Up sampled		0.816	0.521	0.380
downsampled		0.814	0.520	0.377
imbalanced	ploy	0.895	0.422	0.637
Up sampled		0.856	0.583	0.448
downsampled		0.827	0.553	0.402

4.2 Algorithm comparison

For the imbalanced dataset, the RBF kernel achieved the highest accuracy, F1-score, and precision among all kernel functions. The up-sampled dataset generally resulted in lower performance accuracy but higher f1-score compared to the imbalanced dataset across all kernel functions, indicating that model trained on upsampled data is more robust. the down-sampled dataset exhibited lower performance compared to the upsampled dataset but better than imbalanced dataset across all kernel functions. SVM with rbf kernal on upsampled data outperformed other SVM models.

Overall, the MLP model achieved the highest accuracy and precision on the imbalanced dataset, while performance decreased slightly on the up-sampled and down-sampled datasets in terms of accuracy. This suggests that the model was able to generalize well to the imbalanced dataset, but may have struggled to achieve comparable performance on the up-sampled and down-sampled datasets. Additionally, the chosen hyperparameters (momentum and weight decay) were consistent across all datasets, indicating their effectiveness in optimizing the model's performance. But in terms of overall model performance and their reliability to predict different class without being biased, MLP with up sampled is the best because of high f1-score.

The below confusion matrix shows percentage of correct predictions. We can find that best performing model from SVM and MLP is predicting both the class with equal probability. Also, their results are approx. same in terms of accuracy and f1-score.

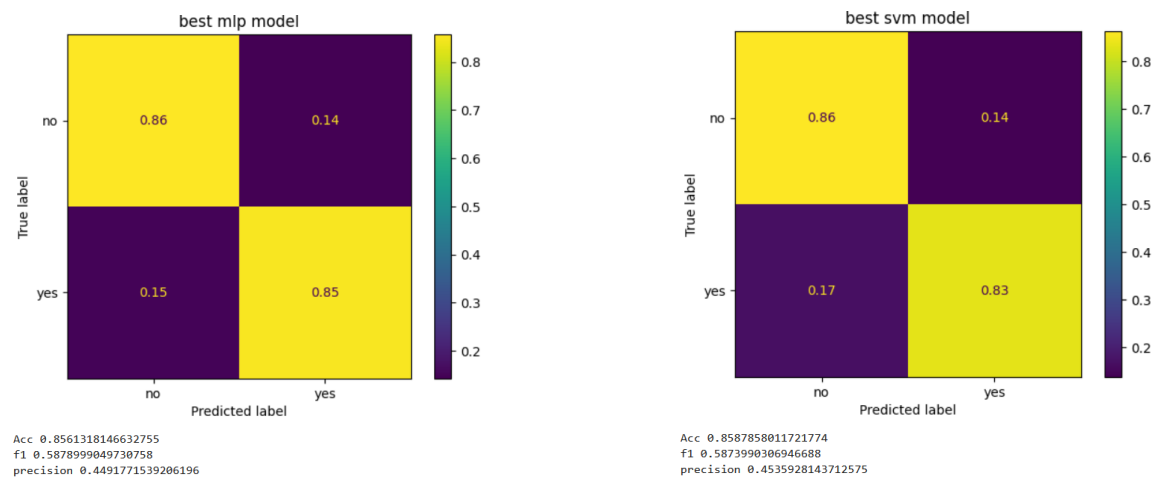


Fig.4: Confusion matrix

5. Conclusion

The study demonstrates the effectiveness of utilizing predictive models to forecast client subscription to a term deposit based on phone-based direct marketing campaigns. By analyzing various features and client demographics, the model was able to provide valuable insights into the likelihood of subscription, aiding the bank in making informed marketing decisions. We applied SVM and MLP to achieve this and found that in our restricted environment both the models performed approx. same. MLP is much faster in training and inference on testing as compared to SVM. Therefore, we can conclude that bank can utilize this as a filter technique to optimize their business.

6. Reference

- [1] Rust, R. T., Moorman, C., & Bhalla, G. (2010). Rethinking marketing. *Harvard business review*, 88(1/2), 94-101.
- [2] Nobibon, F. T., Leus, R., & Spieksma, F. C. (2011). Optimization models for targeted offers in direct marketing: Exact and heuristic algorithms. *European Journal of Operational Research*, 210(3), 670-683.
- [3] Kotler, P. (2009). *Marketing management*. Pearson Education India.
- [4] Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992, July). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory* (pp. 144-152).
- [5] Rani, A., Kumar, N., Kumar, J., & Sinha, N. K. (2022). Machine learning for soil moisture assessment. In *Deep learning for sustainable agriculture* (pp. 143-168). Academic Press.
- [6] Tradigo, G., Calabrese, B., Macrí, M., Vocaturo, E., Lombardo, N., & Veltri, P. (2015, September). Voice signal features analysis and classification: looking for new diseases related parameters. In *Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics* (pp. 589-596).
- [7] Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, 62, 22-31.
- [8] <https://imbalanced-learn.org/stable/index.html>