Applied Analytic Modeling

Group Project- Loan Set Model

Model Assessment-SAS Miner

Submitted to

Prof. David Parent

Babasola Osikoya 301316869

Tamanna Oporajita 301253651

Awara Uchechukwu A. 301333712

Aderonke Comfort Onipede 301338997

TABLE OF CONTENT

INTRODUCTION

The choice to approve a loan is crucial for both banks and borrowers. Efficient loan approval procedures are essential for banks and other organizations to maintain profitability and lower the chance of defaults. Loan approvals are necessary for borrowers to fulfill their financial objectives.

In this project, we make decisions about loan approval for clients using SAS. Our goal is to create a model that can determine how likely a borrower is to repay a loan. The bank can now make better-informed decisions as a result.

This model's successful use will greatly impact banks and borrowers alike. It will result in lower loan default rates, higher loan acceptance rates, and better risk management techniques for banks. It will make credit more accessible to borrowers and make it easier for them to realize their financial goals.

We will use a strict research process in this capstone project that includes data collection, data exploration, model construction, model evaluation, and model deployment. Our goal is to create a reliable and broadly applicable model that can accurately forecast the results of loan approvals and support the banking sector.

OBJECTIVES

Creating a model: This will be used to forecast the approval or denial of a loan.
Boost approval rates: The model will make bank approval rates higher.
Cut down on loan defaults:

Customized loan selections: Make personalized loan decisions by applying the predictive model to inform your decision-making and accounting for the distinct features and financial profiles of each application.

EXPLORATORY ANALYSIS

METHODOLOGY

Reviewing the data, we can confirm that there is no bad data. The data seem to be within the normal range and analysis can be done on it.

We used the Loan Prediction Dataset in our machine learning project to create a predictive model that will automate and improve the loan approval procedure. This dataset provides a comprehensive perspective of the attributes of loan applicants by encapsulating important information about them.

Let us examine the salient characteristics.

1. Loan ID: An exclusive number assigned to every loan application that makes thorough monitoring and referencing easier.

2. Gender: Gain insight into the applicant's gender and identify any possible relationships with loan acceptance.

3. Married: A signal indicating marital status, which is frequently taken into consideration when determining whether to approve a loan.

4. Dependents: Unveiling the number of dependents associated with an applicant, revealing insight into familial duties.

5. Education: A summary of the applicant's educational background that distinguishes between graduates and non-graduates.

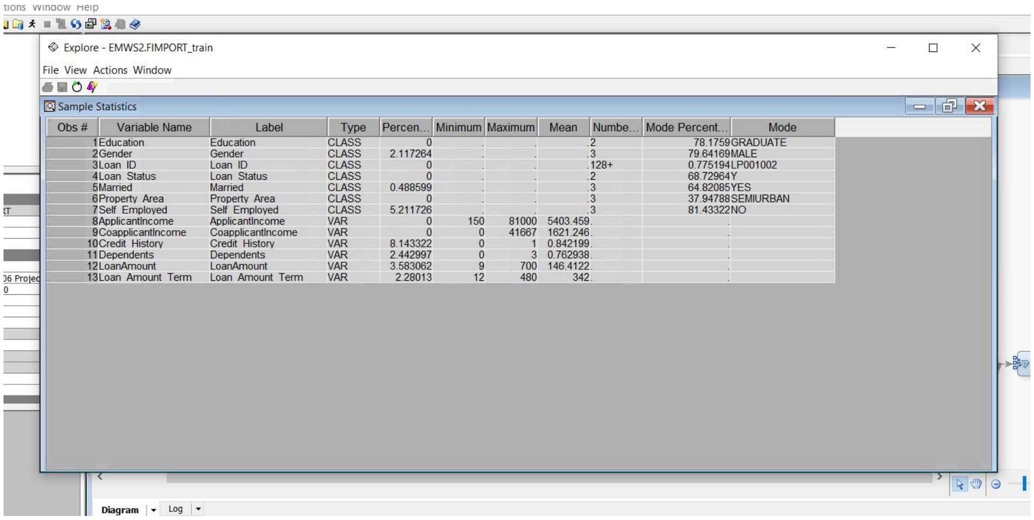
6. Self-Employed: Indicates in binary terms if the applicant works for themselves, providing information about their professional standing.

7. Applicant Income: A numerical depiction of the applicant's overall income, which is a crucial component in determining their financial capability.

