

# Predicting Bank Term Deposit Subscription Using Machine Learning Models

Yanxi Wang \*

Business School, University of Bristol, Bristol, United Kingdom

\* Corresponding Author Email: jo24051@bristol.ac.uk

**Abstract.** Banks are increasingly relying on data-driven target positioning to transfer potential customers into product buyers. By effectively identifying customer characteristics, one can reduce their marketing costs. However, converting purchase propensity models into measurable data remains challenging. This article integrates the advancements in tabular data modeling and marketing analysis to bridge this gap. It first provides an overview of the banking context and dataset characteristics, and then compares the propensity evaluation of basic linear models with ensemble methods, emphasizing the role of interpretability in compliance and on this basis, multiple models were established and evaluated, then the most suitable model was selected based on the data sets. Besides, this article discusses evaluations beyond accuracy: by using ROC/P&R analysis to link threshold selection with operational trade-offs. To support targeted actions, customer segments are defined, groups that are price-sensitive are identified for tailored interventions, so that the customer value are optimized. Thus, this article provides a practical perspective for transfer predictive data into profitable and auditable banking activities.

**Keywords:** Bank Marketing, Purchase Propensity Prediction, Machine Learning, Interpretability.

## 1. Introduction

Recently, bank markets were facing increasing harsh competition and the diversification of customers' demand. For efficiently utilizing marketing resources and increasing product conversion rates, more and more researchers choose to use the efficient and explainable analysis method of predicting customers' purchasing tendencies.

According to Basten and Juelsrud's research, deposits significantly increased the probability of purchasing other financial products in the future. This shows that cross-selling with analysis of customers have important value in banks' strategy [1]. By this way, if we can get a precise prediction of how the customer turns into products like term deposit and investment products from basic products such as balance of their account. And at the same time, Boustani et al. pointed out that bank customer behavior data is an important dimension to reflect customers' preference and is more predictive than simple static characteristics. They also mentioned recently research have also begun to use machine learning methods and incorporate customer transaction data to improve the accuracy of cross-selling models [2]. However, loads of institutions still relieve simple clustering models, and there is a lack of the application of machine learning and interpretable models in terms of customer purchase tendencies. All these always leads to the marketing goals are not precise and the cost to require customers is relatively high. This proves that in recent years, research in the financial market has begun to emphasize the interpretability of machine learning models, especially in high-risk contexts such as customer behavior prediction. For instance, Rudin et al. proposed a decomposable risk model, with visualization tools and explanatory summaries which can enhance the transparency and auditability of the model [3].

According to the situation above, with the requirement of high accuracy and facing constraints such as regulation, combining customer behavior data with machine learning models, and conducting interpretability analysis is the current cutting-edge trend, just like Chen et al. mentioned: Interpretability is an essential prerequisite for the practical application of financial machine learning.

So, in this research, this paper is going to set up prediction framework based on machine learning, finding the most suitable model under customers with different characteristics and behavior. Besides,

the research did model interpretability analysis which can explain how and why the model makes decisions and increasing the model transparency. And how these could give banks ideas on how to make marketing decisions. What's more, the paper is going to talk about how to combine the purchase intention with the value and potential of the customers and then helping them make marketing priority and resources allocation.

## **2. Overview of Mainstream Technology**

In the task of predicting the purchase tendency of bank customers, the mainstream modelling methods are mostly based on machine learning classification models. There are significant differences in model structure and interpretability between different algorithms. Therefore, comparing and analysing the basic principles of these models and their advantages and disadvantages can help this study select the most suitable prediction framework for bank marketing scenarios.

The research in this article involves five machine learning models: Logistic Regression, SVM, Random Forest, KNN and Decision Tree.

### **2.1. Interpretable Model**

Logistic Regression is a linear classification model that maps the results to the function by linearly weighing the input variables to output the probability value. Logistic Regression has strong interpretability and high efficiency, which is suitable for structured data scenarios. However, due to the linear relationship between its characteristics and the results, it has limited performance in dealing with nonlinear or complex interaction features.

The Decision Tree model takes the characterization as the node and divides the sample space through indicators such as information and then forms a tree-like decision-making structure. The structure of this model is intuitive, easy to explain, and can handle nonlinear relationships. However, its disadvantage is that it is easy to over-fit, sensitive to the small disturbances, and the generalization performance of a one tree is usually limited.

