Financial Risk Assessment for Loan Application using Logistic Regression and Factor Analysis

Mr. Kushal Dutia

College Student, B.Sc. Applied Statistics and Analytics, Nilkamal School of Mathematics, Applied Statistics & Analytics, Narsee Monjee Institute of Management Studies (Deemed-to-be) University

Abstract:
Loan officer at any bank would be interested in detecting the factors which can identify people who are likely to default on loans, consequently good and bad credit risks. Moreover, he will also be interested in designing model which can predict chances of default with reasonable accuracy. A crucial step in the loan application process is financial risk assessment, which helps financial institutions minimize credit risk and allocate resources as efficiently as possible. This paper creates a predictive framework for estimating the probability of loan defaults by combining factor analysis and logistic regression. While Factor Analysis lowers dimensionality and finds latent characteristics influencing financial risk, like credit history and debt-to-income ratio. Logistic Regression is a strong classification technique is used to estimate default probabilities by balancing sensitivity and specificity, the suggested model showed excellent predictive accuracy. It offers a data-driven and interpretable way to expedite loan approval decisions, guaranteeing a more fair and effective procedure.

Keywords: Loan Default, Credit Risk, Financial Risk Assessment, Factor Analysis, Logistic Regression, Dimensionality Reduction, Predictive Modelling, Default Probability, Sensitivity and Specificity, Data-Driven Decision Making

1. Introduction:

The overall purpose of this project is to analyse and build a model on loan application data that can be useful for decision-making purposes. Traditional methods rely on credit scoring models that consider a range of financial and demographic variables. However, these models often face challenges due to multicollinearity, high dimensionality, and the lack of interpretability of complex relationships among variables. To address these limitations, advanced techniques of multivariate analysis are applied here: Factor Analysis & Logistic regression. These will help identify patterns, relationships, and predict from the dataset, thus supporting managerial decision making in risk assessment and loan approval processes. This term paper simulates the real-world scenario wherein a financial institution aims to optimize its loan approval system. The analysis would identify the key factors that influence loan decisions, segment customers into meaningful groups, and build predictive models for loan approval within the restrictions of risk management standards.

Objective:

- To identify the key underlying factors that influence loan approval decisions based on borrowers' financial profiles, credit history, and demographic characteristics.

- To predict loan approval likelihood based on financial and demographic factors, identifying significant predictors such as credit score, debt-to-income ratio, homeownership, and loan term to inform loan decision-making.

2. Literature Review:

Statistical and machine learning techniques for evaluating loans have been thoroughly examined in previous studies. Because of its ease of use and interpretability, logistic regression has been a mainstay. Conversely, factor analysis has been used in many fields to extract latent variables from correlated data. In order to improve performance, recent research emphasize the advantages of combining dimensionality reduction approaches with predictive models. By preprocessing loan data using factor analysis before utilizing logistic regression for prediction, this study expands on previous discoveries.

2.1 Credit Assessment:

The process of assessing a person's or an organization's creditworthiness in order to ascertain their capacity to repay debt is known as credit evaluation. To ascertain the degree of risk involved in making a loan to a borrower, it entails examining their income, assets, liabilities, and financial history because it enables them to reduce risk and make well-informed lending decisions, credit evaluation is essential for financial institutions. (Nsengiyera, 2020) Financial firm risk losing money if they lend money to people or organizations that might not be able to repay the loan if their credit is not properly evaluated.

Credit Assessment Types –

There are several methods for evaluating credit, and each has pros and cons of its own. These are a few of the most prevalent kinds:

- **Traditional Credit Assessment:** This kind of evaluation is predicated on a borrower's financial records, payment history, and credit history. The evaluation entails examining the borrower's credit utilization ratio, existing debts, payment history, and credit score. Banks and other financial organizations frequently conduct traditional credit evaluations.
- **Behavioural Credit Assessment:** This method determines a borrower's creditworthiness by analysing information about their habits and behaviour. In order to learn more about the borrower's reliability and character, this method looks at non-financial data including social media activity and internet purchases.
- **Alternative Credit Assessment:** This method of determining a borrower's creditworthiness uses non-traditional data sources, like utility bills, rent payment history, and other information. This strategy is especially helpful for borrowers who do not have a lot of credit history or who have little access to conventional credit information.