8. Co-applicant Income: This feature tracks any co-applicant income like the applicant's income.

9. Loan Amount: The applicant's requested financial amount, which is a deciding factor in the loan

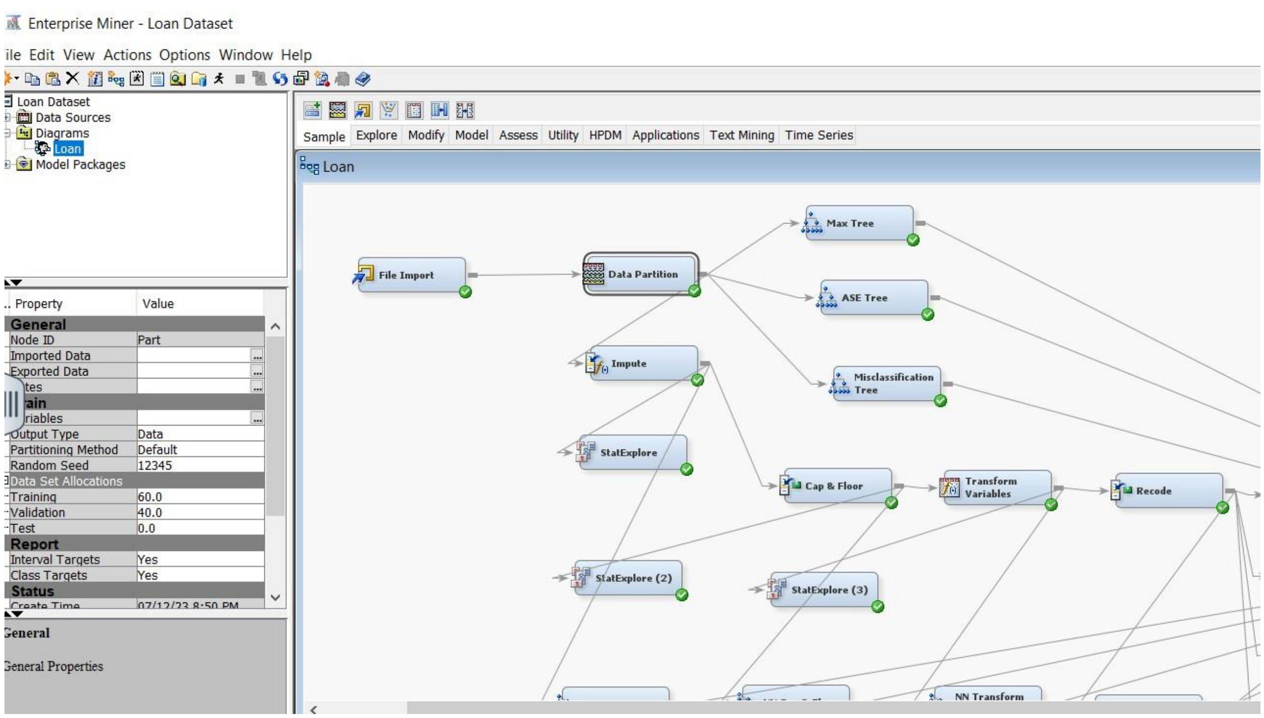
approval process.



The data we used are within acceptable parameters.

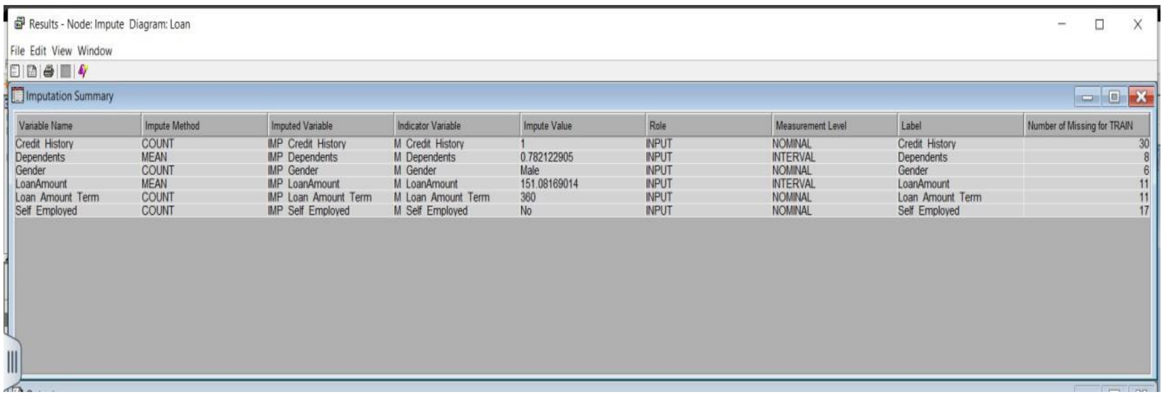
DATA PARTITION

The data was split into 60% training and 40% validation. This ensures that there is more data to train our model on.



