

MLP Neural Networks vs. Logistic Regression: A Comparative Study of Customer Churn Prediction in Bank Marketing

Tomislav Medić

Faculty of Economics and Business, University of Zagreb, Croatia

Antonio Pavlečić

Faculty of Economics and Business, University of Zagreb, Croatia

Amir Topalović

Aisma S.R.L., Milan, Italy

Abstract

Customer churn is a recurring problem in bank marketing, and there are several machine learning approaches that can help. This paper compares two of them: a Multilayer Perceptron (MLP) neural network and logistic regression. We used the UCI Bank Marketing dataset, which has 41,188 client records from direct marketing campaigns run by a Portuguese bank between 2008 and 2013. Training and evaluation were kept the same for both models. We adjusted class weights to deal with the imbalance, since only about 11 percent of clients subscribed. Performance was measured with precision, recall, F1-score, and ROC-AUC. The MLP did better on the main metrics, specifically recall for the positive class and ROC-AUC, but not by much. Logistic regression actually performed better on overall accuracy and a few other measures. It also has the benefit of being easier to interpret. What this suggests is that on structured tabular data, a more complex model does not automatically translate into better results.

Keywords: customer churn, bank marketing, multilayer perceptron, logistic regression, binary classification, machine learning

JEL classification: G21, C45, C25, M31

Paper type: Research article

Received: 1 June 2025

Accepted: 20 July 2025

DOI: 10.54820/entrenova-2025-0088

Introduction

Customer churn, broadly defined as the tendency of clients to disengage from a product or service, is a persistent challenge in the banking industry. Keeping existing clients is generally cheaper than acquiring new ones, so being able to flag at-risk customers early has obvious business value (Guliyev & Yerdelen, 2021). Phone-based direct marketing campaigns, like the ones this study draws on, are one of the main tools banks use to win clients back and promote new products. But running those campaigns effectively depends on knowing who to call. Predictive models offer a more disciplined way to make that call, and can meaningfully cut down on wasted outreach (Buddiga, 2022).

Machine learning has become a go-to in banking and financial services for classification problems like credit scoring, fraud detection, and churn prediction (Heß & Damásio, 2025). Logistic regression and neural networks are two of the more commonly used approaches. The former is valued for its simplicity and the fact that its outputs are easy to interpret. The latter can pick up on complex, non-linear patterns that simpler models miss (Almeida, 2002). That said, a neural network is not always the better choice on structured tabular data, and how much complexity is actually worth the trade-off is still an open question in the literature (Adekunle et al., 2025; Borisov et al., 2022).

This study adds to that conversation by putting an MLP neural network and a logistic regression model head-to-head on the UCI Bank Marketing dataset (Moro et al., 2014), which covers direct marketing campaigns from a Portuguese bank spanning 2008 to 2013. The dataset is a good test case: it has real class imbalance, mixes demographic and macroeconomic features, and uses a binary outcome to indicate whether a client subscribed to a term deposit. The expectation going in was that the MLP would perform better. Performance is measured using recall, precision, F1-score, and ROC-AUC, with recall on the positive class treated as the main metric, since missing a potential subscriber tends to cost more than an unnecessary call.

Methodology

The methodology breaks down into four parts. First, we describe the dataset and the steps taken to prepare it. Then we define the features and the target variable. After that comes the MLP neural network, which is the primary model in this study. Finally, we present logistic regression as the benchmark for comparison.

Both models were implemented in Python. Logistic regression used scikit-learn, and the MLP neural network was built with TensorFlow and Keras. All preprocessing, training, and evaluation steps were conducted in Python 3.12.10.

Dataset and Data Preparation

The dataset used here is the Bank Marketing dataset (Moro et al., 2014), which pulls from phone-based direct marketing campaigns run by a Portuguese bank between 2008 and 2013. Each record represents one client contact and includes twenty variables covering things like age, job type, education level, housing situation, and the type of contact made, among others.

The full dataset has 41,188 observations. Before training, we ran it through a standard preprocessing pipeline. Categorical variables were converted to numerical format using one-hot encoding (Qu et al., 2016), and all numerical features were standardized by centering and scaling to keep values on a comparable scale (Hopkins & Rowlands, 2024). Only about 11 percent of clients accepted the offer, so class imbalance was a real concern. We handled it by adjusting class weights during

training, which keeps the models from simply defaulting to the majority class. The data were then split into training (80 percent) and test (20 percent) sets using stratified sampling, so the class distribution stays consistent across both subsets.

Feature and Target Definition

Once preprocessing is done, the feature set includes everything available on the client and campaign side. Client features are age, job type, marital status, education level, whether they've defaulted on credit, and whether they have housing or personal loans. On the campaign side, we have the type of contact, when the last contact happened (month and day), how long the call lasted, how many times they were contacted during this campaign, how many days passed since the previous campaign, the number of prior contacts, and how that previous campaign turned out. There are also four macroeconomic indicators thrown in: employment variation rate, consumer price index, consumer confidence index, and the Euribor three-month rate.

The target variable is straightforward. A 1 means the client subscribed to a term deposit, and a 0 means they did not. That makes this a binary classification problem (Kumari & Srivastava, 2017), where the task is to pick out which clients are most likely to say yes to future campaigns.

