Original Scientific Paper https://doi.org/10.62598/9thICVA.031



# PERSONAL BANK LOAN APPROVAL FACTORS: A DETAILED EXPLORATION

# ČIMBENICI ODOBRAVANJA OSOBNOG BANKOVNOG ZAJMA: DETALJNO ISTRAŽIVANJE

Xhoxhi, Olsi, Faculty of Economy, University "Ismail Qemali" of Vlora, Albania, olsi.xhoxhi@univlora.edu.al

Dede, Grigor, Faculty of Economy, University "Ismail Qemali" of Vlora, Albania, grigor.dede@univlora.edu.al

Hysi, Arben, Faculty of Economy, University "Ismail Qemali" of Vlora, Albania, grigor.dede@univlora.edu.al

Abstract: This study is going to investigate the complex relationship between demographic and socioeconomic factors and their impacts on bank loan approval, combining decision tree classification with feature importance assessment. The paper focuses on key variables such as age, gender, marital status, level of education, monthly income, loan amount, and interest rate within a robust dataset from financial institutions. Data is collected through questionnaires to 400 people in the city of Vlora, Albania. The present study uses conjointly advanced statistical techniques, particularly decision tree classification, in establishing hierarchical patterns and interactions among these variables that shed light on their influence on loan approval outcomes. In addition, feature importance assessment is used to yield the relative importance of the variables in the prediction of loan approval. The results light up not only the differential effects of demographic and socioeconomic attributes on loan approval but also the prioritization of the most influential factors driving lending decisions. In such a way, this approach provides insights for the optimization of financial institutions' lending practices and, from a broader perspective, for policymakers to promote equal access to credit to include economic development.

**Keywords**: bank loan approval, decision tree classification, demographic and socioeconomic factors, feature importance

Sažetak: Ova će studija istražiti složen odnos između demografskih i socioekonomskih čimbenika i njihovih utjecaja na odobravanje bankovnih kredita, kombinirajući klasifikaciju stabla odlučivanja s procjenom važnosti obilježja. Rad se fokusira na ključne varijable kao što su dob, spol, bračni status, razina obrazovanja, mjesečni prihod, iznos kredita i kamatna stopa unutar robusnog skupa podataka financijskih institucija. Podaci su prikupljeni putem upitnika za 400 ljudi u gradu Vlora, Albanija. Ova studija koristi zajedno napredne statističke tehnike, posebno klasifikaciju stabla odlučivanja, u uspostavljanju hijerarhijskih obrazaca i interakcija među ovim varijablama koje bacaju svjetlo na njihov utjecaj na ishode odobravanja kredita. Osim toga, procjena važnosti značajki koristi se za dobivanje relativne važnosti varijabli u predviđanju odobrenja kredita. Rezultati osvjetljavaju ne samo različite učinke demografskih i socioekonomskih atributa na odobravanje kredita, već i određivanje prioriteta najutjecajnijih čimbenika koji pokreću odluke o kreditiranju. Na takav način, ovaj pristup pruža uvid u optimizaciju prakse kreditiranja financijskih institucija i, iz šire perspektive, kreatorima politika da promiču jednak pristup kreditima kako bi uključili gospodarski razvoj.

*Ključne riječi:* odobravanje bankovnog kredita, klasifikacija stabla odlučivanja, demografski i socioekonomski čimbenici, značaj obilježja

## 1. Introduction

Personal bank loans are one of the most imperative financial instruments in today's society, ensuring that people receive financing to cover needs such as education, buying a home, or debt consolidation. The process of obtaining a personal bank loan undergoes various stages; one of the most important stages of all is the approval stage. Personal bank loan approval is not as simple as filing out an application; rather, it applies to a multifaceted evaluation of many factors that determine the decision of a lender. These range from credit history and income to the economical environment conditions and government policies.

Understanding the multifaceted nature of personal bank loan approval is in itself critical both to borrowers and financial institutions. First, to the borrowers, knowing the factors affecting loan approval is critical in facilitating effective financial planning and increasing the chances of securing good loan terms. For the financial institutions, a deeper understanding of these factors facilitates more accurate risk assessment, leading to better lending practices and lower default rates.