2.2 Logistic Regression:

Credit risk Analysis can be carried out through numerous methods. The logistic regression and neural network models both are alike & worthy. Though the prior one is marginally better. Moreover, the genetic algorithm model is somewhat inferior but efficient. (Gouvea, 2007). Logistic regression and multicriteria decision making have been combined to create a proficient strategy to achieve high performance. Logistics regression is used to find probability of default whereas multicriteria decision is used to classify firms for credit scoring based on predefined criteria. (Jerić, 2009). Credit Risk Analysis is modelled using an artificial neural network. In commercial banks, logistic regression modelling is more effective than logistic regression modelling at identifying problematic clients. Higher credit risk results from higher interest rates and longer repayment terms. Customers in the industry sector are more affected. With a longer customer relationship history with the bank and a longer loan repayment period, it decreases. (Karimi, 2014).

2.3 Factor Analysis:

Factor analysis, which includes methods like principal component analysis and common factor analysis, is a statistical technique for identifying interrelations among a large number of variables. It explains these variables in terms of underlying dimensions or factors. The primary goal is to reduce a large dataset into fewer variables (factors) while retaining most of the original information. By providing a quantitative estimation of the variables' structure, factor analysis offers an objective method to create summarized scales. For instance, in a study of customer ratings for a fast-food restaurant, six variables—food taste, food temperature, freshness, waiting time, cleanliness, and employee friendliness could be analysed. (Joseph F. Hair, 2019) The analysis might reveal that food taste, temperature, and freshness cluster into a single factor, termed "food quality." Similarly, waiting time, cleanliness, and employee friendliness might form another factor, referred to as "service quality." This approach helps researchers better understand the underlying relationships among variables.

3. Methodology:

- **Dataset Definition and Source:** The dataset used for this analysis includes detailed information on loan applications from borrowers and various types of variables- financial and credit profile, demographic characteristics, and loan-specific details.
- **Source:** The dataset has been curated and shared for academic purposes to simulate loan approval data in a financial institution.
- **Dataset Objective:** This dataset will be analysed to identify key factors influencing loan approval decisions, predict loan outcomes, and segment borrowers into actionable groups based on financial and credit profiles.
- **Dataset Description:** The dataset consists of 1,099 records, with the following key variables:

Table 4: Key Variables of the Dataset

Sno.	Variable Name	Description
1.	Current Loan Amount	Total loan amount currently held by the borrower (numeric)
2.	Credit Score	Credit score of the borrower (numeric)
3.	Annual Income	Borrower's annual income (numeric)
4.	Monthly Debt	Total monthly debt obligations of the borrower (numeric)
5.	Years of Credit History	Number of years the borrower has maintained a credit history (numeric)
6.	Number of Open Accounts	Total number of open credit accounts (numeric)
7.	Number of Credit Problems	Number of past credit problems (e.g., defaults, delinquencies)
8.	Current Credit Balance	Total current credit balance (numeric)
9.	Maximum Open Credit	Maximum credit limit of open accounts (numeric)
10.	Bankruptcies	Number of times the borrower has declared bankruptcy (numeric)
11.	Tax Liens	Number of tax liens against the borrower (numeric)
12.	Loan_Term	Duration of the loan (binary/categorical)
13.	HomeOwner	Type of home ownership (e.g., Rent, Home Mortgage) (categorical)
14.	Loan Approval	Loan approval status (binary: Approved/Not Approved)
15.	Purpose	Reason for the loan (e.g., Debt Consolidation, Business Loan) (categorical)

3.1 Factor Analysis:

Objective: To identify the key underlying factors that influence loan approval decisions based on borrowers' financial profiles, credit history, and demographic characteristics.

SPSS Output:

a) **KaiserMeyerOlkin (KMO):** Measures the sampling adequacy to ensure that the data is suitable for factor analysis. A value close to 1 indicates that factor analysis is appropriate.

- KMO Value: 0.553

The KMO value above 0.5, indicates that the dataset is adequate for factor analysis.

b) Bartlett's Test of Sphericity: Tests the null hypothesis that the correlation matrix is an identity matrix (no correlation between variables). A significant result (p-value < 0.05) suggests that the data is suitable for factor analysis.

- Chi-square (Approx.): 4964.935
- Degrees of Freedom (df): 136
- Significance (Sig.): 0.000

The test is highly significant ($p < 0.05$), indicating that the correlation matrix is not an identity matrix. This means there are significant relationships among the variables, justifying the use of factor analysis.