Multilayer Perceptron Neural Network

A Multilayer Perceptron is a type of feedforward neural network that has been around for a while (Popescu et al., 2009; Ramchoun et al., 2016; Kruse et al., 2022). It is made up of an input layer, one or more hidden layers, and an output layer. Neurons sit in each layer, and the connections between them have weights and biases that get adjusted during training. The network learns by running backpropagation (Dampfhofer et al., 2024), which tweaks those parameters to bring down the loss function on the training data.

Our MLP uses an input layer sized to match the number of features after preprocessing. Two hidden layers follow, both using ReLU (Rectified Linear Unit) activations to introduce non-linearity and help the model learn more complex patterns (Bai, 2022). The output layer has a single neuron with a sigmoid activation, which produces a probability score between 0 and 1 for how likely a positive subscription is. Binary cross-entropy serves as the loss function, which fits naturally with binary classification (Ho & Wookey, 2019).

To keep overfitting in check, we applied dropout regularization within the hidden layers (Salehin & Kang, 2023). Dropout randomly shuts off a portion of neurons on each training pass. Weights are updated using the Adam optimizer (Jais & Ismail, 2019), which adapts the learning rate as training goes on and tends to work well when class distributions are uneven. We also used early stopping, which watches the validation loss across epochs and stops training when no improvement shows up for a set number of consecutive epochs. In this case, that means training halts before the model starts memorizing the training data instead of generalizing from it.

Logistic Regression

Logistic regression is the baseline model in this comparison. It is a well-known method for binary classification that estimates the probability of an outcome based on input features (Hilbe, 2025). The model fits a linear decision boundary and passes the output through a sigmoid function to produce class probabilities. It is simpler than the MLP, but that simplicity makes it a useful reference point for determining how much the added complexity actually helps.

We trained the logistic regression model on the same preprocessed data as the MLP. L2 regularization was applied to reduce overfitting, with the regularization strength chosen by cross-validation on the training set (Ghojogh & Crowley, 2023). The same class weighting approach used for the MLP was applied here as well. Coefficients were estimated with a standard solver for binary classification. Since logistic regression produces direct probability estimates, it can be evaluated with the same metrics as the MLP, which makes for a clean comparison between the two.

Results

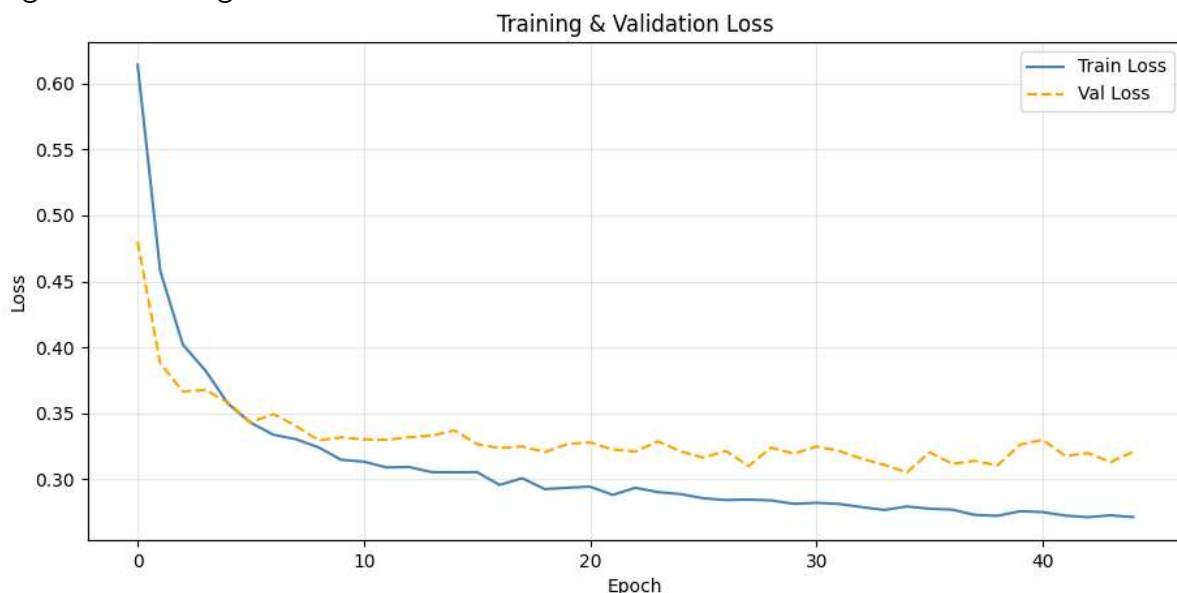
This section covers the results from both models on the UCI Bank Marketing dataset. We start with the MLP neural network, then move to logistic regression, and close with a direct comparison. Along the way, we discuss what the numbers actually mean for the churn prediction task.

We evaluate both models using four metrics: accuracy, precision, recall, and area under the ROC curve (AUC). With only around 11 percent of clients subscribing, accuracy alone does not tell the full story (Dinga et al., 2019). For that reason, recall on the Subscribe class is the primary metric here. Getting those clients right is the whole point of the prediction task. AUC is used to capture overall discriminative ability, since it reflects performance across all possible thresholds rather than at just one (Muschelli, 2020).

