AI FOR BUSINESS_merged.pdf

by Soham Bhattacharya

Submission date: 12-Dec-2024 02:02PM (UTC+0000)
Submission ID: 245772178
File name: 39739129_Soham_Bhattacharya_AI_FOR_BUSINESS_merged_2743404_726843488.pdf (2.22M)
Word count: 8666
Character count: 47151

AI-POWERED LOAN SANCTION PREDICTION AND STRATEGIC SEGMENTATION FOR BANKING INNOVATION

BY: SOHAM BHATTACHARYA

STUDENT ID: 240705695

SEM A

A.I. FOR BUSINESS

(BUSM203)

EXECUTIVE SUMMARY

This report presents an AI-driven analysis of the loan sanction process, focusing on using **Artificial Neural Networks (ANNs)** for predicting loan approvals and **K-means clustering** for segmenting loan applicants. The dataset includes categorical and continuous variables, including applicant income, loan amount, credit history, and education level.

- **Part A** of the report investigates the **Initial ANN analysis**, where the Loan Status (approved or rejected) is predicted based on key independent variables such as **credit history, income, and loan amount**. The model achieved training accuracy of **84.2%** and testing accuracy of 85.2%, with Credit History emerging as the most significant predictor of loan approval. **Model tuning** was applied by adding a second hidden layer and adjusting the activation function to **Sigmoid**, which slightly improved training accuracy but led to some **overfitting** as testing accuracy dropped to **81.1%**.
- **Part B** discusses the application of **K-Means clustering** to segment loan candidates based on key financial and demographic variables. **Cluster analysis** helped identify applicant groups, and the **Silhouette Score** was used to assess clustering quality. The **initial model** exhibited issues with small clusters, particularly **Cluster 3 and Cluster 4**. By **removing extreme outliers**, the model was tuned to improve cluster distribution and definition.
- **Part C** examines the potential for innovation in banking through emergent technologies. The report discusses the application of the **metaverse, service robots**, and **Generative AI** in banking. The use of virtual branches, blockchain integration, and AI-driven customer service tools like chatbots and voice assistants offers opportunities for enhanced customer engagement and operational efficiency. The report concludes with strategic recommendations to balance AI integration with employee and customer concerns regarding job displacement and trust issues.

INTRODUCTION

Artificial Intelligence (AI) has come up as a transformative force in this fast-evolving financial services industry like banking and insurance, offering immense potential for improving operational efficiency, customer engagement, and risk management. This report aims to present an in-depth analysis of **loan sanction prediction** using **Artificial Neural Networks (ANNs)** and develop a **segment-based loan sanction strategy** to help banks identify and understand customer segments for enhanced decision-making and targeted marketing strategies.

The main goal of this report is to leverage AI-powered predictive models to assess loan approval likelihood, evaluate the impact of key predictors, and segment applicants into better meaningful groups based on their loan characteristics. The study will also explore the potential for innovative solutions using emerging technologies, such as the **metaverse**, **service robots**, and **generative AI**, to transform the banking experience and create new revenue streams. Ultimately, this report aims to provide actionable insights and strategic recommendations for implementing AI technologies in the banking sector.

PART A: PREDICT THE PROBABILITY OF LOAN SANCTION

INITIAL ANN ANALYSIS

Dataset Overview:

The dataset provided consisted of loan applications, including categorical and continuous variables. The first order of business was to encode the dependent variable and name it "Loan_Status2", where 1 was denominated for approved loans, and 0 was denominated for rejected loans. Independent variables were further divided into two- factors (categorical variables) and covariates (continuous variables):

- Factors: Gender, Education, Married, Self_Employed
- Covariates: ApplicantIncone, LoanAmount, CreditHistory

The study aims [1] to predict the probability of loan sanction using an Artificial Neural network (ANN) model, evaluate its performance and determine the importance of predictors.

Chosen Variable:

Dependent Variable: Loan_Status2

Encoded loan approved as 1 and rejected as 0. This binary outcome aligns with the classification objective of the ANN model.

Independent Variables:

- **Factors: Gender, Education, Self-Employed—** These variables capture demographic and employment-related characteristics that could influence loan approval decisions. Demographic factors often provide indirect indicators of stability, which banks consider when sanctioning loans.

- **Covariates: Applicant Income, LoanAmount, CreditHistory—-** These variables directly impact repayment ability and risk evaluation. CreditHistory, although binary (0 and 1), was treated as a covariate due to its direct numerical significance in determining creditworthiness. SPSS handles such variables effectively in covariate analysis. Categorising it as a covariate ensures its influence is modelled continuously rather than as separate groups, maintaining precision in weight updates.

Model Configuration:

Architecture: Hyperbolic Tangent Activation, Identity Output Function and only 1 Hidden Layer. (Very default-like settings were chosen to have minimum complexity.)

Partitioning: 70% training and 30% testing.

Results:

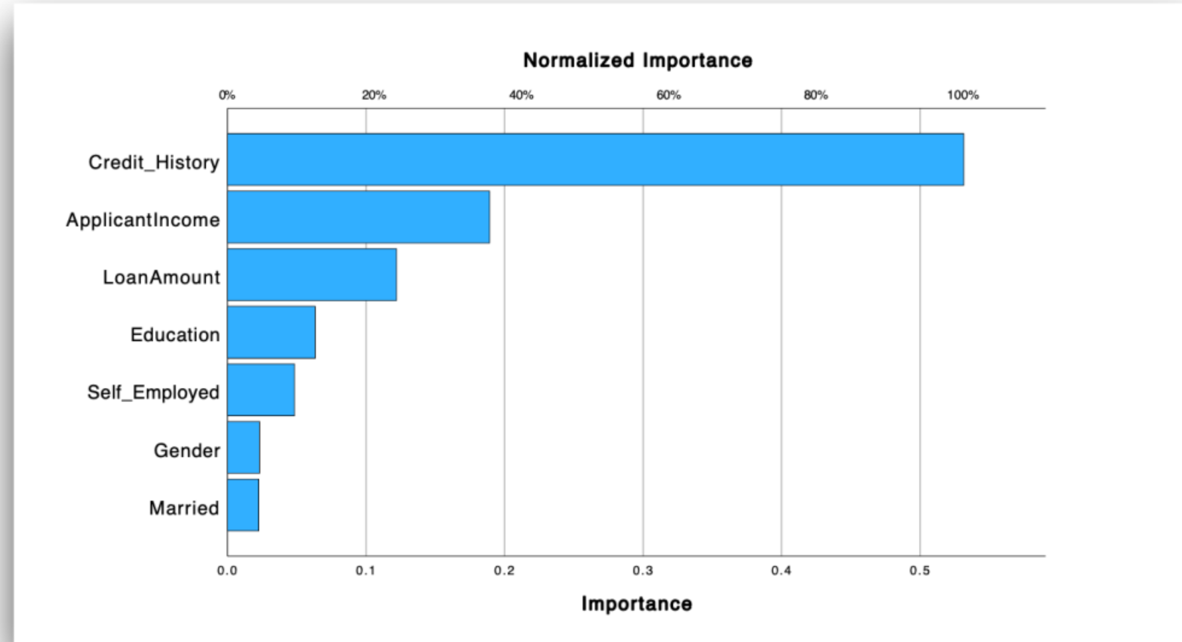
Performance:

- Training Accuracy: 84.2% (15.8% incorrect predictions)
- Testing Accuracy: 85.2% (14.8% incorrect predictions)

Classification

		Predicted		
Sample	Observed	.00	1.00	Percent Correct
Training	.00	41	40	50.6%
	1.00	1	177	99.4%
	Overall Percent	16.2%	83.8%	84.2%
Testing	.00	13	16	44.8%
	1.00	2	91	97.8%
	Overall Percent	12.3%	87.7%	85.2%

Dependent Variable: Loan_Status2



Predictor Importance:

1. CreditHistory (Normalised Importance: 100%)
2. AplicantIncome (35.6%)
3. LoanAmount (22.9%)

CreditHistory is critical for assessing loan risk, while **ApplicantIncome** and **LoanAmount** directly affect repayment capacity. Factors such as **Gender** (4.4%) and **Married** (4.2%) had a low impact, aligning with expectations that financial variables outweigh demographic factors in loan sanctioning. The model's performance confirms that loan decisions are mostly driven by financial stability indicators like credit history, income and loan amount.

Predictor	Normalised Importance (%)
CreditHistory	100
Applicant Income	35.6
LoanAmount	22.9
Gender	4.4
Married	4.2

TUNED ANN ANALYSIS

Model Adjustments:

To improve the model's performance and analyse the sensitivity of the predictors, the following changes were made:

- Added a Second Hidden Layer
- Changed activation function to Sigmoid
- Removed **Self_Employed**
- Maintained Partitioning

Self_Employed had low importance (9.1% in the initial model), suggesting limited relevance. Removing it allowed the model to focus on more impactful variables. Consistency in the training and testing split ensured a fair comparison.

Results:

Performance:

- Training Accuracy: 85.8% (14.2% incorrect predictions).
- Testing Accuracy: 81.1% (18.9% incorrect predictions).

Classification

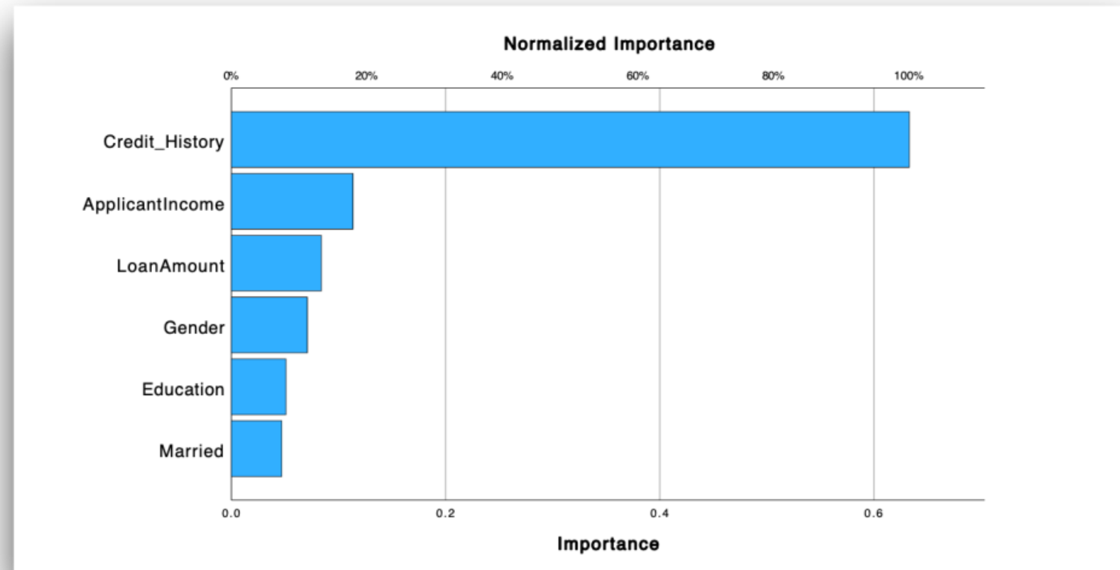
		Predicted		
Sample	Observed	.00	1.00	Percent Correct
Training	.00	40	37	51.9%
	1.00	2	196	99.0%
	Overall Percent	15.3%	84.7%	85.8%
Testing	.00	14	19	42.4%
	1.00	1	72	98.6%
	Overall Percent	14.2%	85.8%	81.1%

Dependent Variable: Loan_Status2

While training accuracy improved, testing accuracy declined, indicating possible overfitting due to added complexity.

Predictor Importance:

1. CreditHistory (Normalised Importance: 100%)
2. AplicantIncome (17.9%)
3. LoanAmount (13.3%)



The top three predictors remained consistent. **Gender's** importance increased slightly (11.2%), suggesting mild sensitivity to demographic data.