DATA IMPUTATION

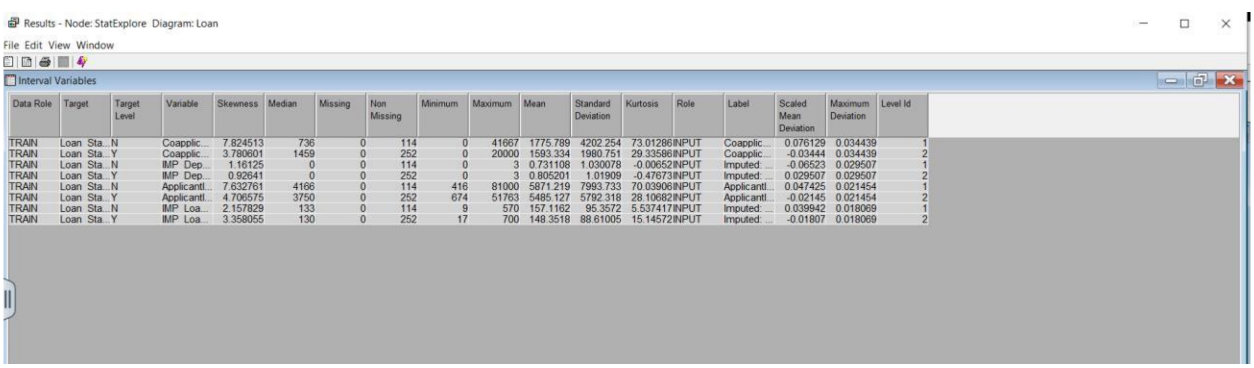
In our commitment to making well-informed loan decisions, we recognize that missing data can compromise the accuracy of our analysis. To address this, we employ a sophisticated technique known as imputation, which utilizes statistical methods to estimate and fill in any gaps in the data. This ensures that we have a more complete and precise understanding of each borrower’s financial profile. By enhancing the quality of our data, we can make more equitable and informed decisions regarding loan approvals.


Results - Node: Impute Diagram: Loan
File Edit View Window
Imputation Summary

Variable Name	Impute Method	Imputed Variable	Indicator Variable	Impute Value	Role	Measurement Level	Label	Number of Missing for TRAIN
Credit History	COUNT	IMP Credit History	M Credit History	1	INPUT	NOMINAL	Credit History	30
Dependents	MEAN	IMP Dependents	M Dependents	0.782122905	INPUT	INTERVAL	Dependents	8
Gender	COUNT	IMP Gender	M Gender	Male	INPUT	NOMINAL	Gender	6
LoanAmount	MEAN	IMP LoanAmount	M LoanAmount	151.08169014	INPUT	INTERVAL	LoanAmount	11
Loan Amount Term	COUNT	IMP Loan Amount Term	M Loan Amount Term	360	INPUT	NOMINAL	Loan Amount Term	11
Self Employed	COUNT	IMP Self Employed	M Self Employed	No	INPUT	NOMINAL	Self Employed	17

SKEWNESS THRESHOLD OF 1

Loan companies rely heavily on data analysis, but biased data can lead to unfair decisions. Describes identifying such bias, where information leans towards an extreme (e.g., mostly high incomes). This skews the model and could disadvantage some borrowers. Cleaning the data, such as skewness techniques helps ensure a fairer assessment, ultimately influencing whether a loan is granted.


Results - Node: StatExplore Diagram: Loan
File Edit View Window
Interval Variables