MLP Neural Network

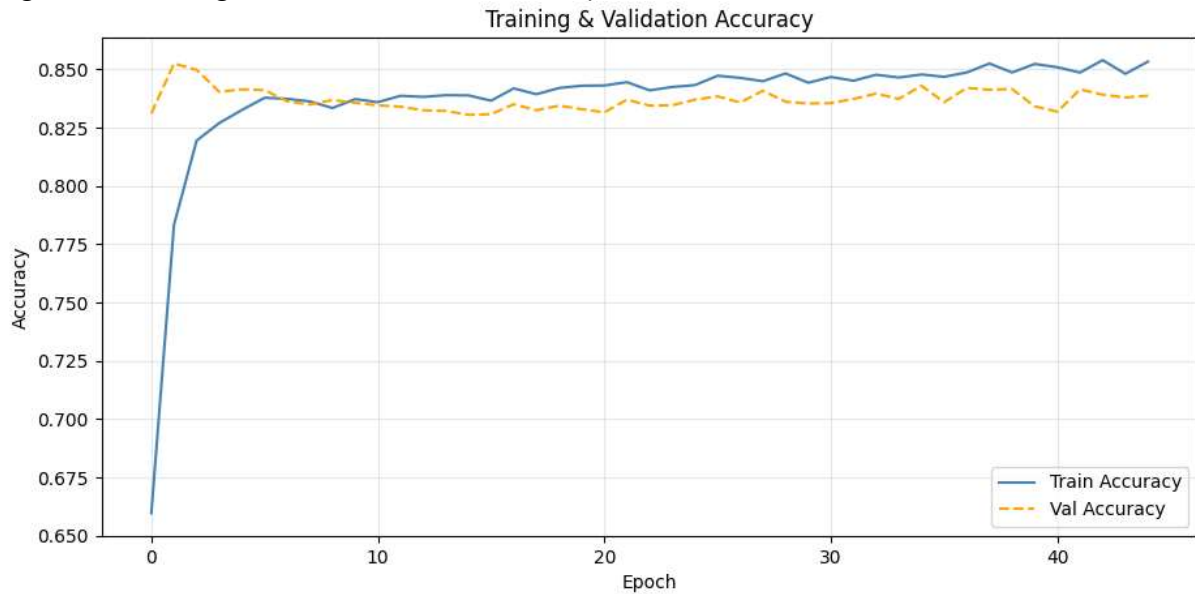
The MLP was trained on 80 percent of the data and evaluated on the remaining 20 percent, which amounts to 8,238 observations. We used early stopping to monitor validation loss during training. The process halted at around epoch 45, which gave the model enough time to learn without pushing into overfitting territory (Bergman et al., 2024).

Figure 1: Training and Validation Loss of the MLP Neural Network



Source: Author's illustration

Figure 2: Training and Validation Accuracy of the MLP Neural Network

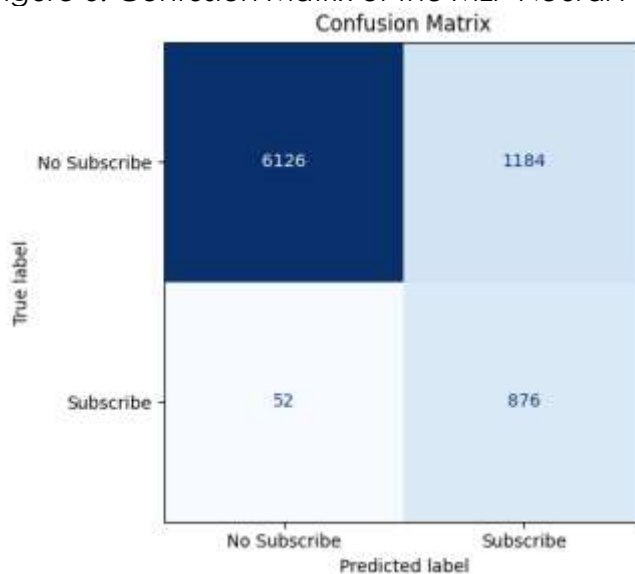


Source: Author's illustration

Figures 1 and 2 show how training and validation loss and accuracy evolved across epochs. The loss curves converge without much divergence, so training was stable. A small gap opens up between training and validation loss around epoch 10, which hints at mild overfitting, but early stopping stepped in before it became a problem. One thing to note: validation accuracy actually starts a bit higher than training accuracy in the early epochs. That might seem odd, but it makes sense when dropout is turned on during training and off during validation (Tsymbalov et al., 2018). By the end, both curves level out and meet up, so the model is generalizing the way it should.

Overall accuracy on the test set came in at 85 percent. With the class imbalance in this dataset, though, that figure does not say much on its own. The confusion matrix breaks things down in more detail.