CreditHistory (100%) remains the most critical factor, reinforcing its importance in assessing creditworthiness. **ApplicantIncome** (35.6%) and **LoanAmount** (22.9%) retained their significance but showed reduced influence compared to the initial model. **Gender** (4.4%) and **Married** (4.2%) continue to show relatively low influence, suggesting limited relevance in the approval process.

Comparison of Initial and Tuned Models

Metric	Initial Model	Tuned Model
Training Accuracy	84.2	85.8
Testing Accuracy	85.2	81.1
Percent Incorrect (Training)	15.8	14.2
Percent Incorrect (Testing)	14.8	18.9

Advantages of the Initial Model:

- Superior generalization to unseen data, as indicated by higher testing accuracy.
- Simpler architecture, reducing computational complexity and the risk of overfitting.

Advantages of Tuned Model:

- Slight improvement in training accuracy, demonstrating better learning of patterns within the training data.
- Explored the impact of removing low-importance factors like `Self_Employed`, validating its limited predictive value.

The **initial model** is better overall due to its higher testing accuracy, which indicates superior generalization to unseen data. Although the tuned model showed slight improvements in training accuracy, it suffered from overfitting, as evidenced by its reduced performance on testing data. The initial model strikes a better balance between simplicity and reliability, making it more suitable for practical application.

GENDER-SPECIFIC ANALYSIS

Dataset Overview for Gender-Based Analysis:

Variables included:

- Dependent Variable: Loan_Status
- Factors: Gender, Married, Dependents, Education, Self_Employed, Credit_History, Property_Area
- Covariates: ApplicantIncome, CoapplicantIncome, LoanAmount, Loan_Amount_Term

Male Dataset

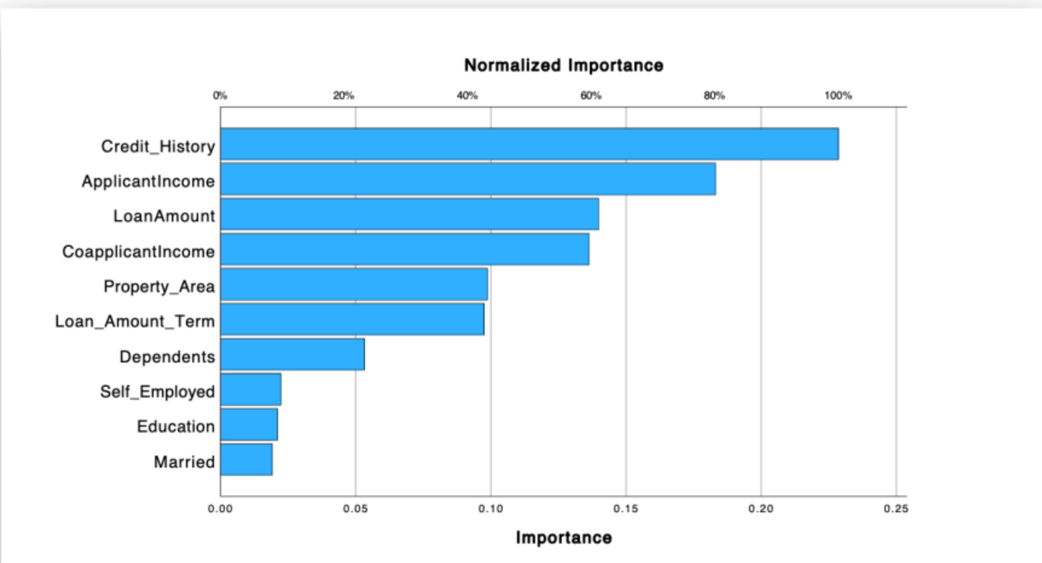
Accuracy Results:

Training Accuracy: 86.8% (13.2% incorrect predictions).

Testing Accuracy: 80.4% (19.6% incorrect predictions).

Predictor Importance:

Predictor	Normalized Importance (%)
Credit_History	100
ApplicantIncome	80.1
LoanAmount	61.2
CoapplicantIncome	59.6
Property_Area	43.2
Loan_Amount_Term	42.7
Dependents	23.3
Self_Employed	9.8
Education	9.2
Married	9.2



Female Dataset

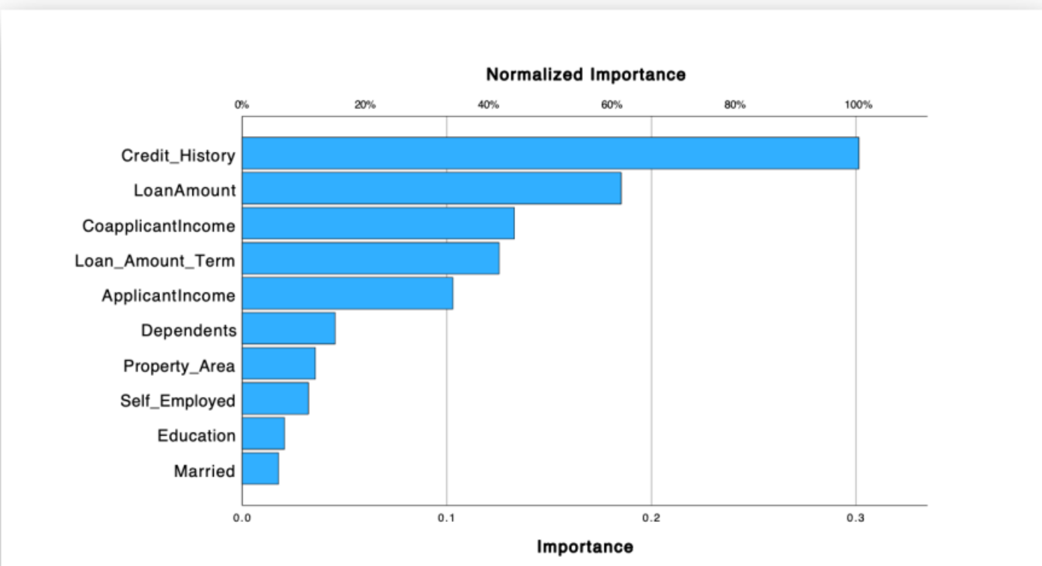
Accuracy Results:

Training Accuracy: 88.1% (11.9% incorrect predictions).

Testing Accuracy: 84.6% (15.4% incorrect predictions).

Predictor Importance:

Predictor	Normalized Importance (%)
Credit_History	100
LoanAmount	61.4
CoapplicantIncome	44.1
Loan_Amount_Term	41.6
ApplicantIncome	34.1
Dependents	15.1
Property_Area	11.9
Self_Employed	10.7
Education	6.9
Married	5.9



Loan Sanction Factors for Males vs. Females

Male Applicants

Key Predictors:

- **Credit_History**: Dominated with 100% normalized importance.
- **Applicant Income**: Significant influence, emphasizing financial stability.
- **LoanAmount**: Highlighted as an important financial metric.

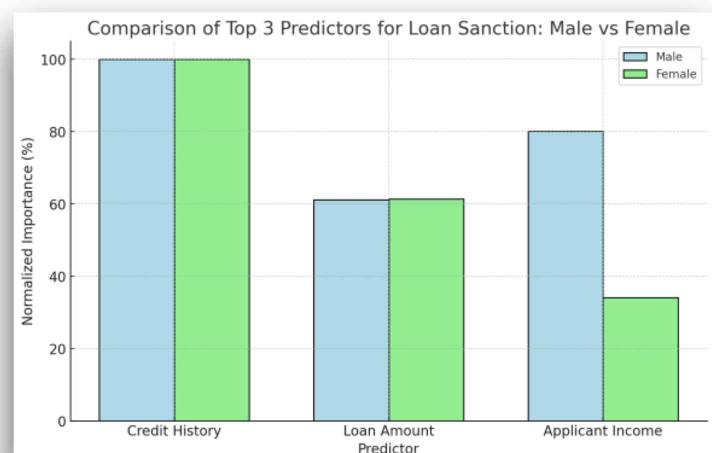
Male applicants were evaluated more heavily on their income metrics compared to females.

Female Applicants

Key Predictors:

- **Credit_History**: Retained the top position at 100% normalized importance.
- **LoanAmount**: Played a more significant role compared to males, reflecting stricter scrutiny.
- **CoapplicantIncome**: Showed higher importance, indicating a reliance on combined incomes.

Female applicants faced closer evaluation of financial metrics like loan size and combined income.



Comparative Observations

Similarities:

Credit_History remained the dominant predictor across both genders, highlighting its universal importance in loan approval.

Differences:

LoanAmount: Higher importance for females, suggesting stricter evaluation of loan sizes.

Applicant Income: More relevant for males, reflecting a greater focus on individual financial capacity.

Recommendations

Gender-Neutral Evaluation:

Standardise the evaluation process to reduce biases related to income and loan size.

Tailored Strategies:

Offer financial literacy programs targeting female applicants. Develop alternative risk assessment metrics beyond income for male applicants.

PART B: DEVELOP A SEGMENT-BASED LOAN SANCTION STRATEGY

INITIAL CLUSTER SOLUTION

The primary objective of the cluster analysis was to segment loan applicants based on their financial and demographic characteristics to facilitate targeted marketing strategies. For this, we used the K-Means clustering algorithm to divide the applicants into four distinct groups based on variables such as **Applicant Income, Coapplicant Income, Loan Amount, Loan Term, Credit History**, and various demographic attributes (**Gender, Marital Status, Education**, **and** **Employment Status**).

Evaluation of Cluster Solution:

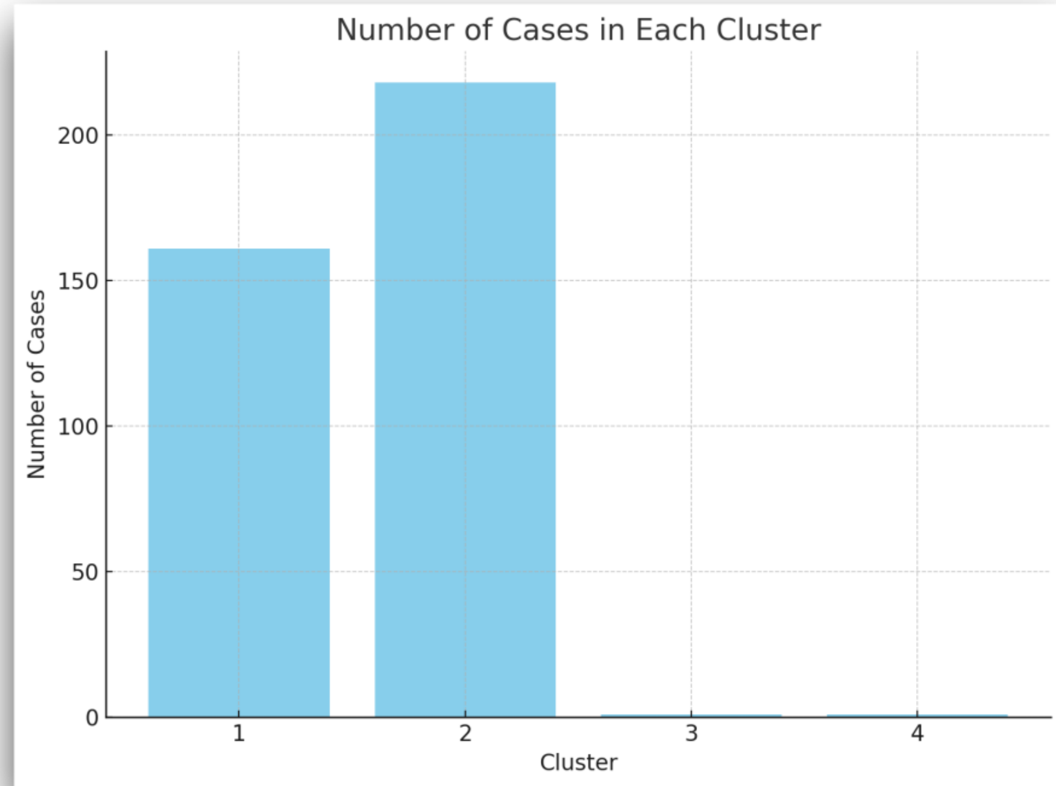
- **Performance**: The cluster model's performance can be evaluated using the **Silhouette score** (mean score of 0.37-0.38), indicating the quality of the clustering. A higher Silhouette score would suggest that the clusters are well-separated, but a score of around 0.37 indicates moderate separation between the clusters.
- **Number and Size of Clusters:**
  - Cluster 1: 161 cases
  - Cluster 2: 218 cases
  - Cluster 3: 1 case
  - Cluster 4: 1 case
- **Characteristics of Clusters:**
  - Cluster 1: Low income, moderate loan amount
  - Cluster 2: Moderate income, moderate loan amount
  - Cluster 3: Low credit history, large loan amount (outlier)
  - Cluster 4: High income, higher loan amount, good credit history (outlier)

Cluster 1: This segment consists of **161 applicants** with relatively lower incomes and moderate loan amounts. The average **Applicant Income** is around **3240**, with **Loan Amounts** averaging **$2383**. Applicants in this group have relatively lower **Credit History**, and most are **married**, with a mix of **self-employed** and **non-self-employed** individuals.