The topic "Personal Bank Loan Approval Factors: A Detailed Exploration" summarizes the general objective of this research study, which is to explore the complexities of personal bank loan approval and reveal the many variables that tend to influence this process. The paper does this through a detailed analysis of the literature, statistical analysis, and possibly empirical studies with the intention of shedding valuable insight into the dynamics of personal bank loan approval.

Before proceeding with the factors in personal bank loan approval, a contextual backdrop is needed regarding the significance of personal loans within modern financial systems. Personal loans or unsecured loans refer to credit given to individuals by financial institutions without any collateral security requirement. They can be used for a wide array of purposes, such as debt consolidation, home improvement, medical expenses, or even to finance a vacation.

Demand for personal loans has been at an increasing trend over the years, drawing from various socio-economic factors. People have to meet the increasing costs of living, increasing consumerism, and changing lifestyles. Besides that, digital lending platforms and the easing of loan application procedures have made access to personal loans much easier to reach a larger part of the population.

With the growth in demand, competition in the personal lending market also became fiercer, with more and more financial institutions competing for market share. As a result, the industry developed new loan products, competitive pricing strategies, and approval processes as the lenders sought to distinguish their products and services to attract more borrowers.

While there is abundant literature on personal finance and credit risk assessment, there lies a gap in research focused on the factors affecting personal bank loan approval. While some studies have looked into individual aspects, such as credit scoring models or demographic variables, few have provided a comprehensive analysis of the holistic approval process.

Understanding the reasons why a lender will approve or reject a personal loan application is important for several reasons: enabling borrowers to tailor their applications according to the criteria which the lender places the most value on, hence improving their probabilities of approval; allowing transparency in the practices of lending, fostering trust between financial institutions and consumers; and enabling policy and regulatory bodies to assess the effectiveness of current regulations on lending and point out gaps that need to be improved.

This is why the current research wishes to probe in detail the factors of approval for personal bank loans in an attempt to fill up the lacuna left in the literature and make substantial contributions to academia, industry practitioners, policymakers, and consumers. This paper will try to determine all the various factors of loan approval through a systematized analysis of the determinants underlying the personal lending market in order to enhance our understanding of the dynamics at play in the personal lending market and provide practical recommendations for both borrowers and lenders.

### 2. Theoretical and conceptual background/framework

This section will comprehensively cover the multifaceted realm of personal bank loan approval factors. We will bring into the open the intricacies that drive the decisions of financial institutions from a varied array of studies by the respective esteemed authors. From credit history to stability in income, and a lot more, we will look in-depth into the variables that shape approval or rejection in loan applications. Join us on a journey through many scholars' findings as we go forth into the world of personal finance with precision and depth.

In their paper, Dansana et al. (2023) highlight the importance that loans play in enhancing the income of the financial sector; however, they also expose the enormous amount of financial risk associated with them. Focusing on the interest generated from loans as one of the greatest bank assets, they also stress that the demand for loans is increasing globally, which, in turn, forces organizations to develop more effective business strategies to attract clients. Although several loan applications are received in organizations daily from all walks of life, not all of them are approved because of the possibility of default. Cases of defaults, though not unknown, highly cost the banks. The most important point in their paper is how to determine whether to offer loans to certain individuals or organizations. Using the Random Forest Regressor model to evaluate both performance and suitability, the paper suggests that banks should diversify and not remain embedded in wealthy clients but instead consider various customer features that are critical to credit scoring and default prediction. Their paper explores multiple parameters in the loan approval process, including gender, level of education, employment status, nature of business, the period for which a loan is taken, and marital status. In addition, the paper critically analyzes the statistics of the approved, disbursed, and rejected loans, and these provide valuable insights into the loan approval and prediction process.

Meshref (2020) expounds on the predominant use of the Bank Marketing dataset available on Kaggle, which is primarily used to predict the probability that bank clients will subscribe to long-term deposits. They however argue that the dataset can provide far deeper insight in the area of predicting the approval of loans, which is one of the critical decisions made by those holding positions in banking leadership. They point out that high model accuracy is key to making reliable predictions and the interpretability of these predictions, particularly since the decision-making about loan approval is very crucial. Meshref's study applies different ensemble machine learning techniques such as Bagging and Boosting, which lead to a loan approval prediction model of an accuracy rate of 83.97%, some 25% above state-of-the-art models. Importantly, the efforts in the interpretability of the model shed light on critical scenarios that decision-makers within the bank face, instilling confidence in the predictive abilities of the model. To Meshref, both accuracy and interpretability are imperative to achieve a very delicate balance between security and reliability in financial lending systems to ensure that decision-makers are well-equipped to sail through complex risk management landscapes.