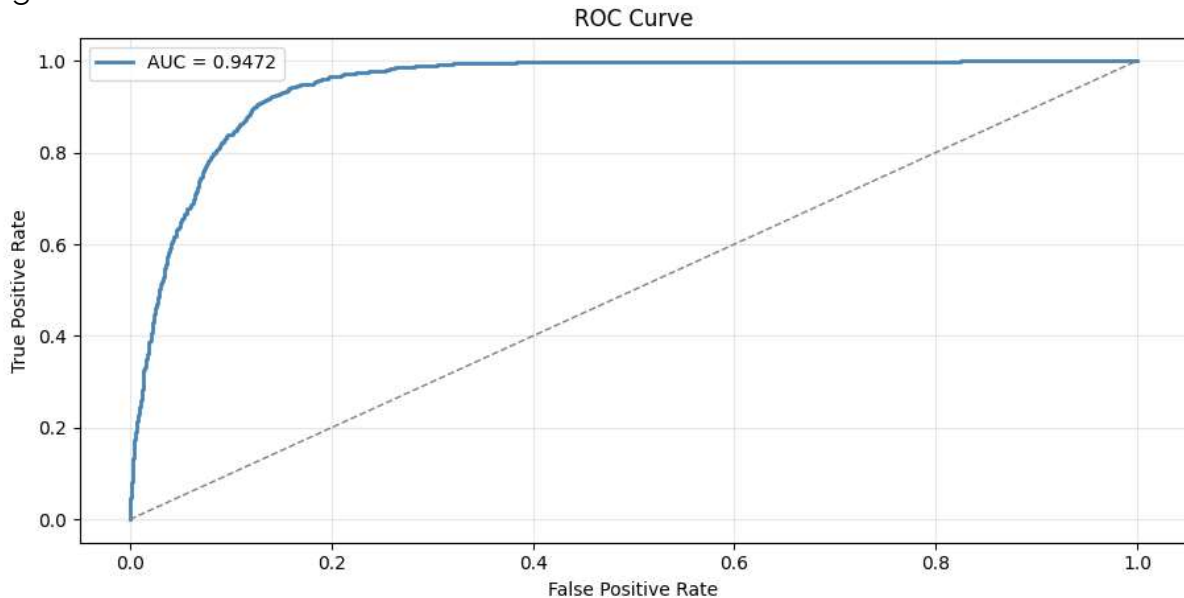
Figure 3: Confusion Matrix of the MLP Neural Network



Source: Author's illustration

Figure 3 shows the MLP caught 876 out of 928 actual subscribers, which works out to a recall of 0.94 on the Subscribe class. That means only 52 were missed. The high recall here is not accidental. Using balanced class weights during training forces the model to focus harder on the minority class. The downside is precision drops to 0.43, so the model ended up flagging 1,184 non-subscribers as potential converts. For a direct marketing campaign, that is fine. Calling someone who does not subscribe costs less than missing someone who would have (Correa Bahnsen et al., 2015).

Figure 4: ROC Curve of the MLP Neural Network



Source: Author's illustration

Figure 4 shows the ROC curve for the MLP. It shoots up fast toward the top-left and then levels off. That is the shape you want to see for a classifier that works well. AUC came out to 0.9472. What that means is the model consistently ranks people who actually subscribed higher than those who did not, and it does that regardless of what threshold you pick (Hand, 2009). All the results are laid out in Table 1.

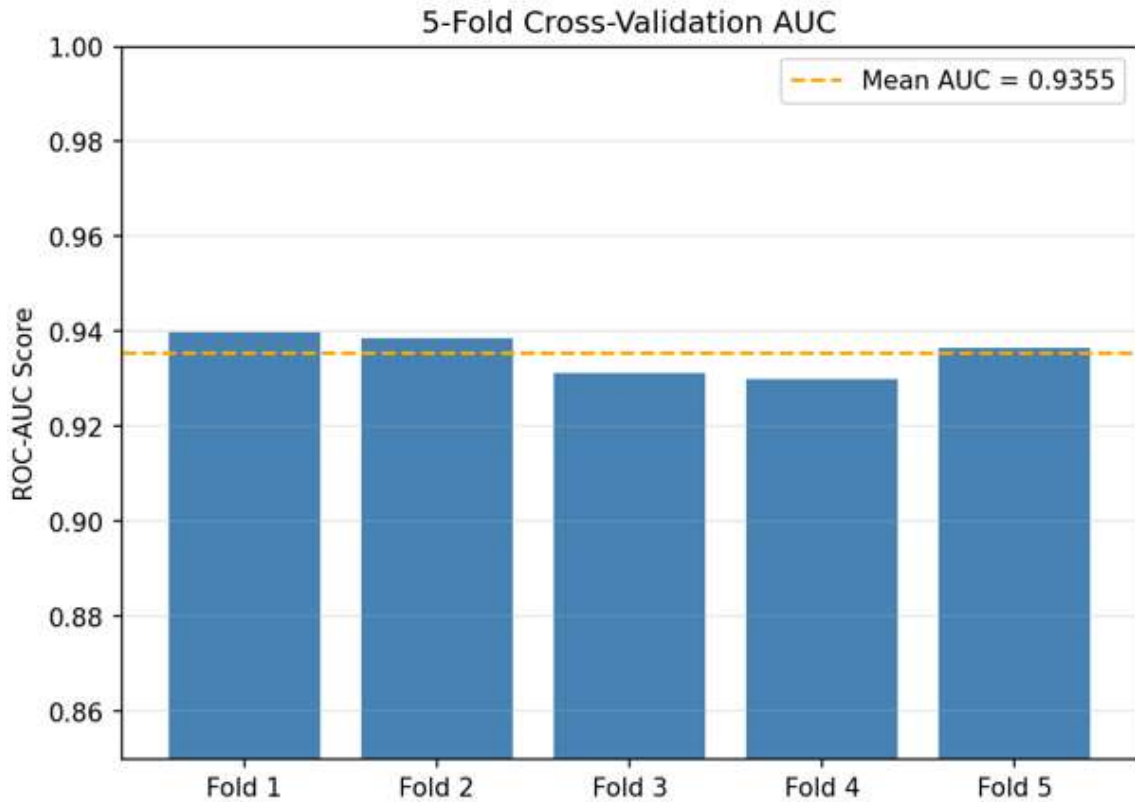
Logistic Regression

For logistic regression, we kept everything the same. Same split, same preprocessing. L2 regularization was turned on to prevent overfitting. Class weights were balanced to deal with the skew in the target variable. One thing we did before testing was run 5-fold stratified cross-validation on the training set. The idea was to check whether the model would hold up consistently (Almonteros & Matias, 2024).