Chen et al. pointed out that in financial risk forecasting, transparent and explainable models such as logistic regression have significant advantages in compliance and trustworthiness and are not necessarily inferior to complex black box models in accuracy [4].

### **2.2. Random Forest**

Random Forest is an integrated learning model composed of multiple decision trees. It significantly improves the stability and accuracy of the model by introducing randomness at the sample and feature levels and voting on the results. And it has strong anti-fitting ability, can handle high-dimensional characteristics, and has good generalization ability. However, its interpretability is weak and cannot clearly show the direct relationship between variables and results.

### **2.3. SVM and KNN**

These are two more complex models, which have more advantages in the accuracy of output results.

The basic principle of SVM is to deal with nonlinear classification problems by finding an optimal hyperplane to maximise the sample spacing of different categories. SVM has high classification accuracy and strong generalization ability, which is especially suitable for high dimensional data sets. However, it is relatively complex and sensitive to parameter selection. In addition, the decision-making boundaries of it are difficult to explain, and it is not easy to meet the requirements of model transparency of business requirements.

KNN based on the assumption that 'similar samples have similar results', selecting the nearest K neighbors for voting prediction by calculating the distance between the sample to be predicted and the sample in the training set. Its advantage is that it is simple to realize, does not require training, but is sensitive to data scale.

### 3. Experimental Results

#### 3.1. Data Description

The dataset used in this study is the Bank Marketing dataset from the UCI Machine Learning Repository which contains 11,162 records and 17 attributes. Among them, 7 are numerical features (e.g., age, balance, duration) and 10 are categorical (e.g., job, education, marital). The target variable deposit indicates whether a client subscribed to a term deposit (“yes” or “no”) [5].

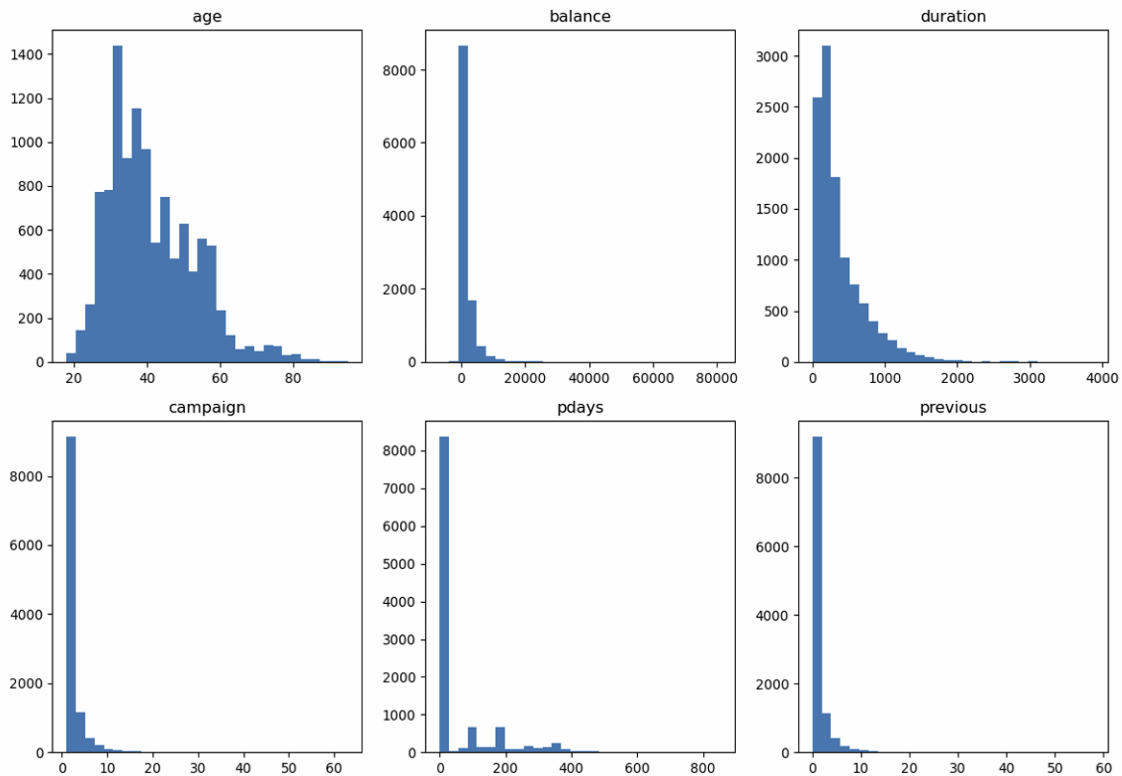
#### 3.2. Processing Data

In the initial data processing, data integrity checks, data identification and feature classification, target variables distribution analysis, and feature statistics were conducted.