The research undertaken by Azam et al. (2012) seeks to establish the impact of various socio-economic attributes of loan applicants on personal loan decisions. They tested six hypotheses: region, age, gender, income, residence status, and years with the current organization, about which a candidate sends his/her application. Their analysis of the data showed that region, age, gender, income, residence status, and years at the current organization positively impact the decision to accept or refuse a loan. The results show that regions with the highest population, such as Karachi, were most likely to give out loans. The applicant above the age of 40 years would have lower success rates. Besides, income

was a crucial factor, which, with its increase, increases the possibility of disbursing the loan. Also, the place of residence, including whether it is rented or owned, and the duration of residence lowered the risk of default. However, this does not appear to be the only factor that dictates whether the loan would be disbursed or not because other factors, such as bad credit history and those whose debt burden ratios were higher, were considered. The authors especially underlined the role of the region, residence status, and years with the organization for a good prediction of the loan decision. They added that the companies should treat those variables as the most important that significantly impact on making decisions based on their analysis.

In his study, Himali (2020) discusses a phenomenon known as Loan Default, which he defines as a failure of the applicants of loans to settle their loan obligations. The research intended to explore what determines personal loan default, contrasting between the effectiveness of the proportional hazards model and that of random survival forests models. Using data collected from 1500 customers of a large financial institution in Sri Lanka through questionnaires, Himali established that the personal loan default was determined by the factors attributed to the customer, such as occupation, monthly income, and loan purpose. The Random Survival Forest listed monthly income, occupation, loan purpose, and loan amount, while the Cox Proportional Hazard model showed that other liabilities and payment frequency were prominent in relation to personal loan default. In his conclusion, Himali stresses the need for government intervention in dealing with economic strains not only to facilitate economic growth but also to reduce the factors related to the customer, which cause loan default.

According to Tiwari & Bapat (2020), the study sheds light on the practice of community policing by private commercial banks through the practice of loan appraisals, pledged lending, sending reminder notes to defaulters, credit follow-up, and planning for loan and credit protection. They enumerate several causes of credit crime, to wit: bad selection of loan clients, insufficient follow-up, tight bank competition, inadequate appraisal of project feasibility, high rate of interest, insufficient funds, financial instability, and lack of symmetry in information. To complement these findings, the study also enumerates the strategies to remedy the loan performance, such as strict observance of regulations, calling efforts after disbursement, careful appraisal of securities, regular guidance to clients, lending to highly reputable businesses, calling efforts after loan, competitive rates of interest, and a comprehensive documentation before disbursement. The data was gathered from 79 questionnaires administered to bank clients and subsequently, SPSS analyzed. Furthermore, the study emphasizes that investor-owned businesses significantly affect loan repayment.

In their research, Wang & Xie (2023) investigate the increasing trend of personal loans with growth in the economy, conceding an increase in spending. By using factor, t-test, and ANOVA tests in their analysis, they found out that age, experience, and above all, income, are the significant factors that would affect personal loan activities, indicating that increased personal consumption is related to an increase in personal loans. Their findings point to income as a critical determinant of personal loan decisions and provide a much-needed contribution to the issue of how banks can best persuade customers to patronize their services and give them more business. In their application of techniques such as factor analysis and comparative analysis, they further establish that all the variables that are relevant to the model, in particular income and CCAvg, are relevant to personal loan activities, thereby enhancing the fulfillment of the purpose and objectives of the study. They call for the consideration of undergraduate students and small family sizes as targets in this group, considering that even the presence of mortgage loans influences personal loan decisions, especially for large mortgage holders in order to live a better life.