Figure 5 shows how that went. AUC scores stayed pretty tight across all five folds. They ranged from 0.929 to 0.939. Mean came in at 0.9355. Standard deviation was low. So, the model was not sensitive to which subset it trained on, which is a good sign. After cross-validation, we trained on the full set and tested on the holdout data.

Logistic regression hit 87 percent accuracy on the test set. Out of 928 actual subscribers, it caught 846, so recall on the Subscribe class was 0.91. That left 82 subscribers misclassified as non-subscribers.

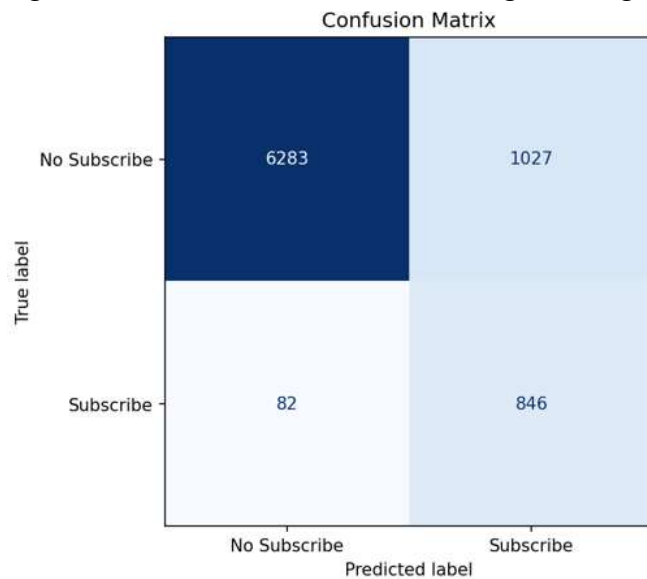
Figure 5: Cross-Validation AUC Scores of the Logistic Regression Model



Source: Author's illustration

Figure 6 breaks down where the errors happened. The model flagged 1,027 non-subscribers as likely converts, which puts precision at around 0.45 for the Subscribe class. That is a bit higher than what the MLP got, but logistic regression also missed more actual subscribers. You see this tension between precision and recall all the time when classes are imbalanced (Juba & Le, 2019). ROC-AUC on the test set came in at 0.9438.

Figure 6: Confusion Matrix of the Logistic Regression Model

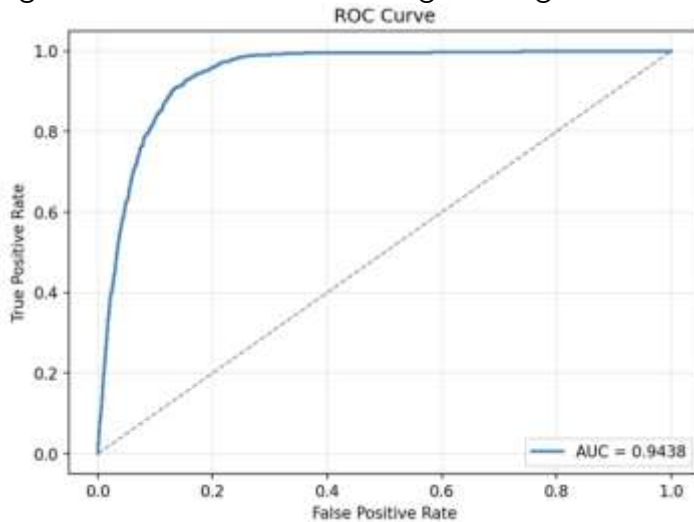


Source: Author's illustration

Figure 7 shows the ROC curve for logistic regression. The shape looks similar to what we saw for the MLP, which confirms it has solid discriminative ability overall. The AUC is slightly lower, though, and that difference points to the MLP picking up on patterns that a linear decision boundary just cannot capture (Vlassopoulos et al., 2020).

One clear advantage logistic regression has over the MLP is interpretability. The coefficients tell you directly which features push the prediction up or down, and by how much.

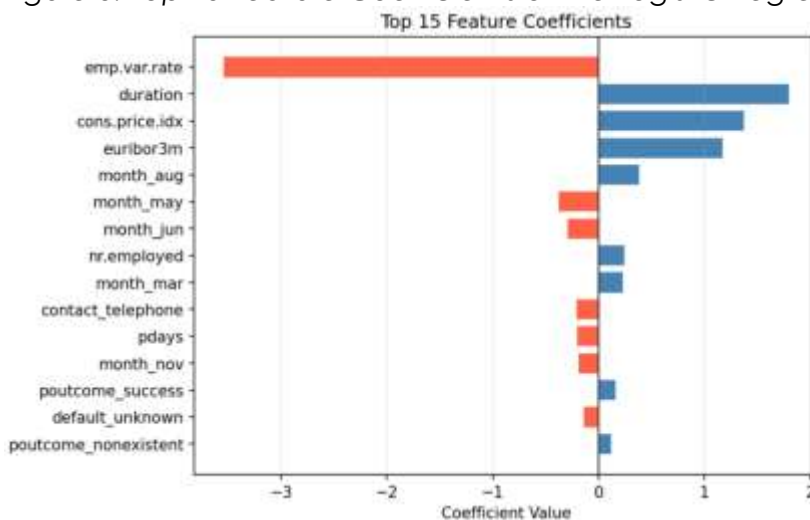
Figure 7: ROC Curve of the Logistic Regression Model