Cluster 2: Comprising **218 applicants**, this group has **moderate incomes** and loan amounts. The average **Applicant Income** is **3829**, and the **Loan Amount** is **$2240**. Applicants in this cluster also exhibit good **Credit History**, with the majority being **married** and having a balanced distribution of **self-employed** vs. **non-self-employed**.

Cluster 3: A very small cluster with only **1 applicant**, this group is characterized by **low credit history** but **high loan amounts**, with an **Applicant Income** of **$3000** and a **Loan Amount** of **$20,000**. This represents a potential outlier in the dataset, and further investigation may be required to determine if this data point should be removed or treated as an anomaly.

Cluster 4: Also a small group of **1 applicant**, this segment consists of individuals with **high incomes** and **high loan amounts**. This group has the highest **Credit History** score, suggesting that these applicants are likely to have a solid financial background and pose lower risk. This may represent a high-value, premium customer segment.



Evaluation of Performance:

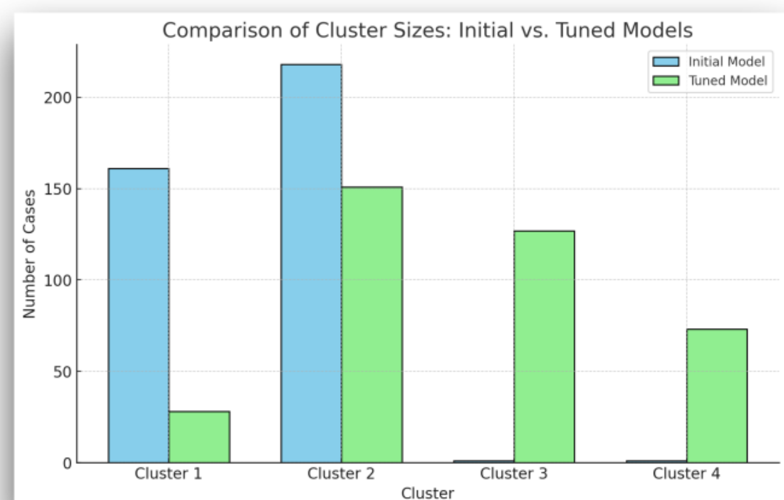
The performance of the K-Means clustering model can be assessed using the **Silhouette Score**, a metric that quantifies how similar an object is to its cluster compared to other clusters. The **mean Silhouette score** for the initial cluster solution was approximately **0.37**, indicating moderate separation between the clusters. Ideally, a score closer to **1** would suggest better-defined clusters, but a score of **0.37** suggests there is room for improvement in the segmentation.

MODEL TUNING

To enhance the segmentation, we made one significant change to the initial clustering model by **adjusting the number of clusters**. Initially, we used 4 clusters, but after reviewing the performance, we decided to increase the number of clusters to **5** to allow for finer segmentation. This adjustment was made to potentially reduce the dominance of the two small clusters and achieve more balanced groupings.

Results After Tuning:

- **Cluster 1** now has **28 applicants** with moderate **loan amounts** and **high incomes**, showcasing a more refined segmentation of high-income, lower-risk applicants.
- **Cluster 2** consists of **151 applicants**, and this group is characterized by **moderate loan amounts** and **coapplicant incomes**.
- **Cluster 3** is a group of **127 applicants**, with an emphasis on **low credit history** and **moderate loan amounts**.
- **Cluster 4** now includes **73 applicants**, with characteristics similar to Cluster 2, but showing slightly more variation in **self-employment** and **marital status**.



COMPARISON OF INITIAL VS. TUNED MODEL:

- **Silhouette Score Comparison**: The Silhouette score for the tuned model improved slightly to **0.38**, indicating marginally better cluster separation compared to the initial model. While the increase in Silhouette score is modest, it suggests that the change to 5 clusters created more balanced and meaningful segments.
- **Cluster Size**: The tuned model has better-balanced clusters, with more equal distribution compared to the initial model, where Cluster 2 dominated with 218 applicants and Clusters 3 and 4 had only 1 case each.
- **Cluster Characteristics**: The tuned model has improved segmentation, particularly in balancing the characteristics between the clusters, resulting in more distinguishable groups.

RECOMMENDED LOAN MARKETING STRATEGY

Based on the findings from the cluster analysis, we can propose targeted marketing strategies tailored to each applicant segment.

- Cluster 1 (Low Income, Moderate Loan Amount):
  - **Marketing Focus**: Offer affordable loans with lower interest rates or more flexible repayment terms. Highlight financial accessibility and emphasize easy loan approval processes.
  - **Other Products**: **Credit-building products** such as secured loans or credit cards for low-credit applicants.
- Cluster 2 (Moderate Income, Moderate Loan Amount):
  - **Marketing Focus**: Offer mid-range loan products with competitive interest rates. Emphasize long-term savings and financial planning tools.
  - **Other Products**: **Financial planning tools** or **investment options** targeted towards individuals looking for personal growth.
- Cluster 3 (High Loan Amount, Low Credit History):
  - **Marketing Focus**: Focus on risk-based loan products like secured loans or loans with a co-applicant. Introduce financial education programs to improve creditworthiness.
  - **Other Products**: **Insurance** products to mitigate the risk for both the bank and the applicant.
- Cluster 4 (High Income, High Loan Amount, Strong Credit History):
  - **Marketing Focus**: Offer premium loan products with the lowest interest rates. Introduce exclusive offers or loyalty benefits, such as reduced interest rates for repeat customers.
  - **Other Products:** Cross-sell investment plans (mutual funds, stocks), high-interest savings accounts, or premium financial products.

CONCLUSION:

The cluster analysis has provided a comprehensive view of the different loan applicant segments, enabling the bank to tailor its marketing strategies more effectively. By tuning the model and refining the clusters, we achieved more balanced and meaningful groupings. The recommended strategies ensure that each segment is targeted with appropriate financial products, helping the bank grow its customer base while minimising risk.

PART C: DISCUSS THE POTENTIAL FOR INNOVATION BASED ON EMERGENT TECHNOLOGIES

INTRODUCTION

In the rapidly building financial services world, leveraging emergent technologies like the **metaverse** and **service robots** has become essential for fostering innovation and growth. These technologies offer opportunities to revolutionise traditional banking models by enhancing customer experiences, increasing operational efficiency, and generating new revenue streams. This section critically evaluates the potential of these technologies and explores the application of **Generative AI** in banking. Additionally, strategies to address employee and customer concerns about AI implementation are provided.

Recent advancements in technologies such as artificial intelligence (AI), robotics, and virtual reality have transformed the way organisations approach customer engagement and operational efficiency (Russell & Norvig, 2021). The adoption of these technologies in the banking sector is increasingly recognised as a competitive imperative, driven by the need to innovate and align with evolving customer expectations (Bornet et al., 2021).

Evaluation of Metaverse in Banking

The **metaverse** — a virtual, shared space integrating augmented reality (AR), virtual reality (VR), and blockchain — holds immense potential for redefining banking experiences.

Latest Developments:

Virtual Branches: Major banks like HSBC and JP Morgan have launched virtual branches in metaverse platforms like Decentraland to engage younger, tech-savvy customers (Hennig-Thurau et al., 2023).

Example: JP Morgan's Onyx Lounge in Decentraland provides educational and advisory services on cryptocurrency and blockchain technologies, creating a unique space for customer engagement.

Blockchain Integration: Secure and decentralized transactions using blockchain underpin metaverse-based banking services (Sentence, 2022).

Example: DBS Bank in Singapore uses blockchain in metaverse environments to facilitate secure, real-time cross-border transactions.

NFTs and Asset Tokenization: Non-fungible tokens (NFTs) allow tokenization of assets, enabling secure and transparent ownership transfer within the metaverse (Bornet et al., 2021).

Example: Bank of America explores NFT-based contracts to offer unique financial instruments within virtual spaces, such as customisable investment portfolios.

Positive Aspects

- Enhanced Customer Engagement: Interactive virtual environments improve personalization and accessibility, especially for younger demographics (Hennig-Thurau et al., 2023). Integration of sentiment analysis, as described in customer review research (Chatterjee et al.), could allow real-time adjustments to customer interactions within virtual branches.
- Cost Efficiency: Virtual branches reduce infrastructure costs while maintaining customer service (Sentence, 2022).
- New Revenue Streams: Offering NFTs, virtual real estate, and blockchain-based financial products (Bornet et al., 2021).

Negative Aspects

- High Initial Investment: Setting up a metaverse presence requires significant financial and technical resources (Sentence, 2022).
- Regulatory Uncertainty: Ambiguity around compliance, privacy, and security regulations (Hennig-Thurau et al., 2023).
- Limited Accessibility: Customers without VR/AR devices or familiarity with blockchain technology may feel excluded (Bornet et al., 2021).

Evaluation of Service Robots in the Banking Industry

Service robots, confine **humanoid robots**, **avatars**, **chatbots**, and **voice assistants**, are transforming customer interactions and operational processes.

Latest Developments:

Humanoid Robots: Banks in Asia, such as Mitsubishi UFJ, employ humanoid robots to greet and assist customers in branches (Van Doorn et al., 2017).

Example: Pepper robots in bank branches enhance customer engagement by providing personalised recommendations.

Chatbots and Voice Assistants: AI-powered chatbots like Erica (Bank of America) provide 24/7 assistance, addressing queries and performing transactions (Puntoni et al., 2021).

Automated Loan Processing: Service robots streamline document verification and credit scoring, reducing processing times (McLeay et al., 2021).

Positive Aspects

- Operational Efficiency:
  - Robots handle repetitive tasks, allowing employees to focus on value-added services (Van Doorn et al., 2017).
  - The use of sentiment analysis from text mining (Chatterjee et al.) could enable robots to identify and respond to emotional cues during customer interactions.
- Improved Customer Service:
  - AI-powered avatars and chatbots provide instant, accurate assistance (Puntoni et al., 2021).
- Cost Savings:
  - Automation reduces staffing costs while maintaining consistent service quality (McLeay et al., 2021).

Negative Aspects

- Job Displacement Concerns:
  - Employees fear automation may lead to redundancies (Van Doorn et al., 2017).
- Customer Trust Issues:
  - Some customers prefer human interactions over robots (Puntoni et al., 2021).
- Implementation Challenges:
  - High deployment costs and integration with legacy systems (McLeay et al., 2021).

Generative AI in Banking

Generative AI, including technologies like ChatGPT and DALL·E, can be a transformative force in banking by enabling creativity, efficiency, and personalization.