Through the processing, there is no missing data, and the range of each feature is reasonable, but the value range and standard deviation of ‘balance’, ‘duration’ and ‘pdays’ are too high, which means these features need to be standardized. Meanwhile, the category proportion of the target variable was got to the balance (no – 0.526, yes – 0.474). Besides, due to the sample size is big enough, this data is highly suitable for binary classification modeling.

At the same time, through the observation of the mean and standard deviation, the coefficient of variation (CV) of balance ( $\approx 2.11$ ) and duration ( $\approx 0.93$ ) can be got, this means that these two features have significant dispersion, so it has greater potential to distinguish different customers.

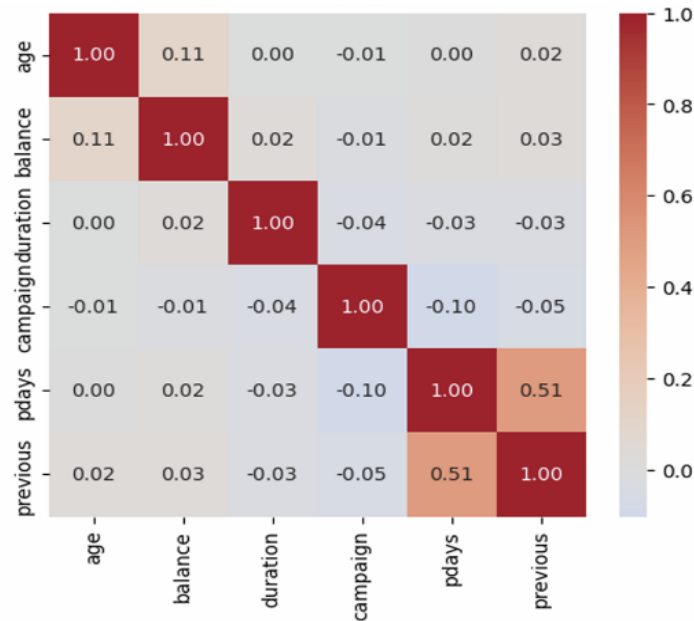
#### 3.3. Visualization of Numerical Feature Distribution



**Fig 1.** Distributions of Key Numerical Features.

In Figure 1, it can be noticed that some features obviously have extreme values. In particular, the maximum value of ‘balance’, ‘duration’, ‘campaign’, ‘pdays’, and ‘previous’ are much higher than the 75% percentile, and the distribution is very skewed to the right, which means extreme samples existed. This kind of situation is also mentioned by Cowan et al., it’s normal in banks’ data. In the industry, extrem value impacts are usually mitigated through annualized aggregation and segmented modeling, thereby enhancing the stability of subsequent clustering and transfer predictions [6].

### 3.4. Correlation between Numeric Variables



**Fig 2.** Person Correlation Matrix.

**Table 1.** Correlation between Numerical Variables.

Variable 1:	Variable 2:	Correlation
Previous	Pdays	0.507272
Balance	Age	0.112300
Pdays	Campaign	0.102726
Previous	Campaign	0.049699
Campaign	Duration	0.041557

In this part this article constructed a Person correlation matrix to examine the linear relationship among numerical variables, like Figure 2 and the results are shown in Table 1. Through the heatmap, it can be found out that most pairs of variables are weakly correlated ( $|r| < 0.1$ ), except ‘pdays - previous’ ( $|r| = 0.51$ ) and ‘age - balance’ ( $|r| = 0.11$ ), the former show a moderate positive correlation and the latter show a weak positive correlation. Besides, duration is almost completely independent of others, which means its effect can be independently reflect the models, it is one of the most important predictive features. In conclusion, there is almost no multicollinearity problem among the various features.