Table 4.1.1: KMO and Barlett's Test Result

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.553
Bartlett's Test of Sphericity	Approx. Chi-Square	4964.935
	df	136
	Sig.	.000

c) Principal Component Analysis:

- **Eigenvalues and Factor Retention:** Components with eigenvalues above 1 are typically considered significant and considered for further analysis in Factor Analysis. In this case, seven components have an eigenvalue above 1, so they were taken as the significant factors. This states, complexity in your dataset can be reduced to about seven core factors without information loss being too consequential
- **Scree Test:** A visual inspection of the eigenvalues in a scree plot. The point where the curve flattens helps to determine the number of factors to retain.
- **Variance Explained by Each Factor:** The initial eigenvalues column shows that the first component explains 16.507% of the variance, while the cumulative variance explained by the seven factors is 64.764%. This means that these seven components capture around 65% of the total variance in the dataset.

Table 4.1.2: Total Variance Explained

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.806	16.507	16.507	2.806	16.507	16.507	1.992	11.717	11.717
2	1.934	11.378	27.885	1.934	11.378	27.885	1.907	11.217	22.934
3	1.472	8.657	36.542	1.472	8.657	36.542	1.724	10.139	33.073
4	1.348	7.927	44.469	1.348	7.927	44.469	1.503	8.841	41.914
5	1.229	7.227	51.696	1.229	7.227	51.696	1.438	8.459	50.373
6	1.148	6.754	58.450	1.148	6.754	58.450	1.305	7.675	58.047
7	1.073	6.314	64.764	1.073	6.314	64.764	1.142	6.717	64.764
8	.998	5.870	70.634						
9	.906	5.330	75.964						
10	.846	4.976	80.940						
11	.760	4.473	85.413						
12	.733	4.311	89.724						
13	.665	3.912	93.636						
14	.518	3.045	96.680						
15	.326	1.920	98.601						
16	.143	.843	99.444						
17	.095	.556	100.000						

Extraction Method: Principal Component Analysis.