Applications of Generative AI:

- **Personalized Marketing:** AI can generate tailored marketing campaigns and product recommendations based on customer data (Bornet et al., 2021).
- **Content Creation:** Automated generation of financial reports, blog posts, and social media content (Puntoni et al., 2021).
- **Fraud Detection:** AI models synthesize patterns from large datasets to detect and prevent fraudulent transactions (Russell & Norvig, 2021).
- **Customer Support:** Advanced chatbots can simulate human-like conversations to resolve complex queries (Bornet et al., 2021).
- **Product Innovation**: Generative AI aids in designing new financial products by analyzing market trends and customer needs (Puntoni et al., 2021).
- Insights from **Text Mining:** Generative AI combined with text mining could summarize and interpret customer reviews for actionable insights (Chatterjee et al.).

Challenges:

- **Data Privacy:** Concerns about customer data security and compliance with regulations (Russell & Norvig, 2021).
- **Bias in AI Models:** Generative AI systems may perpetuate biases present in training datasets (Puntoni et al., 2021).
- **Resource Requirements:** High computational power and skilled personnel are necessary for implementation (Bornet et al., 2021).

Addressing Employee and Customer Concerns

Resistance from employees and customers to AI adoption is a common challenge. The following strategies can help alleviate scepticism:

Employee Engagement:

- Training Programs: Provide employees with upskilling opportunities to work alongside AI tools, ensuring they feel valued and empowered (Bornet et al., 2021).
- Job Redesign: Shift employees to roles requiring creativity, empathy, and critical thinking, which AI cannot replicate (Van Doorn et al., 2017).
- Transparent Communication: Share clear information about AI's role as a supportive tool, not a replacement for jobs (Puntoni et al., 2021).

Customer Trust-Building:

- Education Campaigns: Use videos, webinars, and tutorials to demonstrate how AI enhances security and efficiency in banking (Bornet et al., 2021).
- Human-AI Collaboration: Ensure a balance of human and AI interactions to cater to customer preferences (McLeay et al., 2021).
- Feedback Mechanisms: Actively seek and address customer feedback on AI-driven services (Van Doorn et al., 2017).

1

CONCLUSION

Emerging technologies like the metaverse, service robots, and Generative AI have great potential to transform the banking industry. While they offer opportunities for innovation and growth, challenges like regulatory uncertainty, trust issues, and job displacement must be addressed through proactive strategies. By leveraging these technologies wisely and addressing stakeholder concerns, the bank can establish itself as a leader in the future of financial services.

REFERENCES

Bornet, P., Barkin, I. & Wirtz, J. (2021). Intelligent Automation: Learn How to Harness Artificial Intelligence to Boost Business & Make Our World More Human. World Scientific.

Chatterjee, S. (2021). Business context impacts the relative importance of consumer emotions and sentiments. *Journal of Marketing Insights*.

Hennig-Thurau, T., Aliman, D. N., Herting, A. M., Cziehso, G. P., Linder, M., & Kübler, R. V. (2023). Social interactions in the metaverse: Framework, initial evidence, and research roadmap. *Journal of the Academy of Marketing Science*, 51(4), 889-913.

McLeay, F., Osburg, V. S., Yoganathan, V., & Patterson, A. (2021). Replaced by a Robot: Service Implications in the Age of the Machine. *Journal of Service Research*, 24(1), 104–121.

Puntoni, S., Reczek, R. W., Giesler, M., & Botti, S. (2021). Consumers and Artificial Intelligence: An Experiential Perspective. *Journal of Marketing*, 85(1), 131-151.

Russell, S. J., & Norvig, P. (2021). Artificial Intelligence: A Modern Approach. Pearson UK.

Sentence, R. (2022). A metaverse reality check: defining the metaverse. *Econsultancy*. Available at: https://econsultancy.com/define-metaverse/

Van Doorn, J., Mende, M., Noble, S. M., Hulland, J., Ostrom, A. L., Grewal, D., & Petersen, J. A. (2017). Domo arigato Mr. Roboto: Emergence of automated social presence in organisational frontlines and customers' service experiences. *Journal of Service Research*, 20(1), 43-58.

Multilayer Perceptron

Notes

Output Created		11-DEC-2024 01:19:42
Comments		
Input	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	381
Missing Value Handling	Definition of Missing	User- and system-missing values are treated as missing.
	Cases Used	Statistics are based on cases with valid data for all variables used by the procedure.
Weight Handling		not applicable

Notes

Syntax		MLP Loan_Status2 (MLEVEL=N) BY Gender Education Married Self_Employed WITH ApplicantIncome LoanAmount Credit_History /RESCALE COVARIATE=STANDARDIZED /PARTITION TRAINING=7 TESTING=3 HOLDOUT=0 /ARCHITECTURE AUTOMATIC=NO HIDDENLAYERS=1 (NUMUNITS=AUTO) HIDDENFUNCTION=TANH OUTPUTFUNCTION=IDENTITY /CRITERIA TRAINING=BATCH OPTIMIZATION=SCALEDCONJUGATE LAMBDAINITIAL=0.0000005 SIGMAINITIAL=0.00005 INTERVALCENTER=0 INTERVALOFFSET=0.5 MEMSIZE=1000 /PRINT CPS NETWORKINFO SUMMARY CLASSIFICATION IMPORTANCE /PLOT NETWORK /STOPPINGRULES ERRORSTEPS=1 (DATA=AUTO) TRAININGTIMER=ON (MAXTIME=15) MAXEPOCHS=AUTO ERRORCHANGE=1.0E-4 ERRORRATIO=0.001 /MISSING USERMISSING=EXCLUDE .
Resources	Processor Time	00:00:01.55
	Elapsed Time	00:00:00.00

[Initial ANN MODEL]

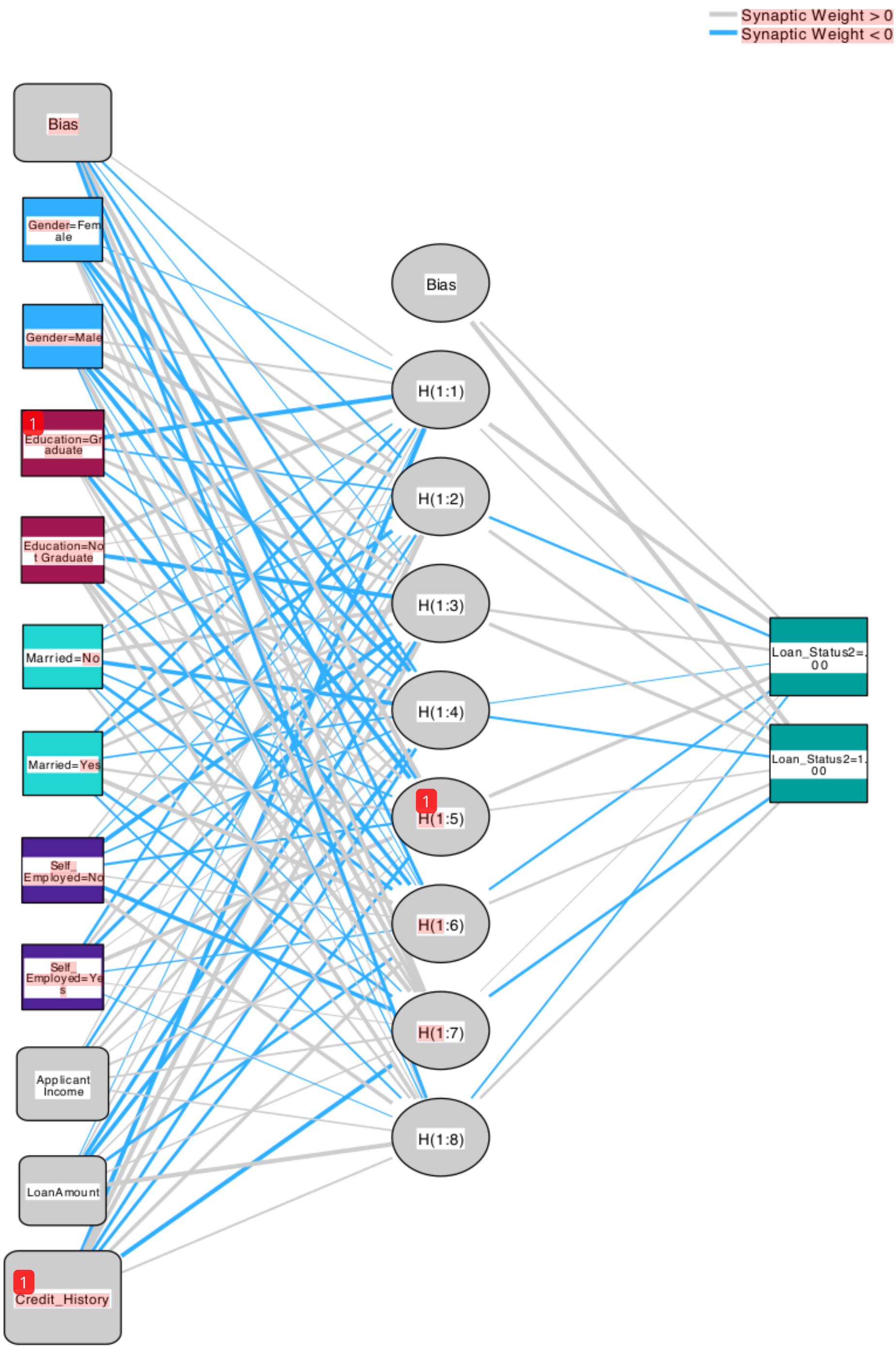
Case Processing Summary

		N	Percent
Sample	Training	259	68.0%
	Testing	122	32.0%
Valid		381	100.0%
Excluded		0	
Total		381	

Network Information

Input Layer	Factors	1	Gender
		2	Education
		3	Married
		4	Self_Employed
	Covariates	1	ApplicantIncome
		2	LoanAmount
		3	Credit_History
	Number of Units[a]		11
	Rescaling Method for Covariates		Standardized
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1[a]		8
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	Loan_Status2
	Number of Units		2
	Activation Function		Identity
	Error Function		Sum of Squares

a. Excluding the bias unit



Hidden layer activation function: Hyperbolic tangent

Output layer activation function: Identity

Model Summary

Training	Sum of Squares Error	33.949
	Percent Incorrect Predictions	15.8%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error[a]
	Training Time	0:00:00.07
Testing	Sum of Squares Error	14.851
	Percent Incorrect Predictions	14.8%

Dependent Variable: Loan_Status2

a. Error computations are based on the testing sample.