Alagic, et al. (2024) note in their study that loan requests in the world have been rising exponentially; hence, there is an enormous stake in the sector for credit approval. Despite the rich information on

customer behavior gathered from banking transactions, the process of approving loans by financial institutions is still quite tiring and highly complicated. In the year 2022, over 20 million Americans were in debt, owing about USD 178 billion, yet over a fifth of applications for loans were rejected. The authors use machine learning techniques for estimating loan risks to augment traditional statistical methods. Using mental health and loan approval datasets that contain responses from 1991 individuals, they survey different machine learning algorithms that predict credit risk. Their analysis has comprehensively established that XGBoost outperforms the rest with an 84% accuracy rate in dataset one, while for dataset two, it is the random forest. In addition, precision, recall, and other metrics assessing the general algorithmic performance also support XGBoost and random forest as the two with better predictive capability in the respective datasets. These results have implications for researchers and industry practitioners in the refining and optimization of classification models for loan risk assessment.

Artificial intelligence algorithms and ML models have pervaded most contemporary fields over the past few decades, from industry to education, healthcare, and entertainment. In the paper written by Végh, et al. (2023), the focus was on the use of ML algorithms for the bank loan approval process. This paper commences with a concise overview of the current state of literature on loan approval prediction using ML models. The authors then make use of the loan approval prediction dataset retrieved from Kaggle to serve as the benchmark in the evaluation of a number of ML classification models. They highlight that credit scores and loan terms are the two major driving forces. Moreover, the dataset is divided into an 80% train set and a 20% test set. They then train 27 different ML models in MATLAB where three models are further optimized using Bayesian Optimization to optimize the hyperparameters for the minimization of the error. To avoid overfitting, 5-fold cross-validation is used in the train process. Afterward, the performance of the trained models on the test set is evaluated, which shows their performance on unseen data. Validation and test data reveal that these models have the best accuracy, exceeding 98%, using neural networks and ensemble classification models.

## 3. Methodology

The research methodology applied in this paper involved a critical exploration of key variables influencing the approval of personal bank loans. The sample size was determined using the Slovin formula, resulting in a cohort of 400 individuals. We used a decision tree classifier, a strong machine learning algorithm with regard to interpretability and handling of categorical data, to predict whether or not a person's loan would be approved. Variable selection was done based on their assumed importance in loan approval processes. Such variables include age, gender, marital status, level of education, monthly income, loan amount, and interest rate, all of which are key inputs in the decision-making processes of lending institutions. We wished to try to shed light on the dynamic of the loan approval processing, and feature engineering to create a predictive model that captures the essence of the factors and provides valuable insights for both borrowers and lenders. This provides the methodological bedrock upon which the underlying complex and multifaceted determinants of personal bank loan approval processes can be subjected to a rigorous examination with very important implications for both financial institutions and policymakers, as well as the people who seek financial assistance.

### 4. Results

In this section, we present the results of our research that had the main purpose of making a detailed investigation in factors that influence the acceptance of personal bank loans. Our research elaborates on the complex dynamics that shape the decision-making process within lending institutions, focusing particularly on defining the key drivers that influence the results of loan approval. Before we start explaining and interpreting our findings, it is worth pointing out the nature and structure of the data under analysis.

Table 1 comprises an overview of variables, each comprising 400 non-null entries. The variables are: Loan Status, Age, Gender, Marital Status, Education Level, Monthly Income, Loan Amount, and Interest Rate. Loan Status, Gender, Marital Status, and Education Level are categorical variables represented as objects, Age and Monthly Income are integers; Loan Amount is similarly represented as an integer, with Interest Rate as a floating-point number. All these characteristics give a better view of the data and represent the nature of the data in the best possible way for further analyses and interpretations.

Nr	Column	Non-Null Count	Dtype		
0	Loan Status	400 non-null	object		
1	Age	400 non-null	int64		
2	Gender	400 non-null	object		
3	Marital Status	400 non-null	object		
4	Education Level	400 non-null	object		
5	Monthly Income	400 non-null	int64		
6	Loan Amount	400 non-null	int64		
7	Interest Rate	400 non-null	float64		
dtypes: float64(1), int64(3), object(4)					

Table 1. Overview of Data Types and Non-null Values

Source: Author's calculation

The data depicted by means of the heatmaps in Figure 1 is a visualization of educational attainment across demographics specifically, gender and marital status. For females, the heatmap includes 52 high school graduates, 74 bachelor's degree holders, and 72 master's degree recipients. Conversely, males include 58 high school graduates, 72 bachelor's degree holders, and 72 master's degree earners. By marital status, singles are characterized by 58 high school graduates, 60 bachelor's degree earners, and 62 master's degree earners. On the other hand, married, by marital status, is characterized by 52 high school graduates, 86 bachelor's degree holders, and 82 master's degree holders. As shown, the visualization of this educational landscape is short and highlights in a quick view the intersectionality of gender and marital status within educational achievement.