Table 4.1.3: Scree Plot

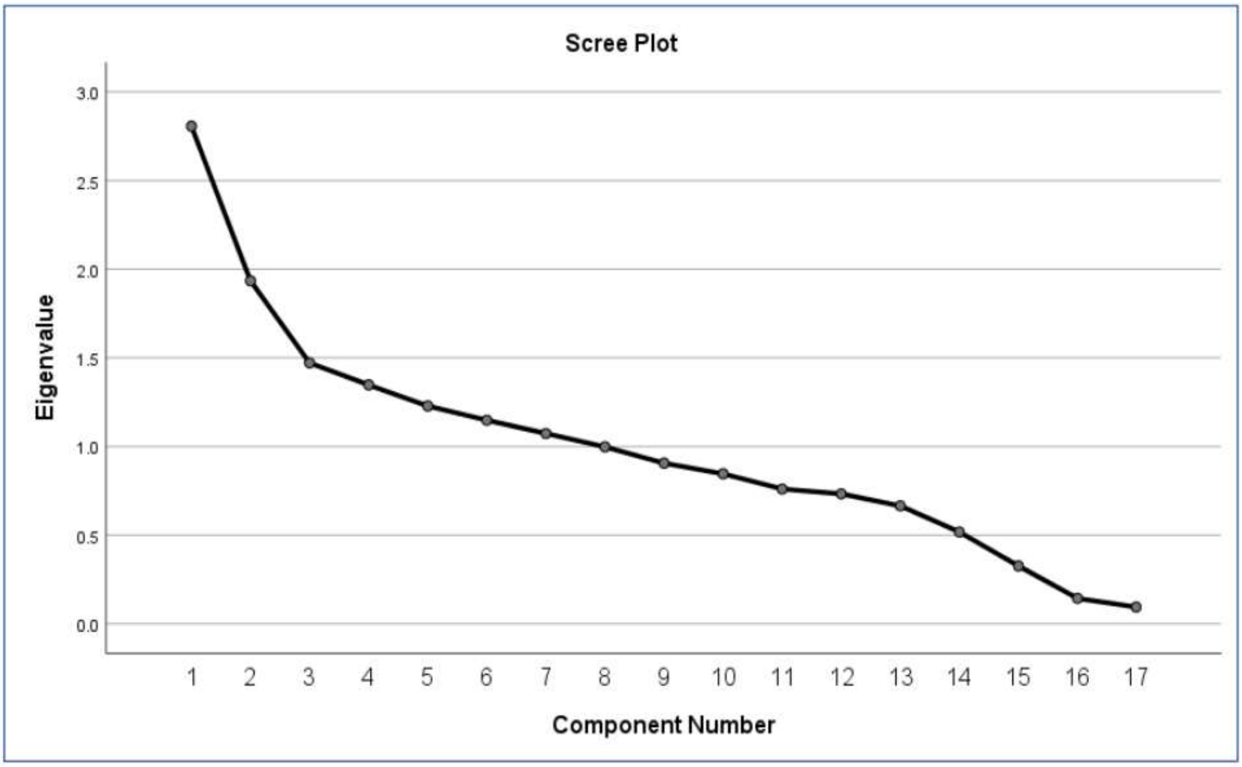

Table 4.1.4: Rotated Component Matrix

Rotated Component Matrix[a]

	Component						
	Credit Balance and Utilization	Loan Characteristics	Income and Debt	Credit History and Stability	Loan Terms	Credit Utilization and Purpose	Negative Credit Events
Current Loan Amount		.972					
Annual Income			.593				
Monthly Debt			.765				
Years of Credit History				.641			
Loan_To_Income		.969					
Months since last delinquent							
Number of Open Accounts			.733				
Current Credit Balance	.918						
Maximum Open Credit	.899						
Bankruptcies							.727
Purpose_Target						.524	
Tax Liens							.671
Creditupdate					-.821		
Loan_Term					.804		
Years_Job				.729			
Howner				.622			
Credit_Utilization						.820	

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.[a]

a. Rotation converged in 8 iterations.

d) Communalities Analysis:

Variables with High Communalities: These variables are strongly captured by the factors, indicating they share a lot of variances with the latent constructs.

Variables with Moderate Communalities: These are reasonably well explained by the factors but could be influenced by other unmeasured variables.

Variables with Low Communalities: These are poorly explained by the factors, suggesting that either these variables are not highly related to the latent constructs or there are additional underlying factors not captured in this analysis.

Table 4.1.5: Communalities Table

Communalities

	Initial	Extraction
Current Loan Amount	1.000	.948
Annual Income	1.000	.547
Monthly Debt	1.000	.760
Years of Credit History	1.000	.475
Loan_To_Income	1.000	.950
Months since last delinquent	1.000	.261
Number of Open Accounts	1.000	.634

Current Credit Balance	1.000	.877
Maximum Open Credit	1.000	.871
Bankruptcies	1.000	.648
Purpose_Target	1.000	.334
Tax Liens	1.000	.565
Creditupdate	1.000	.741
Loan_Term	1.000	.699
Years_Job	1.000	.566
Howner	1.000	.436
Credit_Utilization	1.000	.699

Extraction Method: Principal Component Analysis.

Interpretation:

a) Factor 1: Credit Balance and Utilization

This factor reflects the borrower's credit capacity and current usage of available credit limits. High values in this factor might indicate individuals with high credit balances and greater credit limits, which in turn may suggest a strong credit profile.

b) Factor 2: Loan Characteristics

This factor describes the size and burden of the loan relative to the borrower's income.

It captures how well borrowers are able to handle their loan obligations relative to what they earn.

c) Factor 3: Income and Debt

This factor evaluates the creditworthiness of the borrower. It emphasizes income and current obligations. The number of accounts open represents the borrower's activity in the credit system.

d) Factor 4: Credit History and Stability

This would reflect the ability to commit to savings and dependability. A longer credit history, stable employment, and homeownership suggest a borrower is likely more financially dependable.

e) Factor 5: Loan Terms

This reflects the terms of the loan and recent updates in the credit profile of the borrower. The negative loading on "Credit update" suggests the relationship between recent updates to credit files and length of terms of loans might be inversely related.

f) Factor 6: Credit Utilization and Purpose

This factor focuses on how credit is being used for specific goals and the proportion of credit being utilized. It ties the borrower's credit behaviour to their purpose for borrowing.

g) Factor 7: Negative Credit Events

This factor measures the presence of adverse financial events, including bankruptcies and tax liens, which directly impact creditworthiness and the borrower's ability to secure loans.