**Table 2.** Correlation between Numerical Variables and The Target Variable.

age	$r = 0.035$	$p = 0.000226$
balance	$r = 0.081$	$p = 9.13e-18$
duration	$r = 0.452$	$p = 0$
campaign	$r = -0.128$	$p = 4.83e-42$
pdays	$r = 0.152$	$p = 2.27e-58$
previous	$r = 0.140$	$p = 7.13e-50$

In Table 2 the correlation between each numerical variable and the target variable deposit was calculated. The conclusion is: ‘duration’ is the most important feature, which has the strongest positive correlation with whether the customer decides to make a purchase. Besides, ‘campaign’ has negative correlation, and all p values less than 0.001, which means all these relationships have statistical significance.

### 3.5. Multi-model Training Evaluation

**Table 3.** Model Performance Summary.

	model	accuracy	precision	recall	f1	approved
0	SVM	0.855352	0.827832	0.877127	0.851767	
1	Random Forest	0.853112	0.823009	0.879017	0.850091	
2	Logistic Regression	0.826243	0.822115	0.808129	0.815062	
3	KNN	0.808777	0.817724	0.767486	0.791809	
4	Decision Tree	0.780116	0.772334	0.759924	0.766079	

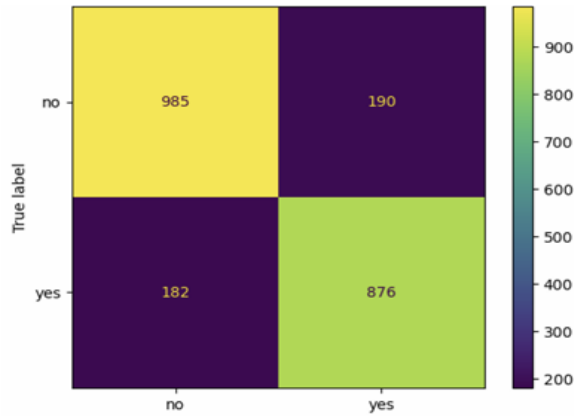
Gerling and Lessman pointed out that the field of banking marketing analysis is gradually developing from a single algorithm to multi-model integration and hybrid learning, which makes multi-model comparison experiments practical [7]. By this way, in order to compare the predictive capabilities of different models and verify the stability of the model, this research set up a multi-model training, and all the results are shown in Table 3. All model's accuracy is between 0.78- 0.86, and F1 scores are all above 0.75, which shows that the overall framework can accurately predict whether customers will purchase term deposits.

By comparing all the models, the paper can discover that the SVM and the Random Forest have best performance, which indicates that the customers' purchasing tendency is not simple linear relationship but rather more complex.

On the other hand, when searching for the task of marketing analytics that the data are structured and interpretable, logistic regression seems better. As noted in research, for tabular data, complex models rarely performed better than the Logistic regression in predictive accuracy [3]. In this point we can also noticed in the results: the predicted performance of SVM and Random Forest is only slightly higher than Logistic Regression (about 2.5-4%), which means that most of the data can be explained by the linear relation. And in reality, banks may be more curious about what features can affect the customers, just as Biehler et al. pointed out that the banking industry is subject to regulatory and system restrictions and usually tends to choose explainable models such as logical regression or shallow decision trees. Even if complex models can slightly improve accuracy, it is difficult to meet the requirements of regulatory transparency [8]. This is the reason why this research chose logistic regression but not SVM and RF.

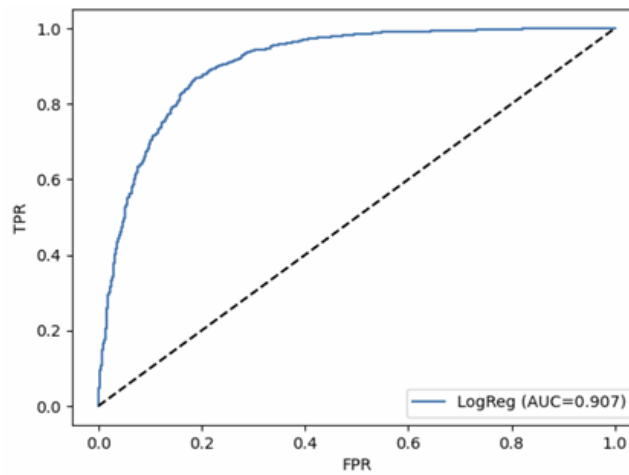
Overall, although SVM and RF indicate that there are some nonlinear characteristics, the improvement about it is limited. From the perspectives of interpretability and compliance, shallow interpretable models are more suitable for implementation in banking scenarios. Meanwhile, using tree models as the base line and integrating them with deep models on a small scale when necessary is a good choice that balances accuracy and cost [9]. Therefore, in this project, it is more appropriate to use the Logistic Regression model.