Data Role	Target	Target Level	Variable	Skewness	Median	Missing	Non Missing	Minimum	Maximum	Mean	Standard Deviation	Kurtosis	Role	Label	Scaled Mean Deviation	Maximum Deviation	Level Id
TRAIN	Loan Sta...	N	Coapplic...	7.824513	736	0	114	0	41667	1775.789	4202.254	73.01286	INPUT	Coapplic...	0.076129	0.034439	1
TRAIN	Loan Sta...	Y	Coapplic...	3.780601	1459	0	252	0	20000	1593.334	1980.751	29.33586	INPUT	Coapplic...	-0.03444	0.034439	2
TRAIN	Loan Sta...	N	IMP Dep...	1.16125	0	0	114	0	3	0.731108	1.030078	-0.00652	INPUT	Imputed: ...	-0.06523	0.029507	1
TRAIN	Loan Sta...	Y	IMP Dep...	0.92641	0	0	252	0	3	0.805201	1.01909	-0.47673	INPUT	Imputed: ...	0.029507	0.029507	2
TRAIN	Loan Sta...	N	Applicantl...	7.632761	4166	0	114	416	81000	5871.219	7993.733	70.03906	INPUT	Applicantl...	0.047425	0.021454	1
TRAIN	Loan Sta...	Y	Applicantl...	4.706575	3750	0	252	674	51763	5485.127	5792.318	28.10682	INPUT	Applicantl...	-0.02145	0.021454	2
TRAIN	Loan Sta...	N	IMP Loa...	2.157829	133	0	114	9	570	157.1162	95.3572	5.537417	INPUT	Imputed: ...	0.039942	0.018069	1
TRAIN	Loan Sta...	Y	IMP Loa...	3.358055	130	0	252	17	700	148.3518	88.61005	15.14572	INPUT	Imputed: ...	-0.01807	0.018069	2

OUTLIERS AND EXTREME VALUES

Our initial data exploration revealed a potential bias in the loan application data set. The values skewed towards one extreme, which could lead to inaccurate assessments and unfair lending practices. To address this, we're implementing 'Cap & Floor' techniques to establish reasonable limits on the data points. While this is a positive step, further data cleaning procedures may be necessary to guarantee a fully unbiased and reliable data set for informed loan approval decisions.

Results - Node: Cap & Floor Diagram: Loan

File Edit View Window

Interval Variables

Variable	Replace Variable	Lower limit	Upper Limit	Label	Limits Method	Replacement Method	Lower Replacement Value	Upper Replacement Value
ApplicantIncome	REP ApplicantIncome	-14041	25251.78	ApplicantIncome	STDDEV	COMPUTED	-14041	25251.78
CoapplicantIncome	REP CoapplicantIncome	-6925.93	10226.26	CoapplicantIncome	STDDEV	COMPUTED	-6925.93	10226.26
IMP Dependents	REP IMP Dependents	-2.28295	3.847194	Imputed: Dependents	STDDEV	COMPUTED	-2.28295	3.847194
IMP LoanAmount	REP IMP LoanAmount	-121.093	423.2567	Imputed: LoanAmount	STDDEV	COMPUTED	-121.093	423.2567

Results - Node: StatExplore (2) Diagram: Loan

File Edit View Window

Interval Variables

Data Role	Target	Target Level	Variable	Skewness ▼	Median	Missing	Non Missing	Minimum	Maximum	Mean	Standard Deviation	Kurtosis	Role	Label	Scaled Mean Deviation	Maximum Deviation	Level Id
TRAIN	Loan Sta...	Y	REP Ap...	2.700742	3750	0	252	674	25251.78	5216.79	4215.807	8.396406	INPUT	Replace...	-0.00978	0.009779	2
TRAIN	Loan Sta...	N	REP Ap...	2.421778	4166	0	114	416	25251.78	5382.2	4127.058	6.880759	INPUT	Replace...	0.021618	0.009779	1
TRAIN	Loan Sta...	Y	REP IM...	2.006979	130	0	252	17	423.2567	144.6119	70.65879	5.294827	INPUT	Replace...	-0.01988	0.019879	2
TRAIN	Loan Sta...	N	REP Co...	1.957465	736	0	114	0	10226.26	1490.575	1988.882	5.359575	INPUT	Replace...	-0.0287	0.012985	1
TRAIN	Loan Sta...	N	REP IM...	1.617521	133	0	114	9	423.2567	154.0287	84.31311	2.629961	INPUT	Replace...	0.043944	0.019879	1
TRAIN	Loan Sta...	Y	REP Co...	1.493124	1459	0	252	0	10226.26	1554.549	1693.801	3.486603	INPUT	Replace...	0.012985	0.012985	2
TRAIN	Loan Sta...	N	REP IM...	1.16125	0	0	114	0	3	0.731108	1.030078	-0.00652	INPUT	Replace...	-0.06523	0.029507	1
TRAIN	Loan Sta...	Y	REP IM...	0.92641	0	0	252	0	3	0.805201	1.01909	-0.47673	INPUT	Replace...	0.029507	0.029507	2

DATA TRANSFORM

We identified a bias in the loan application data, where certain variables were skewed towards one extreme. To mitigate this and ensure a more balanced representation, we're employing a mathematical transformation technique on key variables like income and loan amount. This process effectively evens out the data distribution, creating a more accurate picture of each applicant's financial profile. By using unbiased and reliable data, we can make confident and responsible loan approval decisions.