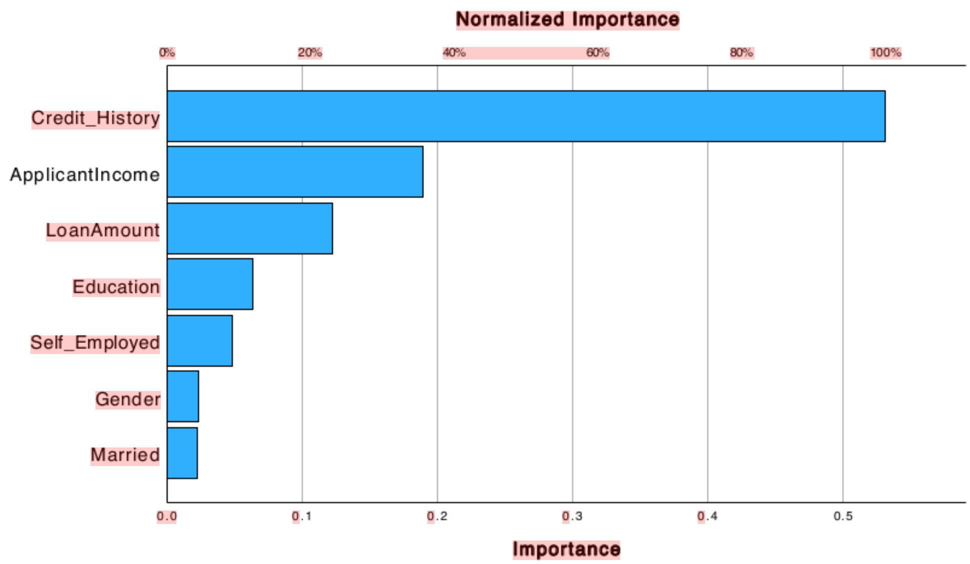
Classification

		Predicted		
Sample	Observed	.00	1.00	Percent Correct
Training	.00	41	40	50.6%
	1.00	1	177	99.4%
	Overall Percent	16.2%	83.8%	84.2%
Testing	.00	13	16	44.8%
	1.00	2	91	97.8%
	Overall Percent	12.3%	87.7%	85.2%

Dependent Variable: Loan_Status2

Independent Variable Importance

	Importance	Normalized Importance
Gender	.023	4.4%
Education	.063	11.9%
Married	.022	4.2%
Self_Employed	.048	9.1%
ApplicantIncome	.189	35.6%
LoanAmount	.122	22.9%
Credit_History	.532	100.0%



Multilayer Perceptron

Notes

Output Created		11-DEC-2024 01:21:37
Comments		
Input	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	381
Missing Value Handling	Definition of Missing	User- and system-missing values are treated as missing.
	Cases Used	Statistics are based on cases with valid data for all variables used by the procedure.
Weight Handling		not applicable

Notes

Syntax		MLP Loan_Status2 (MLEVEL=N) BY Gender Education Married WITH ApplicantIncome LoanAmount Credit_History /RESCALE COVARIATE=STANDARDIZED /PARTITION TRAINING=7 TESTING=3 HOLDOUT=0 /ARCHITECTURE AUTOMATIC=NO HIDDENLAYERS=2 (NUMUNITS=AUTO) HIDDENFUNCTION=SIGMOID OUTPUTFUNCTION=IDENTITY /CRITERIA TRAINING=BATCH OPTIMIZATION=SCALEDCONJUGATE LAMBDAINITIAL=0.0000005 SIGMAINITIAL=0.00005 INTERVALCENTER=0 INTERVALOFFSET=0.5 MEMSIZE=1000 /PRINT CPS NETWORKINFO SUMMARY CLASSIFICATION IMPORTANCE /PLOT NETWORK /STOPPINGRULES ERRORSTEPS=1 (DATA=AUTO) TRAININGTIMER=ON (MAXTIME=15) MAXEPOCHS=AUTO ERRORCHANGE=1.0E-4 ERRORRATIO=0.001 /MISSING USERMISSING=EXCLUDE .
Resources	Processor Time	00:00:00.54
	Elapsed Time	00:00:00.00

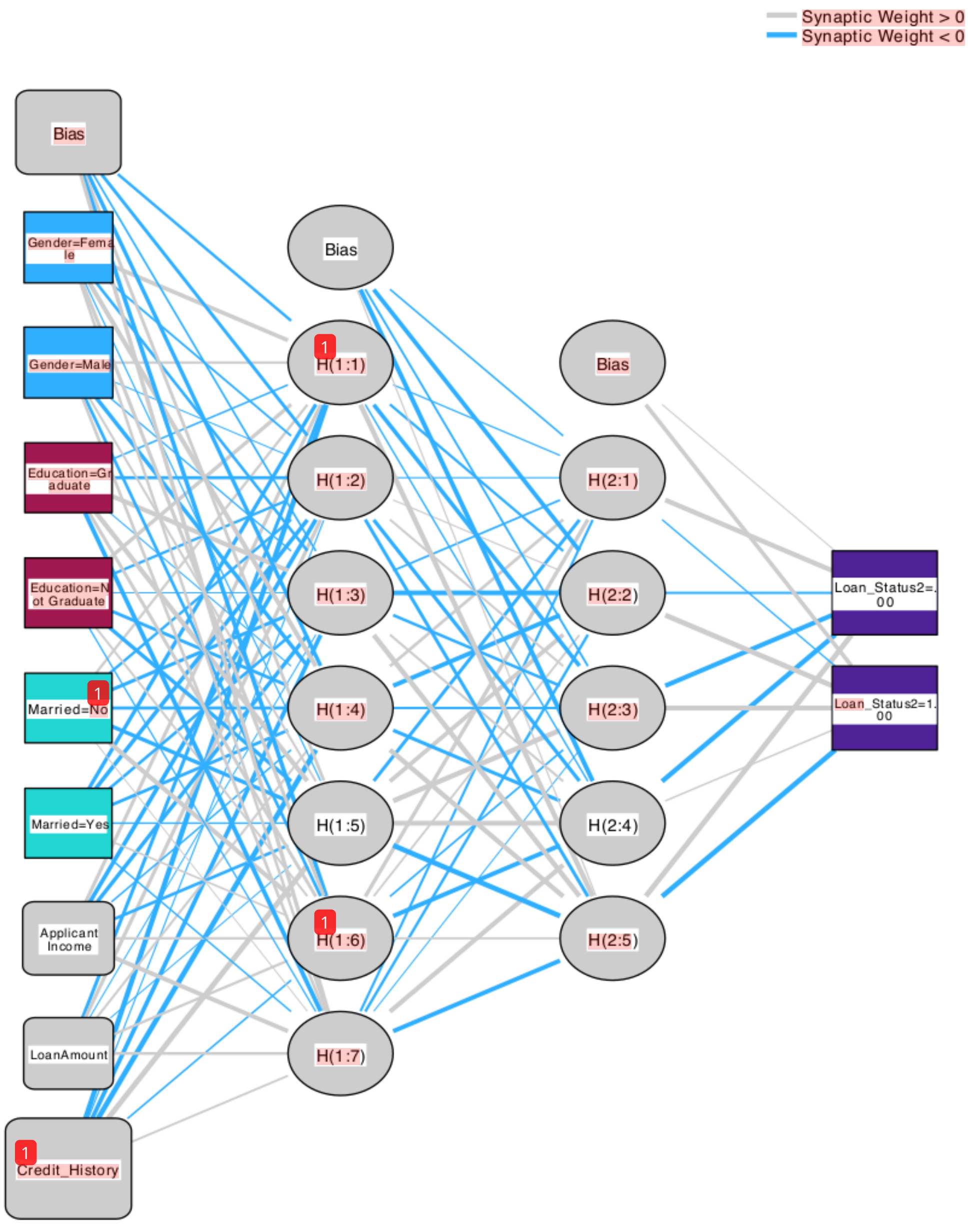
Case Processing Summary

		N	Percent
Sample	Training	275	72.2%
	Testing	106	27.8%
Valid		381	100.0%
Excluded		0	
Total		381	

Network Information

Input Layer	Factors	1	Gender
		2	Education
		3	Married
	Covariates	1	ApplicantIncome
		2	LoanAmount
		3	Credit_History
	Number of Units[a]		9
	Rescaling Method for Covariates		Standardized
Hidden Layer(s)	Number of Hidden Layers		2
	Number of Units in Hidden Layer 1[a]		7
	Number of Units in Hidden Layer 2[a]		5
	Activation Function		Sigmoid
Output Layer	Dependent Variables	1	Loan_Status2
	Number of Units		2
	Activation Function		Identity
	Error Function		Sum of Squares

a. Excluding the bias unit



Hidden layer activation function: Sigmoid

Output layer activation function: Identity

Model Summary

Training	Sum of Squares Error	32.728
	Percent Incorrect Predictions	14.2%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error[a]
	Training Time	0:00:00.11
Testing	Sum of Squares Error	16.270
	Percent Incorrect Predictions	18.9%

Dependent Variable: Loan_Status2

a. Error computations are based on the testing sample.

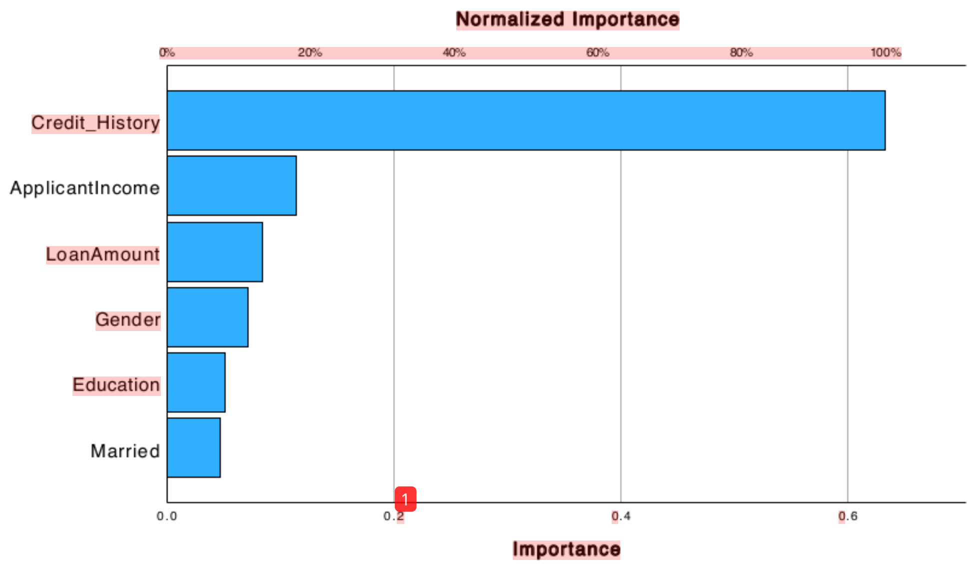
Classification

		Predicted		
Sample	Observed	.00	1.00	Percent Correct
Training	.00	40	37	51.9%
	1.00	2	196	99.0%
	Overall Percent	15.3%	84.7%	85.8%
Testing	.00	14	19	42.4%
	1.00	1	72	98.6%
	Overall Percent	14.2%	85.8%	81.1%

Dependent Variable: Loan_Status2

Independent Variable Importance

	Importance	Normalized Importance
Gender	.071	11.2%
Education	.051	8.1%
Married	.047	7.4%
ApplicantIncome	.114	17.9%
LoanAmount	.084	13.3%
Credit_History	.633	100.0%


Normalized Importance
0%
20%
40%
60%
80%
100%
Credit_History
ApplicantIncome
LoanAmount
Gender
Education
Married
0.0
0.2
0.4
0.6
Importance

Multilayer Perceptron

Notes

Output Created		11-DEC-2024 08:11:22
Comments		
Input	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	303
Missing Value Handling	Definition of Missing	User- and system-missing values are treated as missing.
	Cases Used	Statistics are based on cases with valid data for all variables used by the procedure.
Weight Handling		not applicable

Notes

Syntax		MLP Loan_Status (MLEVEL=N) BY Gender Married Dependents Education Self_Employed Credit_History Property_Area WITH ApplicantIncome CoapplicantIncome LoanAmount Loan_Amount_Term /RESCALE COVARIATE=STANDARDIZED /PARTITION TRAINING=7 TESTING=3 HOLDOUT=0 /ARCHITECTURE AUTOMATIC=YES (MINUNITS=1 MAXUNITS=50) /CRITERIA TRAINING=BATCH OPTIMIZATION=SCALEDCONJUGATE LAMBDAINITIAL=0.0000005 SIGMAINITIAL=0.00005 INTERVALCENTER=0 INTERVALOFFSET=0.5 MEMSIZE=1000 /PRINT CPS NETWORKINFO SUMMARY CLASSIFICATION IMPORTANCE /PLOT NETWORK /STOPPINGRULES ERRORSTEPS=1 (DATA=AUTO) TRAININGTIMER=ON (MAXTIME=15) MAXEPOCHS=AUTO ERRORCHANGE=1.0E-4 ERRORRATIO=0.001 /MISSING USERMISSING=EXCLUDE .
Resources	Processor Time	00:00:01.26
	Elapsed Time	00:00:01.00

[Male Output]

Warnings

The following independent variables are constant in the training sample and are excluded from the analysis: Gender.