4.2 Logistic Regression:

Objective: To predict loan approval likelihood based on financial and demographic factors, identifying significant predictors such as credit score, debt-to-income ratio, homeownership, and loan term to inform loan decision-making.

SPSS Output:

a) Case Processing Summary:

Table 4.2.1: Case Processing Summary

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	1099	100.0
	Missing Cases	0	.0
	Total	1099	100.0
Unselected Cases		0	.0
Total		1099	100.0

a. If weight is in effect, see classification table for the total number of cases.

b) Categorical Variables Codings:

Table 4.2.2: Categorical Variables Codings

Categorical Variables Codings

		Frequency	Parameter coding (1)	(2)	(3)
Home Ownership	HaveMort	1	1.000	.000	.000
	Home Mor	499	.000	1.000	.000
	Own Home	85	.000	.000	1.000
	Rent	514	.000	.000	.000
Term	Long Ter	303	1.000		
	Short Te	796	.000		

c) Baseline Comparison: The baseline accuracy of 74.0% serves as a benchmark for evaluating the logistic regression model.

Table 4.2.3: Classification Table

Block 0: Beginning Block

Classification Table[a,b]

Observed			Predicted Loan_approval_status 0	1	Percentage Correct
Step 0	Loan_approval_status	0	0	286	.0
		1	0	813	100.0
	Overall Percentage				74.0

a. Constant is included in the model.

b. The cut value is .500

d) Omnibus Test of Significance:

Table 4.2.4: Omnibus Test of Significance

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	897.759	16	.000
	Block	897.759	16	.000
	Model	897.759	16	.000

The p-value = 0.000 (less than 0.05), which indicates that the model with predictors significantly improves the prediction compared to the null model.

e) Hosmer and Lemeshow Test:

Table 4.2.5: Hosmer and Lemeshow Test

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	12.770	8	.120

Interpretation:

1. **Chi Square Value (12.770):** This measures the difference between observed and predicted probabilities of the dependent variable (loan approval). A smaller value indicates a better fit.
2. **Significance (p = 0.120):** Since p=0.120p = 0.120p=0.120 is greater than 0.05, there is no statistically significant evidence to suggest a lack of fit. This means the model's predictions are consistent with the observed outcomes.
3. **Classification Table:**

Table 4.2.6: Classification Table

Classification Table[a]

Observed			Predicted: Loan_approval_status 0	Predicted: Loan_approval_status 1	Percentage Correct
Step 1	Loan_approval_status	0	241	45	84.3
		1	20	793	97.5
	Overall Percentage				94.1

a. The cut value is .500

The Classification Table provides the model's prediction accuracy for the dependent variable (in this case, Loan Approval Status) based on the logistic regression model with predictors. Here's how to interpret the results:

- **Loan Approval Status = 0** (Not Approved):
- Observed = 286, of which:
- ▪ Correctly predicted = 241.
- ▪ Incorrectly predicted = 45 (predicted as approved when not approved).
- Accuracy for Not Approved cases = 84.3%.
- **Loan Approval Status = 1** (Approved):
- Observed = 813, of which:
- Correctly predicted = 793.
- Incorrectly predicted = 20 (predicted as not approved when actually approved).
- Accuracy for Approved cases = 97.5%.
- **Overall Percentage:**
- 94.1% of all cases (both approved and not approved) were correctly classified by the model.

Table 4.2.7: Variables in the Equation

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Creditscore_new	.119	.010	148.616	1	.000	1.127
	Current Loan Amount	.000	.000	.180	1	.671	1.000
	Credit_Utilization	.010	.007	1.949	1	.010	1.011
	Purpose_Personal_Improvement	.922	.760	1.474	1	.225	2.515
	Purpose_Asset_Purchase	1.218	1.349	.815	1	.367	3.379
	Purpose_Financial_Restructuring	-.087	.439	.039	1	.843	.917
	Purpose_Business	.053	1.036	.003	1	.959	1.054
	Loan_Term	.348	.336	1.074	1	.300	1.416
	DTI	-7.616	1.708	19.879	1	.000	.000
	Bankruptcies	-25.659	2891.860	.000	1	.993	.000
	Tax Liens	-24.064	3162.480	.000	1	.994	.000
	Monthly Debt	.000	.000	3.568	1	.012	1.000
	Home Ownership			3.640	3	.303	
	Have Mortgage	16.067	40192.969	.000	1	1.000	9504785.235
	Home Mortgage	-.608	.319	3.631	1	.043	.544
	Own Home	-.318	.581	.299	1	.585	.728
	Constant	-80.967	6.810	141.362	1	.000	.000