Results - Node: StatExplore (3) Diagram: Loan

File Edit View Window

Interval Variables

Data Role	Target	Target Level	Variable	Skewness	Median	Missing	Non Missing	Minimum	Maximum	Mean	Standard Deviation	Kurtosis	Role	Label	Scaled Mean Deviation	Maximum Deviation	Level Id
TRAIN	Loan Status	N	LOG REP CoapplicantIncome	-0.00828	6.602588	0	114	0	9.232812	3.98652	3.951888	-1.99707	INPUT	Transforme...	-0.12155	0.054987	1
TRAIN	Loan Status	Y	LOG REP CoapplicantIncome	-0.47156	7.286192	0	252	0	9.232812	4.787665	3.740447	-1.72981	INPUT	Transforme...	0.054987	0.054987	2
TRAIN	Loan Status	N	LOG REP IMP Dependents	0.770629	0	0	114	0	1.386294	0.401936	0.519234	-0.9992	INPUT	Transforme...	-0.07258	0.032833	1
TRAIN	Loan Status	Y	LOG REP IMP Dependents	0.542026	0	0	252	0	1.386294	0.44762	0.521154	-1.31989	INPUT	Transforme...	0.032833	0.032833	2
TRAIN	Loan Status	N	LOG REP IMP LoanAmount	-0.70085	4.89784	0	114	2.302585	6.050339	4.913768	0.526469	4.603032	INPUT	Transforme...	0.004333	0.00196	1
TRAIN	Loan Status	Y	LOG REP IMP LoanAmount	-0.29296	4.875197	0	252	2.890372	6.050339	4.882975	0.444992	2.792316	INPUT	Transforme...	-0.00196	0.00196	2
TRAIN	Loan Status	N	LOG REP ApplicantIncome	0.225378	8.334952	0	114	6.033086	10.13669	8.385991	0.620769	1.559828	INPUT	Transforme...	0.00271	0.001226	1
TRAIN	Loan Status	Y	LOG REP ApplicantIncome	0.689438	8.229778	0	252	6.514713	10.13669	8.353077	0.600276	0.898077	INPUT	Transforme...	-0.00123	0.001226	2

DATA SIMPLIFICATION

Data efficiency is paramount in loan analysis. "RECODE" tackles this by consolidating a multitude of specific loan terms into just two major categories. This strategic grouping, achieved with specialized tools, simplifies analysis without sacrificing valuable insights. By streamlining data categorization, we gain a clearer view of repayment trends across different loan types. This empowers us to make informed decisions regarding future loan offerings and optimize our overall loan approval criteria.

Replacement Editor-WORK.OUTCLASS

Variable	Formatted Value	Replacement Value	Frequency Count	Type	Character Unformatted Value	Numeric Value
Education	Graduate		283	C	Graduate	.
Education	Not Graduate		83	C	Not Graduate	.
Education	_UNKNOWN_	_DEFAULT_	.	C		.
IMP_Credit_History	1		309	N		1
IMP_Credit_History	0		57	N		0
IMP_Credit_History	_UNKNOWN_	_DEFAULT_	.	N		.
IMP_Gender	Male		303	C	Male	.
IMP_Gender	Female		63	C	Female	.
IMP_Gender	_UNKNOWN_	_DEFAULT_	.	C		.
IMP_Loan_Amount_Term	360	360	317	N		360
IMP_Loan_Amount_Term	180	999	28	N		180
IMP_Loan_Amount_Term	480	999	6	N		480
IMP_Loan_Amount_Term	300	999	5	N		300
IMP_Loan_Amount_Term	84	999	3	N		84
IMP_Loan_Amount_Term	120	999	2	N		120
IMP_Loan_Amount_Term	240	999	2	N		240
IMP_Loan_Amount_Term	12	999	1	N		12
IMP_Loan_Amount_Term	36	999	1	N		36
IMP_Loan_Amount_Term	60	999	1	N		60
IMP_Loan_Amount_Term	_UNKNOWN_	_DEFAULT_	.	N		.
IMP_Self_Employed	No		316	C	No	.
IMP_Self_Employed	Yes		50	C	Yes	.
IMP_Self_Employed	_UNKNOWN_	_DEFAULT_	.	C		.
Loan_Status	Y		252	C	Y	.
Loan_Status	N		114	C	N	.
Loan_Status	_UNKNOWN_	_DEFAULT_	.	C		.
M_Credit_History	0		336	N		0
M_Credit_History	1		30	N		1