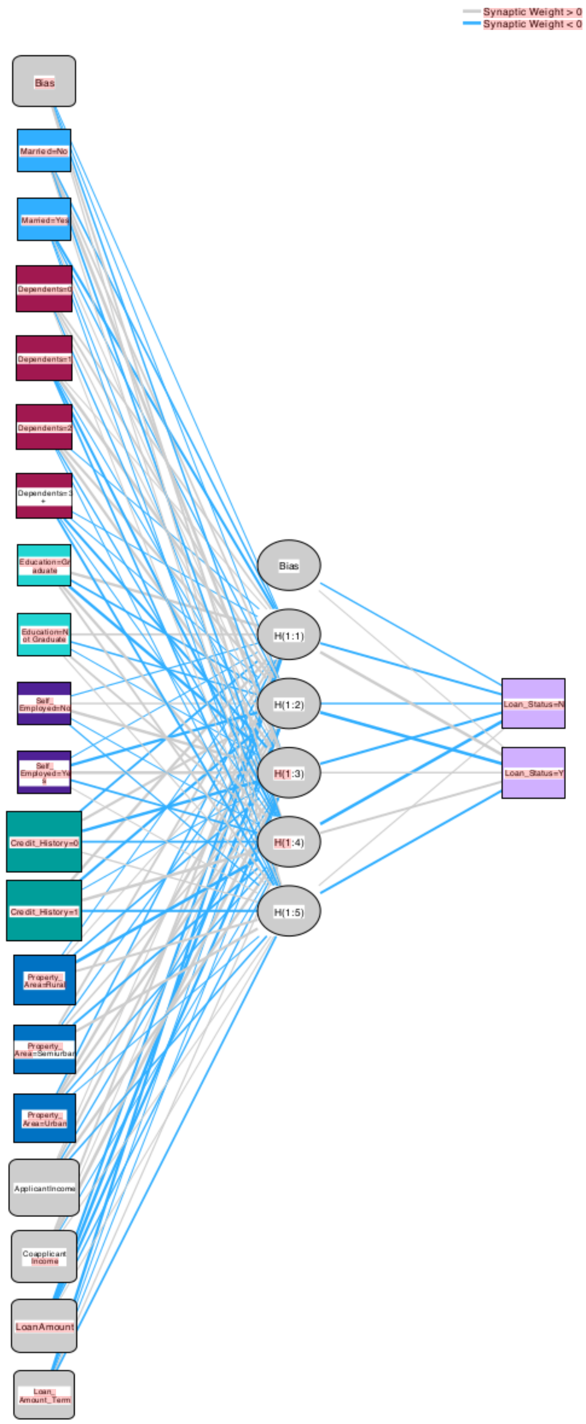
Case Processing Summary

		N	Percent
Sample	Training	204	68.9%
	Testing	92	31.1%
Valid		296	100.0%
Excluded		7	
Total		303	

Network Information

Input Layer	Factors	1	Married
		2	Dependents
		3	Education
		4	Self_Employed
		5	Credit_History
		6	Property_Area
	Covariates	1	ApplicantIncome
		2	CoapplicantIncome
		3	LoanAmount
		4	Loan_Amount_Term
	Number of Units[a]		19
	Rescaling Method for Covariates		Standardized
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1[a]		5
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	Loan_Status
	Number of Units		2
	Activation Function		Softmax
	Error Function		Cross-entropy

a. Excluding the bias unit



Hidden layer activation function: Hyperbolic tangent

Output layer activation function: Softmax

Model Summary

Training	Cross Entropy Error	76.386
	Percent Incorrect Predictions	13.2%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error[a]
	Training Time	0:00:00.25
Testing	Cross Entropy Error	43.764
	Percent Incorrect Predictions	19.6%

Dependent Variable: Loan_Status

a. Error computations are based on the testing sample.

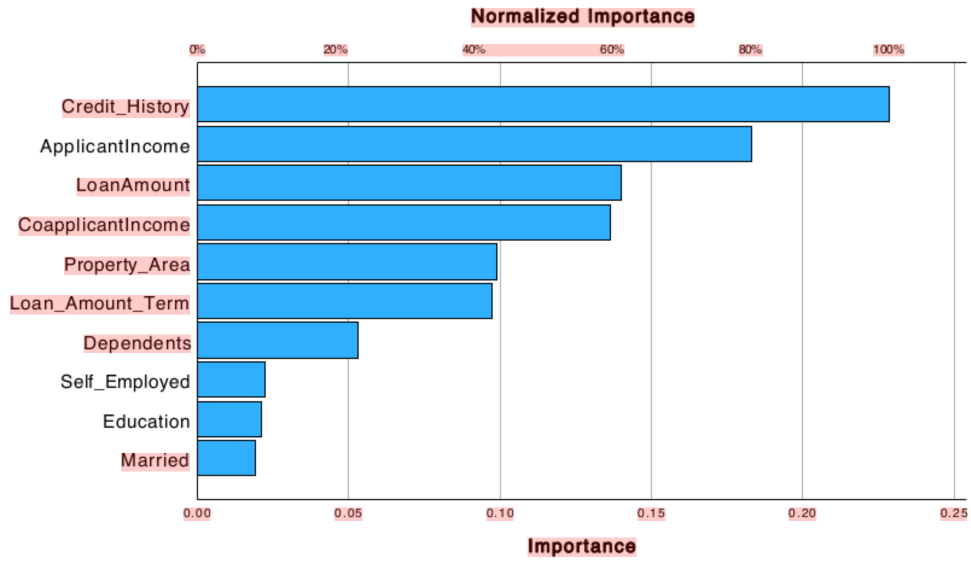
Classification

Sample	Observed	Predicted N	Predicted Y	Predicted Percent Correct
Training	N	29	25	53.7%
	Y	2	148	98.7%
	Overall Percent	15.2%	84.8%	86.8%
Testing	N	13	17	43.3%
	Y	1	61	98.4%
	Overall Percent	15.2%	84.8%	80.4%

Dependent Variable: Loan_Status

Independent Variable Importance

	Importance	Normalized Importance
Married	.019	8.4%
Dependents	.053	23.3%
Education	.021	9.2%
Self_Employed	.022	9.8%
Credit_History	.229	100.0%
Property_Area	.099	43.2%
ApplicantIncome	.183	80.1%
CoapplicantIncome	.136	59.6%
LoanAmount	.140	61.2%
Loan_Amount_Term	.098	42.7%



Multilayer Perceptron

Notes

Output Created		11-DEC-2024 08:18:30
Comments		
Input	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	85
Missing Value Handling	Definition of Missing	User- and system-missing values are treated as missing.
	Cases Used	Statistics are based on cases with valid data for all variables used by the procedure.
Weight Handling		not applicable

Notes

Syntax		MLP Loan_Status (MLEVEL=N) BY Gender Married Dependents Education Self_Employed Credit_History Property_Area WITH ApplicantIncome CoapplicantIncome LoanAmount Loan_Amount_Term /RESCALE COVARIATE=STANDARDIZED /PARTITION TRAINING=7 TESTING=3 HOLDOUT=0 /ARCHITECTURE AUTOMATIC=YES (MINUNITS=1 MAXUNITS=50) /CRITERIA TRAINING=BATCH OPTIMIZATION=SCALEDCONJUGATE LAMBDAINITIAL=0.0000005 SIGMAINITIAL=0.00005 INTERVALCENTER=0 INTERVALOFFSET=0.5 MEMSIZE=1000 /PRINT CPS NETWORKINFO SUMMARY CLASSIFICATION IMPORTANCE /PLOT NETWORK /STOPPINGRULES ERRORSTEPS=1 (DATA=AUTO) TRAININGTIMER=ON (MAXTIME=15) MAXEPOCHS=AUTO ERRORCHANGE=1.0E-4 ERRORRATIO=0.001 /MISSING USERMISSING=EXCLUDE .
Resources	Processor Time	00:00:01.35
	Elapsed Time	00:00:00.00

[DataSet1]

Warnings

The following independent variables are constant in the training sample and are excluded from the analysis: Gender.

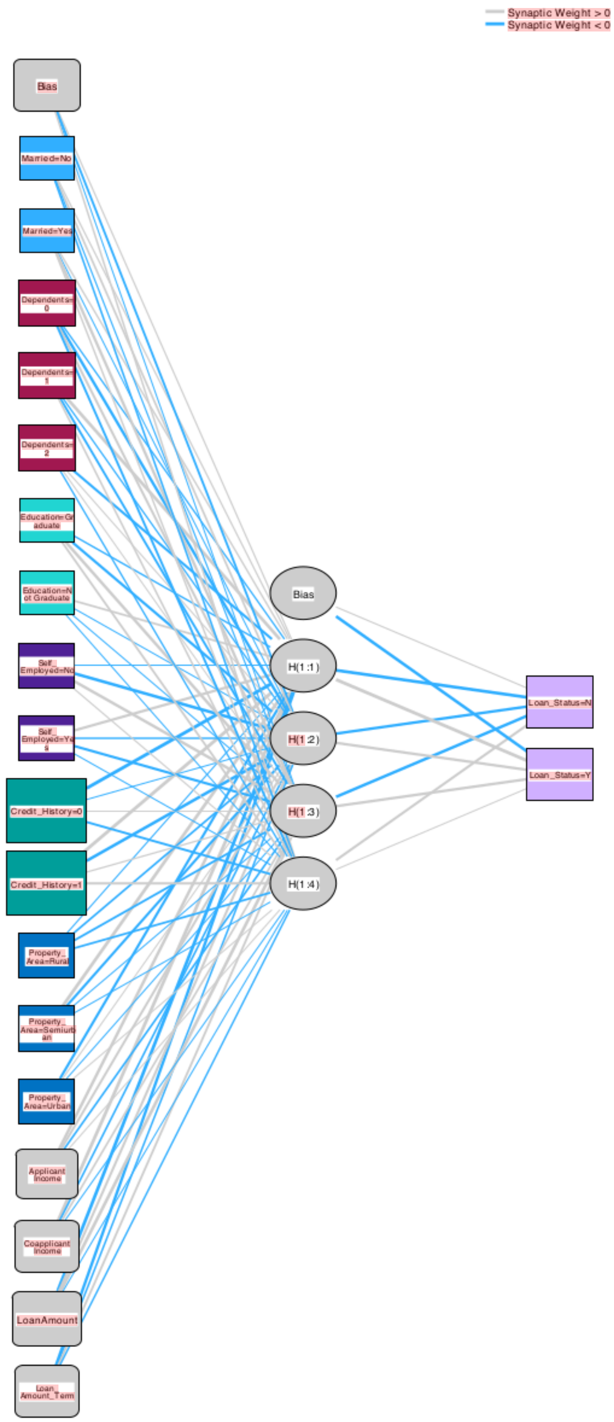
Case Processing Summary

		N	Percent
Sample	Training	59	69.4%
	Testing	26	30.6%
Valid		85	100.0%
Excluded		0	
Total		85	

Network Information

Input Layer	Factors	1	Married
		2	Dependents
		3	Education
		4	Self_Employed
		5	Credit_History
		6	Property_Area
	Covariates	1	ApplicantIncome
		2	CoapplicantIncome
		3	LoanAmount
		4	Loan_Amount_Term
	Number of Units[a]		18
	Rescaling Method for Covariates		Standardized
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1[a]		4
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	Loan_Status
	Number of Units		2
	Activation Function		Softmax
	Error Function		Cross-entropy

a. Excluding the bias unit



Hidden layer activation function: Hyperbolic tangent

Output layer activation function: Softmax

Model Summary

Training	Cross Entropy Error	18.948
	Percent Incorrect Predictions	11.9%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error[a]
	Training Time	0:00:00.04
Testing	Cross Entropy Error	9.645
	Percent Incorrect Predictions	15.4%

Dependent Variable: Loan_Status

a. Error computations are based on the testing sample.