So, in the next step, the Confusion Matrix was made to see the predicted results as Figure 3. From the results, it can be noticed that models make relatively accurate predictions for both types of samples. And because the number of FT and FF is close, the model does not make any significant bias towards both categories. By this way, we can conclude that the logistic regression model performs stably and has highly accuracy and balance in recognizing whether the customers will buy or not.

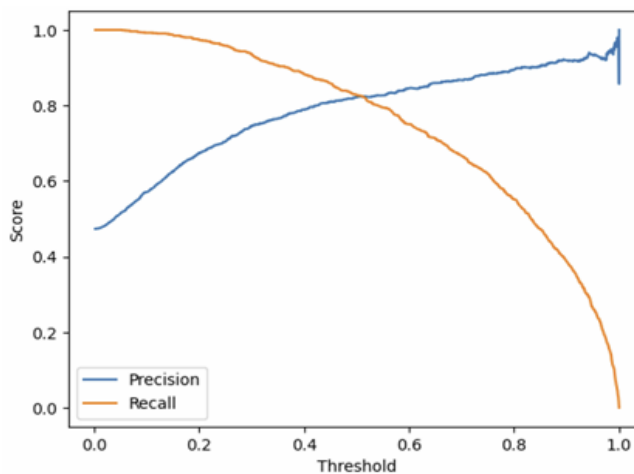


**Fig 3. LogReg Confusion Matrix.**

### 3.6. Deep Analysis



**Fig 4. ROC Curve.**

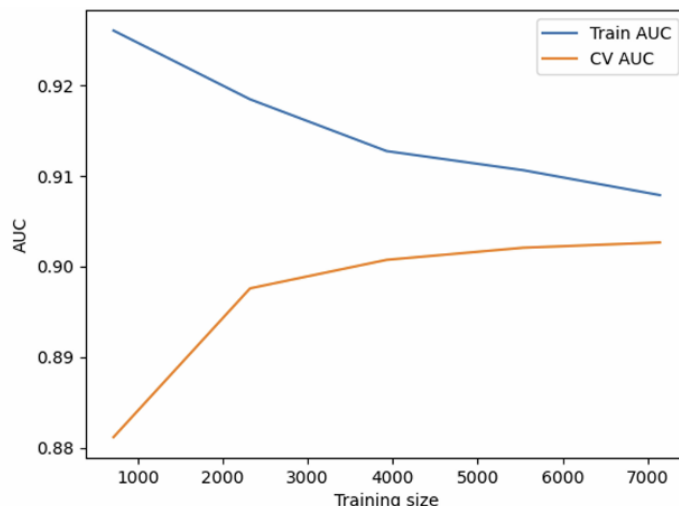


**Fig 5. Precision & Recall vs Threshold.**

In this step, to evaluate the model’s performance stability, the sensitivity to thresholds of probability and explain the feature contribution. Firstly, the ROC Curve (Figure 4) was produced. According to the ROC curve, the curve is higher than the random curve, and the area under the curve (AUC) equals 0.907, which is higher than 0.9. In this case, the model performed stable and reliable. It also proved that logistic regression is very reliable in prediction.

Secondly, the relationship between the changes in precision and recall under different probability thresholds showed as Figure 5. When thresholds nearly to 0.5, precision  $\approx$  recall  $\approx$  0.82, at this point precision and recall get into a balance. From this case, it can be found a high recall and low precision

at a low threshold and get a low recall and high precision at a high threshold. This means this model can be changed to meet different banks' business requirements by changing the threshold. For instance, if the bank wants to expand marketing coverage, it can use low threshold. On the contrary, banks can use high threshold to save marketing costs and increase conversion rates. In real marketing activities, in addition to ROC/PR, marketers usually use improvement or impact analysis to verify the business impact, and accordingly select the Top-k cutoff points, which link the threshold to the incremental benefits directly [10].