a. Variable(s) entered on step 1: Creditscore_new, Current Loan Amount, Credit_Utilization, Purpose_Personal_Improvement, Purpose_Asset_Purchase, Purpose_Financial_Restructuring, Purpose_Business, Loan_Term, DTI, Bankruptcies, Tax Liens, Monthly Debt, Home Ownership.

Significant Predictors ($p < 0.05$)

- **Credit Score:**
  - $B = 0.119$, $p < 0.001$, $Exp(B) = 1.127$
  - A one-unit increase in the credit score increases the odds of loan approval by 12.7%.
  - Insight: High credit scores are positively associated with loan approvals.
- **Debt-to-Income Ratio (DTI):**
  - $B = -7.616$, $p < 0.001$, $Exp(B) = 0.000$
  - A higher DTI decreases the odds of loan approval to almost zero.
  - Insight: Applicants with manageable DTI are significantly more likely to get loans.
- **Monthly Debt:**
  - $B = 0.000$, $p = 0.012$, $Exp(B) = 1.000$
  - Monthly debt, though appearing small, is significant in predicting approvals.
  - Insight: Applicants with consistent monthly debt management may positively influence approval decisions.
- **Credit Utilization:**
  - $B = 0.010$, $p = 0.010$, $Exp(B) = 1.011$
  - A slight increase in credit utilization increases the odds of loan approval by 1.1%.
  - Insight: Moderate credit utilization indicates responsible financial behaviour.
- **Home Mortgage:**
  - $B = -0.608$, $p = 0.043$, $Exp(B) = 0.544$
  - Applicants with mortgages have 45.6% lower odds of approval compared to renters.
  - Insight: Mortgage holders may be seen as riskier due to existing liabilities.

Non-Significant Predictors ($p \geq 0.05$)

- **Loan Purpose (e.g., Personal Improvement, Asset Purchase, Financial Restructuring):**
  - These purposes show no significant impact on loan approval decisions individually.
  - Action: Reassess purpose-specific lending policies or use alternate modelling techniques.
- **Loan Term:**
  - $B = 0.348$, $p = 0.300$, $Exp(B) = 1.416$
  - Longer-term loans slightly increase the odds of approval but are not statistically significant.
- **Bankruptcies and Tax Liens:**
  - Both predictors were not significant, with $p > 0.99$.
  - Action: While these variables may not directly predict approvals, they could influence
  - other loan terms (e.g., interest rates).
- **Current Loan Amount:**
  - $p = 0.671$, indicating no significant impact on approvals.
- **Home Ownership Categories (Other than Mortgage):**
  - Categories such as "Own Home" did not significantly affect loan approvals.

Final Logistic Regression Equation:

$Log(P/1-P) = -80.967+(0.119\times\text{Credit Score}) -(7.616\times\text{DTI}) +(0.000\times\text{Monthly Debt}) +(0.010\times\text{Credit Utilization}) -(0.608\times\text{Home Mortgage})$.

5. Result:

A) Factor Analysis:

- This factor is critical for assessing creditworthiness and risk of default, as it shows whether the borrower is nearing their credit limit or managing balances responsibly.
- A balanced mix of income and debt obligations may be healthy, indicating a stable financial situation. A high debt-to-income ratio might indicate potential difficulties in discharging the debt and consequently, in servicing the debt balance.
- These factors strongly predict trustworthiness and consistency of credit behaviour. Borrowers scoring high here could be considered lower risk, since they have confirmed stability over time.
- A high score here may mean that there are significant loan obligations relative to income, potentially signifying financial strain. This factor is vital for assessing affordability of loans and the debt-to-income ratio, which is an essential metric for lenders.
- A borrower who has several new credits may have secured some new financial obligations or changes that have altered his credit profile. This factor brings in the implication of the current credit events on the long-term obligation.

B) Logistic Regression:

- **Creditworthiness Assessment -** Credit Score is the most significant predictor.

Action: Develop exclusive products or expedited processes for applicants with high credit scores.