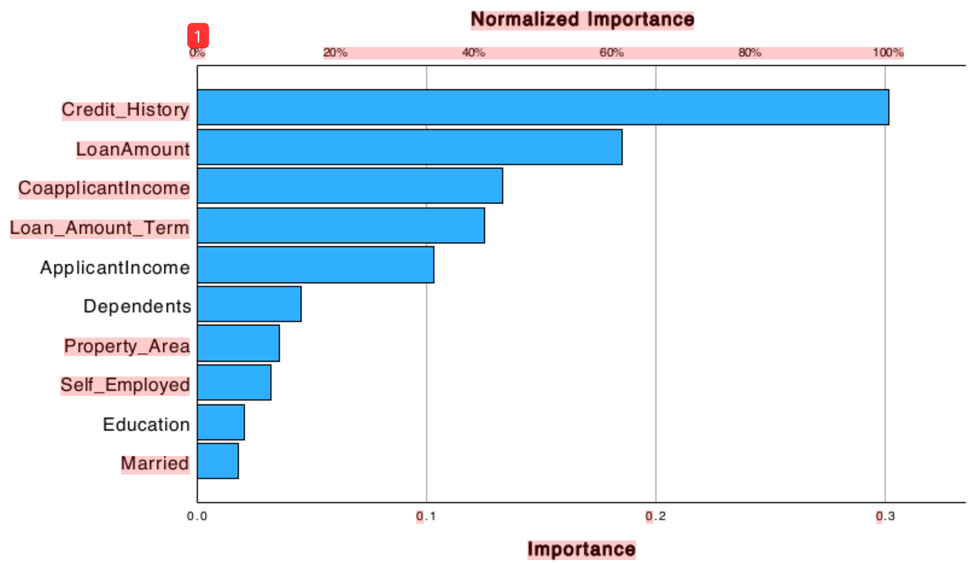
Classification

Sample	Observed	Predicted N	Predicted Y	Predicted Percent Correct
Training	N	14	4	77.8%
	Y	3	38	92.7%
	Overall Percent	28.8%	71.2%	88.1%
Testing	N	4	4	50.0%
	Y	0	18	100.0%
	Overall Percent	15.4%	84.6%	84.6%

Dependent Variable: Loan_Status

Independent Variable Importance

	Importance	Normalized Importance
Married	.018	5.9%
Dependents	.045	15.1%
Education	.021	6.9%
Self_Employed	.032	10.7%
Credit_History	.301	100.0%
Property_Area	.036	11.9%
ApplicantIncome	.103	34.1%
CoapplicantIncome	.133	44.1%
LoanAmount	.185	61.4%
Loan_Amount_Term	.126	41.6%


Normalized Importance
0%
20%
40%
60%
80%
100%
Credit_History
LoanAmount
CoapplicantIncome
Loan_Amount_Term
ApplicantIncome
Dependents
Property_Area
Self_Employed
Education
Married
0.0
0.1
0.2
0.3
Importance

Quick Cluster

Notes

Output Created		12-DEC-2024 10:17:59
Comments		
Input	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	381
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics are based on cases with no missing values for any clustering variable used.
Syntax		QUICK CLUSTER ApplicantIncome CoapplicantIncome LoanAmount Loan_Amount_Term Credit_History Gender2 Married2 Education2 Self_Emplyed2 /MISSING=LISTWISE /CRITERIA=CLUSTER(4) MXITER(100) CONVERGE(0) /METHOD=KMEANS(NOUPDATE) /SAVE CLUSTER DISTANCE /PRINT INITIAL.
Resources	Processor Time	00:00:00.03
	Elapsed Time	00:00:00.00
	Workspace Required	2072 bytes
Variables Created or Modified	QCL_1	Cluster Number of Case
	QCL_2	Distance of Case from its Classification Cluster Center

Initial K-means Cluster

Initial Cluster Centers

	Cluster 1	2	3	4
ApplicantIncome	2583	4917	3000	4583
CoapplicantIncome	8980.00000	.000000000	20000.0000	33837.0000
LoanAmount	120	130	66	128
Loan_Amount_Term	360	360	360	360
Credit_History	1	0	1	1
Gender2	1.00	1.00	1.00	1.00
Married2	1.00	.00	1.00	1.00
Education2	.00	1.00	1.00	1.00
Self_Emplyed2	.00	.00	1.00	.00

Iteration History[a]

Iteration	Change in Cluster Centers 1	2	3	4
1	3391.527	1685.532	.000	.000
2	1073.731	84.783	.000	.000
3	725.382	104.029	.000	.000
4	582.834	149.481	.000	.000
5	442.845	194.484	.000	.000
6	221.752	129.662	.000	.000
7	151.295	106.718	.000	.000
8	107.756	87.070	.000	.000
9	14.652	11.263	.000	.000
10	.000	.000	.000	.000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 10. The minimum distance between initial centers is 9278.365.

Final Cluster Centers

	Cluster 1	2	3	4
ApplicantIncome	3240	3829	3000	4583
CoapplicantIncome	2383.73168	224.881284	20000.0000	33837.0000
LoanAmount	103	107	66	128
Loan_Amount_Term	346	338	360	360
Credit_History	1	1	1	1
Gender2	1.21	1.23	1.00	1.00
Married2	.62	.58	1.00	1.00
Education2	.72	.73	1.00	1.00
Self_Emplyed2	.09	.09	1.00	.00

Number of Cases in each Cluster

Cluster	1	161.000
	2	218.000
	3	1.000
	4	1.000
Valid		381.000
Missing		.000

Quick Cluster

Notes

Output Created		12-DEC-2024 10:17:59
Comments		
Input	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	381
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics are based on cases with no missing values for any clustering variable used.
Syntax		QUICK CLUSTER ApplicantIncome CoapplicantIncome LoanAmount Loan_Amount_Term Credit_History Gender2 Married2 Education2 Self_Emplyed2 /MISSING=LISTWISE /CRITERIA=CLUSTER(4) MXITER(100) CONVERGE(0) /METHOD=KMEANS(NOUPDATE) /SAVE CLUSTER DISTANCE /PRINT INITIAL.
Resources	Processor Time	00:00:00.03
	Elapsed Time	00:00:00.00
	Workspace Required	2072 bytes
Variables Created or Modified	QCL_1	Cluster Number of Case
	QCL_2	Distance of Case from its Classification Cluster Center

Initial K-means Cluster

Initial Cluster Centers

	Cluster 1	2	3	4
ApplicantIncome	2583	4917	3000	4583
CoapplicantIncome	8980.00000	.000000000	20000.0000	33837.0000
LoanAmount	120	130	66	128
Loan_Amount_Term	360	360	360	360
Credit_History	1	0	1	1
Gender2	1.00	1.00	1.00	1.00
Married2	1.00	.00	1.00	1.00
Education2	.00	1.00	1.00	1.00
Self_Emplyed2	.00	.00	1.00	.00

Iteration History[a]

Iteration	Change in Cluster Centers 1	2	3	4
1	3391.527	1685.532	.000	.000
2	1073.731	84.783	.000	.000
3	725.382	104.029	.000	.000
4	582.834	149.481	.000	.000
5	442.845	194.484	.000	.000
6	221.752	129.662	.000	.000
7	151.295	106.718	.000	.000
8	107.756	87.070	.000	.000
9	14.652	11.263	.000	.000
10	.000	.000	.000	.000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 10. The minimum distance between initial centers is 9278.365.

Final Cluster Centers

	Cluster 1	2	3	4
ApplicantIncome	3240	3829	3000	4583
CoapplicantIncome	2383.73168	224.881284	20000.0000	33837.0000
LoanAmount	103	107	66	128
Loan_Amount_Term	346	338	360	360
Credit_History	1	1	1	1
Gender2	1.21	1.23	1.00	1.00
Married2	.62	.58	1.00	1.00
Education2	.72	.73	1.00	1.00
Self_Emplyed2	.09	.09	1.00	.00

Number of Cases in each Cluster

Cluster	1	161.000
	2	218.000
	3	1.000
	4	1.000
Valid		381.000
Missing		.000

Quick Cluster

Notes

Output Created		12-DEC-2024 10:23:52
Comments		
Input	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	379
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics are based on cases with no missing values for any clustering variable used.
Syntax		QUICK CLUSTER ApplicantIncome CoapplicantIncome LoanAmount Loan_Amount_Term Credit_History Gender2 Married2 Education2 Self_Emplyed2 /MISSING=LISTWISE /CRITERIA=CLUSTER(4) MXITER(100) CONVERGE(0) /METHOD=KMEANS(NOUPDATE) /SAVE CLUSTER DISTANCE /PRINT INITIAL.
Resources	Processor Time	00:00:00.03
	Elapsed Time	00:00:00.00
	Workspace Required	2072 bytes
Variables Created or Modified	QCL_3	Cluster Number of Case
	QCL_4	Distance of Case from its Classification Cluster Center

Initial Cluster Centers

	Cluster			
	1	2	3	4
ApplicantIncome	2583	4917	150	9703
CoapplicantIncome	8980.00000	.000000000	1710.00000	1516.00000
LoanAmount	120	130	135	112
Loan_Amount_Term	360	360	360	360
Credit_History	1	0	1	1
Gender2	1.00	1.00	1.00	1.00
Married2	1.00	.00	1.00	1.00
Education2	.00	1.00	1.00	1.00
Self_Emplyed2	.00	.00	.00	.00

Iteration History[a]

	Change in Cluster Centers			
Iteration	1	2	3	4
1	2874.759	1335.203	2009.366	1545.672
2	1012.183	182.722	269.547	1129.454
3	698.132	187.116	129.988	396.451
4	484.371	245.970	118.364	309.332
5	196.576	297.788	244.565	224.480
6	.000	162.980	171.454	101.012
7	99.092	50.572	55.930	53.634
8	65.658	20.311	14.185	53.154
9	39.450	36.734	25.965	47.969
10	.000	20.899	12.573	67.392
11	.000	29.417	12.244	45.777
12	.000	38.243	11.110	64.923
13	.000	20.861	.000	41.291
14	.000	.000	.000	.000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 14. The minimum distance between initial centers is 5020.396.

Final Cluster Centers

	Cluster 1	2	3	4
ApplicantIncome	3568	2962	3042	5793
CoapplicantIncome	4057.82143	67.9398675	1980.00000	787.219178
LoanAmount	109	101	102	117
Loan_Amount_Term	347	339	347	334
Credit_History	1	1	1	1
Gender2	1.18	1.23	1.24	1.19
Married2	.57	.58	.64	.58
Education2	.68	.70	.75	.77
Self_Emplyed2	.07	.03	.08	.23

Number of Cases in each Cluster

Cluster	1	28.000
	2	151.000
	3	127.000
	4	73.000
Valid		379.000
Missing		.000

Cluster Silhouettes

Notes

Output Created		12-DEC-2024 10:36:43
Comments		
Input	File Label	Aggregated File
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	2
Syntax		BEGIN PROGRAM '# '.
Resources	Processor Time	00:00:00.01
	Elapsed Time	00:00:00.00

Silhouette Statistics

Cluster	Case Count	Mean (Statistics)	Minimum	Maximum
1	161.000	.380	.034	.550
2	218.000	.370	-.065	.514
Total	379.000	.374	-.065	.550