Source: Author's illustration

Figure 8 shows the top 15 features ranked by coefficient strength. Employment variation rate came out as the strongest negative predictor, which suggests clients are less inclined to subscribe when the economy looks unstable. Call duration was the strongest positive predictor. That makes sense, since longer calls probably mean the client is more engaged. Macroeconomic indicators like the consumer price index and the Euribor three-month rate also ranked high. The month of contact mattered too. May and June were both tied to lower subscription rates.

Figure 8: Top 15 Feature Coefficients of the Logistic Regression Model



Source: Author's illustration

Model Comparison

This section puts the two models, side by side. Both were trained and tested under identical conditions. Same data split, same preprocessing, same class weighting. Any differences in performance come from the models themselves, not from how the data was handled.

Table 1 summarizes the key metrics for both models on the test set.

Table 1. Model Comparison — Accuracy, Precision, Recall, F1-score, AUC

Metric	MLP	Logistic Regression
Accuracy	0.85	0.87
Precision (No Subscribe)	0.99	0.99
Recall (No Subscribe)	0.83	0.86
F1-Score (No Subscribe)	0.91	0.92
Precision (Subscribe)	0.43	0.45
Recall (Subscribe)	0.94	0.91
F1-Score (Subscribe)	0.59	0.60
Weighted F1-Score	0.87	0.88
ROC-AUC	0.9472	0.9438
False Negatives	52	82

Source: Authors' work

The two models land pretty close on almost every metric. The MLP has the edge on recall, hitting 0.94 versus 0.91 for logistic regression. That translates to 30 fewer missed subscribers. The MLP also edges ahead on AUC. It got 0.9472 versus 0.9438 for logistic regression, which means it does a slightly better job ranking who is likely to subscribe. Now, the F1-scores for the Subscribe class are 0.59 and 0.60. Those numbers might seem modest, but there is a reason for that. We tuned both models to catch as many subscribers as possible. When you do that, recall goes up but precision takes a hit. F1 reflects that trade-off.

Logistic regression does slightly better on precision and overall accuracy. It also has the advantage of being easy to interpret through its coefficients. For the main task of finding potential subscribers, though, the MLP comes out ahead. That said, logistic regression is still a solid choice, especially when you need a model that is easier to explain.

Discussion

The question we started with was whether an MLP neural network would beat logistic regression at predicting customer churn in a bank marketing context. It did, but not by much.

Recall on the Subscribe class was 0.94 for the MLP and 0.91 for logistic regression. AUC came in at 0.9472 versus 0.9438. That translates to 30 more correctly identified subscribers out of the same 8,238 test observations. For a marketing campaign where missing a subscriber is more expensive than making an extra call, those 30 matter. That said, the difference is not dramatic. Logistic regression stayed competitive and has the advantage of being far easier to interpret and cheaper to run.

This fits with what other researchers have found. Neural networks do not always pull ahead by much on tabular data, especially when the dataset is moderate in size and has clear linear patterns baked into it (Jain et al., 2022). When you look at the logistic regression coefficients, call duration and the employment variation rate stand out as the strongest predictors. That makes sense given the context. Call duration probably

reflects how engaged the client is, and economic instability tends to make people less likely to commit to financial products. The fact that these patterns are interpretable and intuitive gives logistic regression real credibility as more than just a benchmark.

There are some limitations here. The data come from one bank in Portugal over a five-year stretch from 2008 to 2013, so it is not clear how well these results would carry over to other settings. Call duration is also a problem. It was the strongest positive predictor, but you do not know how long a call will last until after it happens. Using it in the model probably inflates the performance numbers, which is a classic case of data leakage (Rosenblatt et al., 2024). We only tested two models. Gradient boosting and ensemble methods have done well on similar problems, so a more thorough comparison would include those (Ren et al., 2016).

Conclusion

We set out to compare an MLP neural network and logistic regression for predicting customer churn in a bank marketing setting. The data came from the UCI Bank Marketing dataset, which has 41,188 observations from a Portuguese bank. Both models went through the same training and evaluation process, with class weights adjusted to handle the imbalance in the target variable.

The MLP came out ahead, though not by a lot. AUC landed at 0.9472 for the MLP and 0.9438 for logistic regression. On recall, the MLP hit 0.94 while logistic regression got to 0.91. What that means in practice is the MLP identified 30 more subscribers in the test set. Both models ranked potential subscribers well overall. Logistic regression held up as a solid alternative and has the benefit of being easier to explain, which reinforces something worth remembering: more complexity does not always get you better performance on tabular data like this.

There are a few things to keep in mind about the scope of this study. The data only cover one bank over one time period, so how much this generalizes is unclear. Call duration was one of the strongest predictors, but that variable is not something you have before the call happens. Including it in the model probably pushed the performance numbers higher than they should be.