Figure 2 shows the spread of the data in terms of loan amounts and ages for the dataset. In the set of loan amounts, 1,000,000 is the most common amount, appearing 54 times, followed by 1,200,000, which appears 46 times, and then 1,900,000, which appears 26 times. For the ages, 30 is the most common, appearing 22 times, very closely followed by 25, which appears 20 times. The other ages follow a downward trend, except for several ages that appear at 16 occurrences each and then slowly decrease to 4 occurrences for both 22 and 48. These are the relevant frequencies for the most common loan amounts and ages contained in the dataset, thus able to determine the loan or demographic trend.





Source: Generated using Python

The frequency distributions for the Monthly Income and Interest Rate variables are shown in Figure 3. One can glean the distribution pattern of the data from these. For the Monthly Income data, the data range from \$60,000 to \$241,000, and corresponding frequencies detail how many people fall within an income bracket. Similarly, the Interest Rate data range from 5.00% to 15.00%, where frequencies are the occurrences of each interest rate. These distributions form the basis for exploring the central tendencies, variability, and possible relationships between Monthly Income and Interest Rate to gain a better understanding of the characteristics of this underlying dataset.



#### Figure 3. Distribution of Monthly Income & Interest Rate

Source: Generated using Python

Table 2 summarizes a dataset, which includes information on 400 people who applied for a loan. It further illustrates various key metrics, including age, monthly income, loan amount, and interest rate. The average age of the applicants for loans is 33.45 years, with a standard deviation of 9.33 years; the mean monthly income is \$150,040 with a standard deviation of \$62,100. Loan amounts requested range from \$1,000,000 to \$1,900,000 with an average of \$1,428,000 and a standard deviation of \$284,461. Interest rates offered on these loans vary between 5% and 15%, with an average of 10.25% and a standard deviation of 2.89%. Quartile values add more information about the distribution of these variables among the applicants. Generally, Table 1 contains an overview of those who applied for loans.

	Age	Monthly Income	Loan Amount	Interest Rate
count	400	400	400	400
mean	33,45	150 040,00	1 428 000,00	10,25
std	9,33	62 100,16	284 461,40	2,89
min	18,00	51 000,00	1 000 000,00	5,00
25%	25,00	89 500,00	1 200 000,00	7,75
50%	33,50	153 000,00	1 400 000,00	10,50
75%	41,00	203 250,00	1 700 000,00	12,75
max	50,00	249 000,00	1 900 000,00	15,00

#### **Table 2. Descriptive Statistics**

Source: Author's calculation

The following metrics present an evaluation of a classification model that predicts whether personal banks loan are approved or denied: for the class denied, precision equals 0.76, meaning that from all instances predicted as denied, 76% were classified correctly. Its recall equals 0.86, indicating the model caught 86% of the actually denied cases, while the F1-score, a harmonic mean between the precision and recall for the class denied, is 0.81. Accuracy for approved instances is higher, with precision being 0.88—showing that the model could identify approved cases well, but recall is somewhat weaker at 0.79. The F1-score for approved instances is 0.83. Accuracy stands at 0.82, meaning the model has classified 82% of the instances correctly. The macro average for precision, recall, and F1-score equals 0.82, indicating a balanced performance over the classes; finally, the weighted average, taking into account class imbalance, yields similar values, which indicates uniform performance over the dataset.

	precision	recall	f1-score	support
Denied	0.76	0.86	0.81	44
Approved	0.88	0.79	0.83	56
accuracy			0.82	100
macro avg	0.82	0.82	0.82	100
weighted avg	0.83	0.82	0.82	100

#### **Table 3. Classification Report**

Source: Author's calculation

The confusion matrix shown in the Figure 4 for the test of train shows the performance of a classification model on the training data. With 38 true negatives and 44 true positives, the model seems able to correctly identify both the negative and positive instances. It, however, misclassifies some instances, 12 as false negatives and 6 as false positives. As it appears, while the false negative rate would mean that instances of the positive class are being labeled as negative and the false positive rate means instances of the negative class are labeled as positive, the overall low numbers mean that the model is performing decently enough to distinguish between the two classes. The model has therefore learned the underlying patterns in the training data to some satisfactory degree.