Dissimilarity measure = Euclid

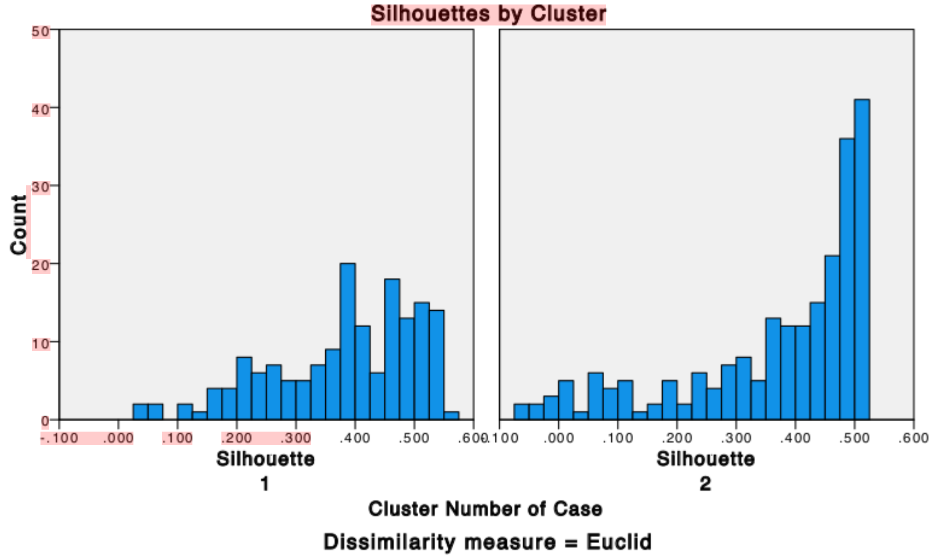
GGraph

Notes

Output Created		12-DEC-2024 10:36:43
Comments		
Input	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	379
Syntax		GGRAPH /GRAPHDATASET NAME="graphdataset" VARIABLES=QCL_1 [LEVEL=nominal] SV [LEVEL=ratio] MISSING=LISTWISE REPORTMISSING=NO /GRAPHSPEC SOURCE=VIZTEMPLATE (NAME="Histogram" [LOCATION=LOCAL] MAPPING("x"="SV" [DATASET=" graphdataset"] "Footnote"=" Dissimilarity measure = Euclid" "Title"=" Silhouettes by Cluster" "Panel_"+ "across"="QCL_1" [DATASET=" graphdataset"] "Summary"="count")) VIZSTYLESHEET=" Traditional" [LOCATION=LOCAL] LABEL='Histogram of Silhouettes' DEFAULTTEMPLATE=NO.
Resources	Processor Time	00:00:01.19
	Elapsed Time	00:00:01.00

[DataSet1]

Silhouettes by Cluster



GGraph

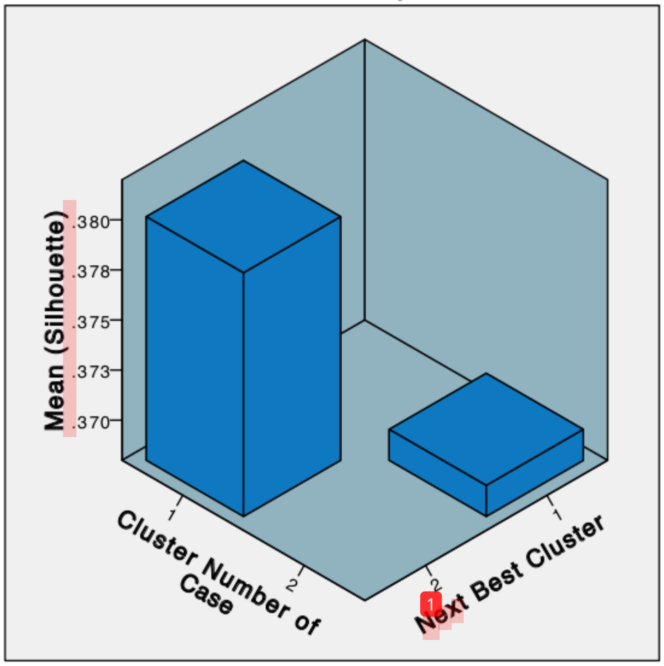
Notes

Output Created		12-DEC-2024 10:36:44
Comments		
Input	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	379

Notes

Syntax		GGRAPH /GRAPHDATASET NAME="graphdataset" VARIABLES=QCL_1 [LEVEL=nominal] NBC [LEVEL=nominal] SV [LEVEL=ratio] MISSING=LISTWISE REPORTMISSING=NO /GRAPHSPEC SOURCE=VIZTEMPLATE (NAME="3-D Bar"[LOCATION=LOCAL] MAPPING("x"="QCL_1"[DATASET=" graphdataset"] "y"="SV"[DATASET=" graphdataset"] "z"="NBC"[DATASET=" graphdataset"] "Summary"="mean" "Footnote"='Dissimilarity measure = Euclid' "Title"="Mean Silhouette by Cluster")) VIZSTYLESHEET=" Traditional"[LOCATION=LOCAL] LABEL='3-D BAR: Silhouettes' DEFAULTTEMPLATE=NO.
Resources	Processor Time	00:00:00.20
	Elapsed Time	00:00:00.00

Mean Silhouette by Cluster



Dissimilarity measure = Euclid

Quick Cluster

Notes

Output Created		12-DEC-2024 10:17:59
Comments		
Input	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	381
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics are based on cases with no missing values for any clustering variable used.
Syntax		QUICK CLUSTER ApplicantIncome CoapplicantIncome LoanAmount Loan_Amount_Term Credit_History Gender2 Married2 Education2 Self_Emplyed2 /MISSING=LISTWISE /CRITERIA=CLUSTER(4) MXITER(100) CONVERGE(0) /METHOD=KMEANS(NOUPDATE) /SAVE CLUSTER DISTANCE /PRINT INITIAL.
Resources	Processor Time	00:00:00.03
	Elapsed Time	00:00:00.00
	Workspace Required	2072 bytes
Variables Created or Modified	QCL_1	Cluster Number of Case
	QCL_2	Distance of Case from its Classification Cluster Center

Initial K-means Cluster

Initial Cluster Centers

	Cluster 1	2	3	4
ApplicantIncome	2583	4917	3000	4583
CoapplicantIncome	8980.00000	.000000000	20000.0000	33837.0000
LoanAmount	120	130	66	128
Loan_Amount_Term	360	360	360	360
Credit_History	1	0	1	1
Gender2	1.00	1.00	1.00	1.00
Married2	1.00	.00	1.00	1.00
Education2	.00	1.00	1.00	1.00
Self_Emplyed2	.00	.00	1.00	.00

Iteration History[a]

Iteration	Change in Cluster Centers 1	2	3	4
1	3391.527	1685.532	.000	.000
2	1073.731	84.783	.000	.000
3	725.382	104.029	.000	.000
4	582.834	149.481	.000	.000
5	442.845	194.484	.000	.000
6	221.752	129.662	.000	.000
7	151.295	106.718	.000	.000
8	107.756	87.070	.000	.000
9	14.652	11.263	.000	.000
10	.000	.000	.000	.000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 10. The minimum distance between initial centers is 9278.365.

Final Cluster Centers

	Cluster 1	2	3	4
ApplicantIncome	3240	3829	3000	4583
CoapplicantIncome	2383.73168	224.881284	20000.0000	33837.0000
LoanAmount	103	107	66	128
Loan_Amount_Term	346	338	360	360
Credit_History	1	1	1	1
Gender2	1.21	1.23	1.00	1.00
Married2	.62	.58	1.00	1.00
Education2	.72	.73	1.00	1.00
Self_Emplyed2	.09	.09	1.00	.00

Number of Cases in each Cluster

Cluster	1	161.000
	2	218.000
	3	1.000
	4	1.000
Valid		381.000
Missing		.000

Quick Cluster

Notes

Output Created		12-DEC-2024 10:23:52
Comments		
Input	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	379
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics are based on cases with no missing values for any clustering variable used.
Syntax		QUICK CLUSTER ApplicantIncome CoapplicantIncome LoanAmount Loan_Amount_Term Credit_History Gender2 Married2 Education2 Self_Emplyed2 /MISSING=LISTWISE /CRITERIA=CLUSTER(4) MXITER(100) CONVERGE(0) /METHOD=KMEANS(NOUPDATE) /SAVE CLUSTER DISTANCE /PRINT INITIAL.
Resources	Processor Time	00:00:00.03
	Elapsed Time	00:00:00.00
	Workspace Required	2072 bytes
Variables Created or Modified	QCL_3	Cluster Number of Case
	QCL_4	Distance of Case from its Classification Cluster Center

Initial Cluster Centers

	Cluster			
	1	2	3	4
ApplicantIncome	2583	4917	150	9703
CoapplicantIncome	8980.00000	.000000000	1710.00000	1516.00000
LoanAmount	120	130	135	112
Loan_Amount_Term	360	360	360	360
Credit_History	1	0	1	1
Gender2	1.00	1.00	1.00	1.00
Married2	1.00	.00	1.00	1.00
Education2	.00	1.00	1.00	1.00
Self_Emplyed2	.00	.00	.00	.00

Iteration History[a]

Iteration	Change in Cluster Centers			
	1	2	3	4
1	2874.759	1335.203	2009.366	1545.672
2	1012.183	182.722	269.547	1129.454
3	698.132	187.116	129.988	396.451
4	484.371	245.970	118.364	309.332
5	196.576	297.788	244.565	224.480
6	.000	162.980	171.454	101.012
7	99.092	50.572	55.930	53.634
8	65.658	20.311	14.185	53.154
9	39.450	36.734	25.965	47.969
10	.000	20.899	12.573	67.392
11	.000	29.417	12.244	45.777
12	.000	38.243	11.110	64.923
13	.000	20.861	.000	41.291
14	.000	.000	.000	.000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 14. The minimum distance between initial centers is 5020.396.

Final Cluster Centers

	Cluster 1	2	3	4
ApplicantIncome	3568	2962	3042	5793
CoapplicantIncome	4057.82143	67.9398675	1980.00000	787.219178
LoanAmount	109	101	102	117
Loan_Amount_Term	347	339	347	334
Credit_History	1	1	1	1
Gender2	1.18	1.23	1.24	1.19
Married2	.57	.58	.64	.58
Education2	.68	.70	.75	.77
Self_Emplyed2	.07	.03	.08	.23

Number of Cases in each Cluster

Cluster	1	28.000
	2	151.000
	3	127.000
	4	73.000
Valid		379.000
Missing		.000

AI FOR BUSINESS_merged.pdf

ORIGINALITY REPORT

43%	26%	21%	42%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Queen Mary and Westfield College Student Paper	40%
2	Submitted to University of Birmingham Student Paper	1%
3	Submitted to University of Northumbria at Newcastle Student Paper	<1%
4	Submitted to UCL Student Paper	<1%
5	Submitted to Coventry University Student Paper	<1%
6	Submitted to University of Wales Institute, Cardiff Student Paper	<1%
7	www.coursehero.com Internet Source	<1%
8	Submitted to Tilburg University Student Paper	<1%

9	Dr. Swagato Chatterjee. "Explaining customer ratings and recommendations by combining qualitative and quantitative user generated contents", Decision Support Systems, 2019 Publication	<1%
10	www.clearcogs.com Internet Source	<1%
11	www.escholar.manchester.ac.uk Internet Source	<1%
12	core.ac.uk Internet Source	<1%
13	www.tojqi.net Internet Source	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		

AI FOR BUSINESS_merged.pdf

GRADEMARK REPORT

FINAL GRADE

74/100

GENERAL COMMENTS

Part	Qtn number	Marks Obtained	Full Marks	C
Part A	a	5	5	
	b	9	10	
	c	8	10	M t
Part B	a	8	10	
	b	8	10	
	c	6	10	A r c s g
Part C	a	9	15	V r
	b	9	15	V r
	c	7	10	V r
Presentation		5	5	V
Total		74	100	

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17

PAGE 18

PAGE 19

PAGE 20

PAGE 21

PAGE 22

PAGE 23

PAGE 24

PAGE 25

PAGE 26

PAGE 27

PAGE 28

PAGE 29

PAGE 30

PAGE 31

PAGE 32

PAGE 33

PAGE 34

PAGE 35

PAGE 36

PAGE 37

PAGE 38

PAGE 39

PAGE 40

PAGE 41

PAGE 42

PAGE 43

PAGE 44

PAGE 45

PAGE 46

PAGE 47

PAGE 48

PAGE 49

PAGE 50

PAGE 51

PAGE 52

PAGE 53

PAGE 54

PAGE 55

PAGE 56

PAGE 57

PAGE 58

PAGE 59

PAGE 60

PAGE 61

PAGE 62

PAGE 63

PAGE 64

PAGE 65

PAGE 66

PAGE 67

PAGE 68

PAGE 69

PAGE 70