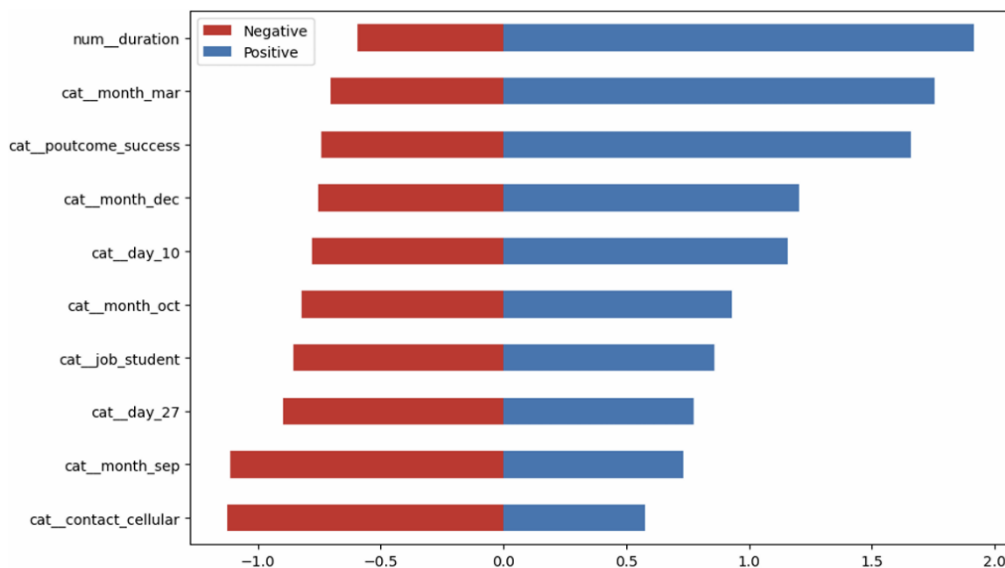


**Fig 6.** Learning Curve.

Thirdly, making a learning curve. In Figure 6, with the increase of the training sample size, the train AUC tends to 0.91 and the Cross-Validation AUC tends to 0.90, only a small gap between these two AUC. Besides, CV AUC shows an upward trend and tends to be stable. These prove that the model is not overfitted, it can learn from more samples and the logistic regression is stable and has strong generalization ability.

In the end, the article extracts the regression coefficients of logistic regression and got a visualization of the top 10 most significant features. In Figure 7, it can be found that the longer the call duration, the higher the probability of purchase, if last marketing is success, the customers are more likely to purchase, and so on.

The analyzation of the regression coefficient verified the previous conclusion again that the duration and the outcome (outcome of the last marketing) are still the key features that have the greatest impact on customers' purchasing will.

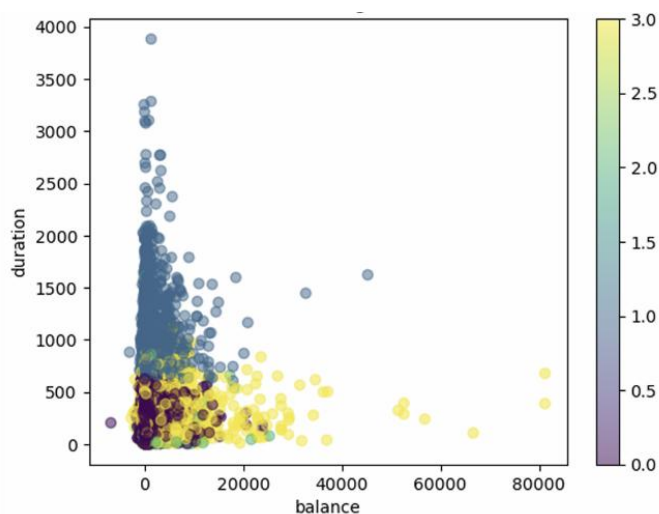


**Fig 7.** Top Positive and Negative Coefficients.

### 3.7. Customer Segments

In the last part, this research did the customer segment by using K-Means clustering. From the previous analysis (3.1&3.3), we already know that duration and balance these two variables are highly sensitive to customers' purchasing tendency. For another, when viewed from the perspective of business, these two variables measure the wealth level and reflect customers' interest and participation in marketing communications. In this way, it is a good idea to segment customer group by these two variables.