- **Debt-to Income Ratio -** Debt-to-Income Ratio shows a strong negative relationship.

Action: Use DTI as a primary eligibility filter in automated approval systems.

- **Loan Purpose Analysis -** Asset Purchase loans show a trend toward approval, although not statistically significant.

Action: Consider creating specialized loan offerings or discounts for asset purchases.

- **Home Ownership Insights -** Mortgage Holders are less likely to receive approvals.

Action: Introduce tailored products like mortgage-backed loans or restructured payment plans for this segment.

- Renters are viewed more favourably.

Action: Explore why renters are approved more frequently and refine criteria for homeowners.

- **Operational Enhancements -** Automate loan approvals based on significant predictors (e.g., high credit score, low DTI, moderate credit utilization).

Action: Deploy a predictive model to flag high-probability approvals and reduce processing time.

- **Risk Management -** High DTI and excessive credit utilization are key risk factors.

Action: Offer credit counselling and risk-based pricing for high-risk customers. Monitor customers with significant monthly debts to identify those suitable for cross-selling other financial products.

- **Policy Refinement -** Reassess the role of non-significant variables (e.g., bankruptcies, tax liens) in risk scoring.

Action: Explore alternative datasets or variable transformations to improve predictive power.

6. Conclusion:

The aim of the project was to combine factor analysis and logistic regression to create a reliable framework for evaluating financial risk in loan applications. The probability of loan defaults was successfully determined using logistic regression, which is known for its resilience in binary classification situations. Factor analysis made it possible to reduce dimensionality and identify important latent factors like debt-to-income ratio, income stability, and credit history that affect financial risk.

The findings showed that combining these techniques can greatly improve loan default prediction interpretability and predictive accuracy. The model provided financial organizations with a useful tool for data-driven decision making by striking a balance between specificity and sensitivity. The framework ensures better financial resource allocation, streamlines loan approval procedures, and reduces credit risk.

7. Recommendation for future research:

- **Integration of advanced machine learning models:** Adding sophisticated algorithms like ensemble techniques or Gradient Boosting Machines (like XGBoost and LightGBM) could improve predicted accuracy and adaptability to complicated datasets.
- **Dynamic factor models:** Using dynamic factor analysis can help models perform better in changing financial markets by taking time varying elements into consideration.
- **Real-time risk assessment:** By integrating streaming data into real-time scoring systems, it may be possible to obtain immediate insights while processing loan applications.
- **External data incorporation:** The model could be enhanced and non-traditional creditworthiness signals could be captured by incorporating alternative data sources such social media activity, transaction patterns, and utility payment histories.
- **Explainability and AI trust:** By providing clear insights into the variables affecting risk projections, explainable AI (XAI) solutions can increase stakeholders' trust.
- **Fairness and Regulatory Compliance:** Future research should concentrate on making sure the model conforms with ethical norms and financial rules, especially when it comes to biases in loan acceptance decisions.
- **Geographic and Demographic Expansion:** By evaluating the model in other geographical locations and demographic groupings, its generalizability can be confirmed and market-specific modifications can be found.
- **Integration with Blockchain Technology:** Investigating blockchain's potential for safe data exchange and improved loan processing transparency can increase public confidence in financial institutions.

References

1. Gouvea, M. A. (2007). Credit Risk Analysis Applying Logistic Regression. Neural Networks and Genetic Algorithms Models. POMS 18th Annual Conference.
2. Jerić, S. &. (2009). Logistic Regression and Multicriteria Decision Making in Credit Scoring,. Proceedings of the 10th International Symposium on Operational Research SOR '09.
3. Karimi, A. (2014). Credit Risk Modeling for Commercial Banks. International Journal of Academic Research in Accounting Finance and Management Sciences, .
4. Joseph F. Hair, W. C. (2019). Multivariate Data Analysis . Cengage Learning.
5. Begiev, Z. (2017). Bank Loan Status Dataset. (kaggle.com/datasets/zaurbegiev, Ed.) Kaggle.
6. Nsengiyera, D. (2020). THE ASSESSMENT OF LOAN MANAGEMENT ON FINANCIAL PERFORMANCE OF BANKING INSTITUTIONS IN RWANDA CASE STUDY: BANK OF KIGALI HEADQUARTER. Daniel Nsengiyera University of Science and Technology of China .