OK Cancel

ANALYSIS

TREE ANALYSIS

Our tree analysis identified credit history as the key driver for model performance, with both the ASE and misclassification trees achieving a strong score of 0.164458. These models outperformed the maximal tree (score: 0.18532) which, while comprehensive, might be overfitting the data (i.e. using more variables). This analysis highlights the optimal balance achieved between model complexity and efficiency.

MAX TREE

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation
Loan Status	Loan Status	NOBS	Sum of Frequencies	366	248
Loan Status	Loan Status	MISC	Misclassification Rate	0.166667	0.241935
Loan Status	Loan Status	MAX	Maximum Absolute Error	0.96	0.96
Loan Status	Loan Status	SSE	Sum of Squared Errors	95.33516	91.91865
Loan Status	Loan Status	ASE	Average Squared Error	0.130239	0.18532
Loan Status	Loan Status	RASE	Root Average Squared Error	0.360887	0.430488
Loan Status	Loan Status	DIV	Divisor for ASE	732	496
Loan Status	Loan Status	DFT	Total Degrees of Freedom	366	.

ASE TREE

Results - Node: ASE Tree Diagram: Loan

File Edit View Window

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Loan Status	Loan Status	NOBS	Sum of Frequencies	366	248	.
Loan Status	Loan Status	MISC	Misclassification Rate	0.177596	0.209677	.
Loan Status	Loan Status	MAX	Maximum Absolute Error	0.929825	0.929825	.
Loan Status	Loan Status	SSE	Sum of Squared Errors	105.3545	81.61603	.
Loan Status	Loan Status	ASE	Average Squared Error	0.143927	0.164548	.
Loan Status	Loan Status	RASE	Root Average Squared Error	0.379377	0.405646	.
Loan Status	Loan Status	DIV	Divisor for ASE	732	496	.
Loan Status	Loan Status	DFT	Total Degrees of Freedom	366	.	.