A few directions could be worth exploring next. Running the same comparison on data from other banks or more recent years would help show whether these results hold up elsewhere. Dropping call duration and retraining both models would give a fairer sense of what they can actually do. Adding gradient boosting or ensemble methods to the comparison would also fill out the picture of what works best for this kind of problem.

References

1. Adekunle, A. A., Fofana, Issouf., Picher, P., Rodriguez-Celis, E. M., Arroyo-Fernandez, O. H., & Zemouri, R. (2025). Optimizing deep learning predictive models: A comprehensive review of RNN and its variant architectures. *Applied Soft Computing*, 185, 114015. <https://doi.org/10.1016/j.asoc.2025.114015>
2. Almeida, J. S. (2002). Predictive non-linear modeling of complex data by artificial neural networks. *Current Opinion in Biotechnology*, 13(1), 72–76. [https://doi.org/10.1016/S0958-1669\(02\)00288-4](https://doi.org/10.1016/S0958-1669(02)00288-4)
3. Almonteros, J. R., & Matias, J. B. (2024). Integration of Stratified KFold Cross Validation to Enhance Prediction Accuracy: A Comparison Study. *2024 5th International Conference on Data Analytics for Business and Industry (ICDABI)*, 81–85. <https://doi.org/10.1109/ICDABI63787.2024.10800425>
4. Bai, Y. (2022). RELU-function and derived function review. *SHS Web of Conferences*, 144, 02006.

5. Bergman, E., Purucker, L., & Hutter, F. (2024). *Don't Waste Your Time: Early Stopping Cross-Validation*. <https://arxiv.org/abs/2405.03389>
6. Borisov, V., Leemann, T., Seßler, K., Haug, J., Pawelczyk, M., & Kasneci, G. (2022). Deep Neural Networks and Tabular Data: A Survey. *IEEE Transactions on Neural Networks and Learning Systems*, *PP*, 1–21. <https://doi.org/10.1109/TNNLS.2022.3229161>
7. Buddiga, S. K. P. (2022). *Improving Targeted Client Acquisition: Predictive Analysis of Retail Bank Direct Marketing Campaigns*.
8. Correa Bahnsen, A., Aouada, D., & Ottersten, B. (2015). A novel cost-sensitive framework for customer churn predictive modeling. *Decision Analytics*, *5*, 1–15. <https://doi.org/10.1186/s40165-015-0014-6>
9. Dampfhofer, M., Mesquida, T., Valentian, A., & Anghel, L. (2024). Backpropagation-Based Learning Techniques for Deep Spiking Neural Networks: A Survey. *IEEE Transactions on Neural Networks and Learning Systems*, *35*(9), 11906–11921. <https://doi.org/10.1109/TNNLS.2023.3263008>
10. Dinga, R., Penninx, B. W. J. H., Veltman, D. J., Schmaal, L., & Marquand, A. F. (2019). Beyond accuracy: Measures for assessing machine learning models, pitfalls and guidelines. *bioRxiv*. <https://doi.org/10.1101/743138>
11. Ghojogh, B., & Crowley, M. (2023). *The Theory Behind Overfitting, Cross Validation, Regularization, Bagging, and Boosting: Tutorial*. <https://arxiv.org/abs/1905.12787>
12. Guliyev, H., & Yerdelen, F. (2021). Customer churn analysis in banking sector: Evidence from explainable machine learning models. *Journal of Applied Microeconometrics*, *1*, 85–99. <https://doi.org/10.53753/jame.1.2.03>
13. Hand, D. J. (2009). Measuring classifier performance: A coherent alternative to the area under the ROC curve. *Machine Learning*, *77*(1), 103–123. <https://doi.org/10.1007/s10994-009-5119-5>
14. Heß, V. L., & Damásio, B. (2025). Machine learning in banking risk management: Mapping a decade of evolution. *International Journal of Information Management Data Insights*, *5*(1), 100324. <https://doi.org/10.1016/j.ijime.2025.100324>
15. Hilbe, J. M. (2025). Logistic Regression. In M. Lovric (Ed.), *International Encyclopedia of Statistical Science* (pp. 1386–1390). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-69359-9_333
16. Ho, Y., & Wookey, S. (2019). The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling. *IEEE Access*, *8*, 4806–4813.
17. Hopkins, W., & Rowlands, D. (2024). Standardization and other approaches to meta-analyze differences in means. *Statistics in Medicine*, *43*. <https://doi.org/10.1002/sim.10114>
18. Jain, V., Goel, M., & Shah, K. (2022). Deep Learning on Small Tabular Dataset: Using Transfer Learning and Image Classification. In A. Dev, S. S. Agrawal, & A. Sharma (Eds), *Artificial Intelligence and Speech Technology* (pp. 555–568). Springer International Publishing.
19. Jais, I. K. M., & Ismail, A. R. (2019). Adam optimization algorithm for wide and deep neural network. *Knowledge Engineering and Data Science*, *2*(1), 10.
20. Juba, B., & Le, H. S. (2019). Precision-Recall versus Accuracy and the Role of Large Data Sets. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(01), 4039–4048. <https://doi.org/10.1609/aaai.v33i01.33014039>
21. Kruse, R., Mostaghim, S., Borgelt, C., Braune, C., & Steinbrecher, M. (2022). Multi-layer perceptrons. In *Computational intelligence: A methodological introduction* (pp. 53–124). Springer.
22. Kumari, R., & Srivastava, S. (2017). Machine Learning: A Review on Binary Classification. *International Journal of Computer Applications*, *160*, 11–15. <https://doi.org/10.5120/ijca2017913083>
23. Moro, S., Rita P., Cortez, P. (2014). *Bank Marketing [Data set]*. UCI Machine Learning Repository. <https://doi.org/10.24432/C5K306>
24. Muschelli, J. (2020). ROC and AUC with a Binary Predictor: A Potentially Misleading Metric. *Journal of Classification*, *37*(3), 696–708. <https://doi.org/10.1007/s00357-019-09345-1>
25. Popescu, M.-C., Balas, V. E., Perescu-Popescu, L., & Mastorakis, N. (2009). Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*, *8*(7), 579–588.