Besides, in this study, the number of clusters (k=4) was selected based on business interpretability and the inspection of data distribution. Four clusters can give the bank a simply model and make them efficiently distinguish the customer group.



**Fig 8.** Customer Segmentation.

In Figure 8, some typical groups can be found, like low balance and short duration groups, high balance and long duration groups and medium groups. The first one means this group has low value for the bank, the second one means this group has high value and high intentions for the bank, the last one is the price-sensitive groups that banks need to use some methods like interest rate or gift incentives to cultivate.

## 4. Discussion

Through data analysis, it was found that call duration and the account balance were the most influential features.

Among the multiple models tested, SVM and Random Forest performed slightly better than the Logistic Regression, but the difference was very small, which indicates that most relationships are linear. In this case, liner models such as Logistic Regression are more suitable. The AUC achieved 0.907, showing that LR model was stable and interpretability.

The results of Figure 7 further indicated call duration and success in last marking had the most influence on the purchasing will, but the times and student\_job were negative factors.

Finally, K-Means clustering (k=4) based on duration and balance identified different customer groups with different purchasing tendencies and can help banks to concentrate on their marketing resources.

## 5. Conclusion

This study shows that the logical regression model is both high-precision and interpretable in the prediction of the purchase tendency of bank customers and is suitable for business decision-making. But there are still many challenges and points for improvement:

In terms of the limitation of data, this study only uses structured data and does not make full use of unstructured information such as communication text, social behavior or transaction time with customers. In the future, it can be combined with deep learning models to improve prediction ability.

For the model generalization, the characteristics of customers in different banks and market environments vary greatly, and the cross-environmental applicability of the model needs to be enhanced through multi-source data or migration learning.

Besides, this kind of handing method has good interpretability. Future research can explore knowledge distillation or use explainable enhancement models, so that complex models can improve accuracy while maintaining transparency.

For Business utilization, it can combine customer value with building a personalized marketing strategy and realise the effective transformation of model results into business practice.

In summary, in the future, continually optimizing the three aspects of model integration, data expansion and practical application transformation to realise a more intelligent and explainable banking marketing decision-making system is a good way.

## References

- [1] Basten C, Juelsrud R. Cross-Selling in Bank-Household Relationships: Mechanisms and Implications for Pricing. *The Review of Financial Studies*, hhad062. 2023.
- [2] Boustani N, Emrouznejad A, Gholami R, Despici O, Ioannou A. Improving the predictive accuracy of the cross-selling of consumer loans using deep learning networks. *Annals of Operations Research*. 2023, 339: 613–630.
- [3] Rudin C, Chen C, Chen Z, Huang H, Semenova L, Zhong C. Interpretable machine learning: Fundamental principles and 10 grand challenges, *Statistics Surveys*, 2022, 16: 1–85.
- [4] Chen C, Wang T, Lin K, Rudin C, Shaposhnik Y, Wang S. A Holistic Approach to Interpretability in Financial Lending: Models, Visualizations, and Summary-Explanations, *arXiv preprint arXiv:2106.02605*. 2022.
- [5] Kaggle. <https://www.kaggle.com/datasets/janiobachmann/bank-marketing-dataset>. 2025.
- [6] Cowan G, Mercuri S, Khraishi R. Modelling customer lifetime-value in the retail banking industry. *arXiv:2304.03038*. 2023.
- [7] Gerling C, Lessmann S. Leveraging AI and NLP for Bank Marketing: A Systematic Review and Gap Analysis, *arXiv:2411.14463*. 2024.
- [8] Biehler M, Guermazi M, Starck C. Using Knowledge Distillation to improve interpretable models in a retail banking context, *arXiv:2209.15496*. 2022.
- [9] Shwartz-Ziv R, Armon A. Tabular Data: Deep Learning is Not All You Need, *Information Fusion*, 2022, 81, 84-90.
- [10] Arevalillo J M. Data science methods for response, incremental response and rate sensitivity to response modelling in banking, *Expert Systems*, 2024, 41(10), e13644. <https://doi.org/10.1111/exsy.13644>.