MISCLASSIFICATION TREE

Results - Node: Misclassification Tree Diagram: Loan

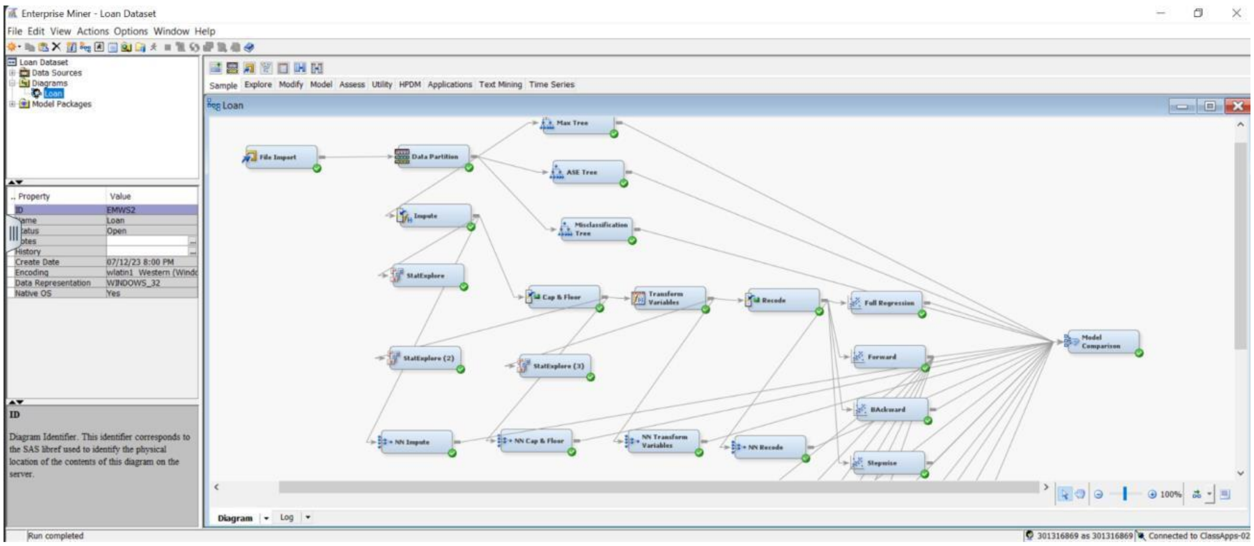
File Edit View Window

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Loan Status	Loan Status	NOBS	Sum of Frequencies	366	248	.
Loan Status	Loan Status	MISC	Misclassification Rate	0.177596	0.209677	.
Loan Status	Loan Status	MAX	Maximum Absolute Error	0.929825	0.929825	.
Loan Status	Loan Status	SSE	Sum of Squared Errors	105.3545	81.61603	.
Loan Status	Loan Status	ASE	Average Squared Error	0.143927	0.164548	.
Loan Status	Loan Status	RASE	Root Average Squared Error	0.379377	0.405646	.
Loan Status	Loan Status	DIV	Divisor for ASE	732	496	.
Loan Status	Loan Status	DFT	Total Degrees of Freedom	366	.	.

REGRESSION ANALYSIS

Our logistic regression analysis explored various approaches to identify the most impactful factors influencing the model. We compared the full regression model to forward selection, backward elimination, and stepwise regression. All models achieved strong performance, with scores ranging from 0.164548 to 0.172466. Notably, the forward selection and stepwise regression yielded identical average squared error (ASE) scores of 0.164548, suggesting a streamlined model can achieve comparable effectiveness. This analysis helps us optimize model complexity while maintaining robust performance.



FORWARD REGRESSION

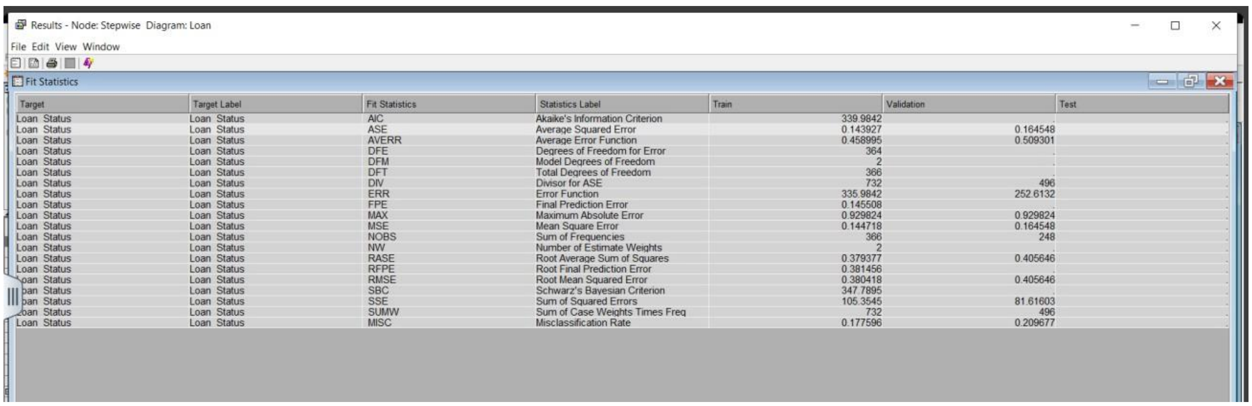
Results - Node: Forward Diagram: Loan

File Edit View Window

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Loan Status	Loan Status	AIC	Akaike's Information Criterion	339.9842	.	.
Loan Status	Loan Status	ASE	Average Squared Error	0.143927	0.164548	.
Loan Status	Loan Status	AVERR	Average Error Function	0.458995	0.509301	.
Loan Status	Loan Status	DFE	Degrees of Freedom for Error	364	.	.
Loan Status	Loan Status	DFM	Model Degrees of Freedom	2	.	.
Loan Status	Loan Status	DFT	Total Degrees of Freedom	366	.	.
Loan Status	Loan Status	DIV	Divisor for ASE	732	496	.
Loan Status	Loan Status	ERR	Error Function	335.9842	252.6132	.
Loan Status	Loan Status	FPE	Final Prediction Error	0.145508	.	.
Loan Status	Loan Status	MAX	Maximum Absolute Error	0.929824	0.929824	.
Loan Status	Loan Status	MSE	Mean Square Error	0.144718	0.164548	.
Loan Status	Loan Status	NOBS	Sum of Frequencies	366	248	.
Loan Status	Loan Status	NW	Number of Estimate Weights	2	.	.
Loan Status	Loan Status	RASE	Root Average Sum of Squares	0.379377	0.405646	.
Loan Status	Loan Status	RFPE	Root Final Prediction Error	0.381456	.	.
Loan Status	Loan Status	RMSE	Root Mean Squared Error	0.380418	0.405646	.
Loan Status	Loan Status	SBC	Schwarz's Bayesian Criterion	347.7895	.	.
Loan Status	Loan Status	SSE	Sum of Squared Errors	105.3545	81.61603	.
Loan Status	Loan Status	SUMW	Sum of Case Weights Times Freq	732	496	.
Loan Status	Loan Status	MISC	Misclassification Rate	0.177596	0.209677	.