Figure 4. Confusion Matrix



Source: Author's calculation

In the ranking of features provided, monthly income acts as the most influential predictor, with a weight of 0.259105. It is of paramount importance in determining the outcome being analyzed. Age comes next, ranking second, with a slightly lower weight of 0.222496; this shows that it has a significant effect on the outcome. The third in place is loan amount, with a feature importance of 0.203758, and the interest rate is subsequently placed with a feature importance of 0.202456. Both attributes lay a similar strong claim to importance. Farther down, there is a place for marital status with a much lower feature importance of 0.051692, which seems to indicate that its contribution to prediction will be much less marked compared to the foregoing features. The education level and

gender follow after that with a still lower feature importance of 0.034430 and 0.026063, respectively, which implies they have only a very minor influence on the outcome.





Source: Author's calculation

## 5. Conclusion

Conclusively, the research carried out to understand personal bank loan approval shows the complex dynamics at play in lending decisions using a strong method that combines statistical analysis and machine learning techniques.By prioritizing key variables such as age, monthly income, loan amount, and interest rate, it was realized that economic stability and the ability to pay were key in the approval of loans. The findings showed that monthly income was the strongest predictor, followed very closely by age, amount of loan, and interest rate, as can be seen from the weights of 0.259105, 0.222496, 0.203758, and 0.202456, respectively.In addition, the study of demographic factors indicated that social attributes and educational level are intersecting dimensions that provide really valuable insights into educational inequalities and their consequences in terms of access to loans. Also, the performance evaluation for the classification model showed promising precision and recall rates with precision rates of 0.76 for the denied and 0.88 for the approved, and recall rates of 0.86 for the denied and 0.79 for the approved, which proves to be effective in predicting loan approval. Overall, our findings provide a deeper understanding of the factors that shape loan approvals and give action-based insights for decision-makers in financial institutions, policymakers, and borrowers, which can lead to greater financial inclusion and equity with informed decision-making.

### 6. References

1. Alagic, A. et al., 2024. Machine Learning for an Enhanced Credit Risk Analysis: A Comparative Study of Loan Approval Prediction Models Integrating Mental Health Data. *Machine Learning and Knowledge Extraction*, 6(1, https://doi.org/10.3390/make6010004), pp. 53-77.

- Azam, R., Danish, M. & Akbar, S. S., 2012. The Significance of Socioeconomic Factors on Personal Loan Decision (A Study of Consumer Banking in Local Private Banks in Pakistan). *Munich Personal RePEc Archive*, Volume http://dx.doi.org/10.2139/ssrn.2167960.
- 3. Dansana, D. et al., 2023. Analyzing the impact of loan features on bank loanprediction using Random Forest algorithm. *Engineering Reports*, 6(2, doi: 10.1002/eng2.12707).
- 4. Himali, L. P., 2020. Determinants of Personal Loan Default and Performance of the Proportional Hazards Model with that ofaRandom Survival Forests Models. *International Journal of Applied Science and Research*, 3(4), pp. 84-93.
- 5. Meshref, H., 2020. Predicting Loan Approval of Bank Direct Marketing Data Using Ensemble Machine Learning Algorithms. *INTERNATIONAL JOURNAL OF CIRCUITS, SYSTEMS AND SIGNAL PROCESSING*, Volume 14, pp. 914-922.
- Tiwari, K. K. & Bapat, H. B., 2020. DETERMINANTS OF LOAN DELINQUENCY IN PERSONAL LOAN. *International Journal of Management (IJM)*, 11(11, DOI: 10.34218/IJM.11.11.2020.241), pp. 2566-2575.
- Végh, L., Czakóová, K. & Takáč, O., 2023. Comparing Machine Learning Classification Models on a Loan Approval Prediction Dataset. *International Journal of Advanced Natural Sciences and Engineering Researches (IJANSER)*, 7(9), pp. 98-103.
- 8. Wang, H. & Xie, Z., 2023. Analyze the influence of various factors on personal loans. s.l., AEMPS (2023) Vol. 6: 448-459. DOI: 10.54254/2754-1169/6/2022186.