26. Qu, Y., Cai, H., Ren, K., Zhang, W., Yu, Y., Wen, Y., & Wang, J. (2016). *Product-based Neural Networks for User Response Prediction*. <https://arxiv.org/abs/1611.00144>
27. Ramchoun, H., Ghanou, Y., Ettaouil, M., & Janati Idrissi, M. A. (2016). *Multilayer perceptron: Architecture optimization and training*.
28. Ren, Y., Zhang, L., & Suganthan, P. N. (2016). Ensemble Classification and Regression-Recent Developments, Applications and Future Directions [Review Article]. *IEEE Computational Intelligence Magazine*, 11(1), 41–53. <https://doi.org/10.1109/MCI.2015.2471235>
29. Rosenblatt, M., Tejavibulya, L., Jiang, R., Noble, S., & Scheinost, D. (2024). Data leakage inflates prediction performance in connectome-based machine learning models. *Nature Communications*, 15(1), 1829. <https://doi.org/10.1038/s41467-024-46150-w>
30. Salehin, I., & Kang, D.-K. (2023). A review on dropout regularization approaches for deep neural networks within the scholarly domain. *Electronics*, 12(14), 3106.
31. Tsymbalov, E., Panov, M., & Shapeev, A. (2018). Dropout-Based Active Learning for Regression. In W. M. P. van der Aalst, V. Batagelj, G. Glavaš, D. I. Ignatov, M. Khachay, S. O. Kuznetsov, O. Koltsova, I. A. Lomazova, N. Loukachevitch, A. Napoli, A. Panchenko, P. M. Pardalos, M. Pelillo, & A. V. Savchenko (Eds), *Analysis of Images, Social Networks and Texts* (pp. 247–258). Springer International Publishing.
32. Vlassopoulos, G., Erven, T. van, Brighton, H., & Menkovski, V. (2020). *Explaining Predictions by Approximating the Local Decision Boundary*. <https://arxiv.org/abs/2006.07985>

About the authors

Tomislav Medić is a Lecturer at Faculty of Economics and Business, University of Zagreb, Croatia. He received his master's degree in Managerial Informatics from the Faculty of Economics and Business in Zagreb in 2019. He is currently attending doctoral studies at the Faculty of Economics and Business – Zagreb. He was employed as an assistant at the Department of Informatics, Faculty of Economics, University of Zagreb from November 2019 to June 2020 and participated in conducting seminar classes in several courses. Since March 2021, he has been employed at the Department of Informatics, Faculty of Economics, University of Zagreb as a lecturer and participates in teaching in the field of knowledge discovery from databases, system dynamics modelling, business information systems, informatics, business application development, artificial intelligence in business and cyber security. Tomislav Medić participated in several professional projects and has published several scientific papers. The author can be contacted at tmedic@net.efzg.hr

Antonio Pavlečić is currently an employee of the Ministry of Physical Planning, Construction and State Property in the field of housing policy and affordable housing. He completed his undergraduate studies at the Faculty of Economics and Business in Zagreb in September 2012. He graduated in April 2014 from the Faculty of Economics and Business in Zagreb, majoring in trade. In May 2019, he graduated from the diplomatic academy as part of his studies at the Ministry of Foreign and European Affairs. Since November 2021, he has been attending doctoral studies in Economics and Global Security at the Faculty of Economics and Business in Zagreb. He has built his professional career through several state institutions of the Republic of Croatia, among which the Ministry of Economy and Sustainable Development and the Ministry of Physical Planning, Construction and State Property stand out. The most prominent position he held was Secretary of the Cabinet of the Minister in the Ministry of Economy and Sustainable Development. He performed tasks related to the activities of the Minister and State Secretaries, coordinated the preparation and supervision of materials for the Government of the Republic of Croatia, the Ministry of Foreign and European Affairs, the Croatian Parliament and other state bodies and legal entities with public authority, especially the emphasis on the energy, environmental protection and economy sectors. He was also an alternate member of the National Security Council in the field of cybersecurity. You can contact the author at antonio.pavlecic@mpgi.hr.

Amir Topalović received his master's degree in finance and risk Management at the University of Parma and he is actually a PhD student in Business and Administration at the University of Sarajevo under the supervision of Prof. Nijaz Bajgoric. His thesis title is "A TOE based Study on Data Mining employment in Italian Small and Medium Enterprises". Currently, he is CEO at Aisma S.L.R., and his research interests are related to data mining in corporate and banking data. The author can be contacted at amir.topalovic@aisma.it