STEPWISE REGRESSION



Results - Node: Stepwise Diagram: Loan

File Edit View Window

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Loan Status	Loan Status	AIC	Akaike's Information Criterion	339.9842	.	.
Loan Status	Loan Status	ASE	Average Squared Error	0.143927	0.164548	.
Loan Status	Loan Status	AVERR	Average Error Function	0.458995	0.509301	.
Loan Status	Loan Status	DFE	Degrees of Freedom for Error	364	.	.
Loan Status	Loan Status	DFM	Model Degrees of Freedom	2	.	.
Loan Status	Loan Status	DFT	Total Degrees of Freedom	366	.	.
Loan Status	Loan Status	DIV	Divisor for ASE	732	496	.
Loan Status	Loan Status	ERR	Error Function	335.9842	252.6132	.
Loan Status	Loan Status	FPE	Final Prediction Error	0.145508	.	.
Loan Status	Loan Status	MAX	Maximum Absolute Error	0.929824	0.929824	.
Loan Status	Loan Status	MSE	Mean Square Error	0.144718	0.164548	.
Loan Status	Loan Status	NOBS	Sum of Frequencies	366	248	.
Loan Status	Loan Status	NW	Number of Estimate Weights	2	.	.
Loan Status	Loan Status	RASE	Root Average Sum of Squares	0.379377	0.405646	.
Loan Status	Loan Status	RFPE	Root Final Prediction Error	0.381456	.	.
Loan Status	Loan Status	RMSE	Root Mean Squared Error	0.380418	0.405646	.
Loan Status	Loan Status	SBC	Schwarz's Bayesian Criterion	347.7895	.	.
Loan Status	Loan Status	SSE	Sum of Squared Errors	105.3545	81.61603	.
Loan Status	Loan Status	SUMW	Sum of Case Weights Times Freq	732	496	.
Loan Status	Loan Status	MISC	Misclassification Rate	0.177596	0.209677	.

FULL REGRESSION

Results - Node: Full Regression Diagram: Loan

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Loan Status	Loan Status	AIC	Akaike's Information Criterion	344.635	.	.
Loan Status	Loan Status	ASE	Average Squared Error	0.131961	0.172466	.
Loan Status	Loan Status	AVERR	Average Error Function	0.4189	0.540719	.
Loan Status	Loan Status	DFE	Degrees of Freedom for Error	347	.	.
Loan Status	Loan Status	DFM	Model Degrees of Freedom	19	.	.
Loan Status	Loan Status	DFT	Total Degrees of Freedom	366	.	.
Loan Status	Loan Status	DIV	Divisor for ASE	732	496	.
Loan Status	Loan Status	ERR	Error Function	306.635	268.1965	.
Loan Status	Loan Status	FPE	Final Prediction Error	0.146412	.	.
Loan Status	Loan Status	MAX	Maximum Absolute Error	0.966576	0.988345	.
Loan Status	Loan Status	MSE	Mean Square Error	0.139186	0.172466	.
Loan Status	Loan Status	NOBS	Sum of Frequencies	366	248	.
Loan Status	Loan Status	NW	Number of Estimate Weights	19	.	.
Loan Status	Loan Status	RASE	Root Average Sum of Squares	0.363264	0.41529	.
Loan Status	Loan Status	RFPE	Root Final Prediction Error	0.382638	.	.
Loan Status	Loan Status	RMSE	Root Mean Squared Error	0.373077	0.41529	.
Loan Status	Loan Status	SBC	Schwarz's Bayesian Criterion	418.7851	.	.
Loan Status	Loan Status	SSE	Sum of Squared Errors	96.59542	85.54308	.
Loan Status	Loan Status	SUMW	Sum of Case Weights Times Freq	732	496	.
Loan Status	Loan Status	MISC	Misclassification Rate	0.180328	0.241935	.

BACKWARD REGRESSION

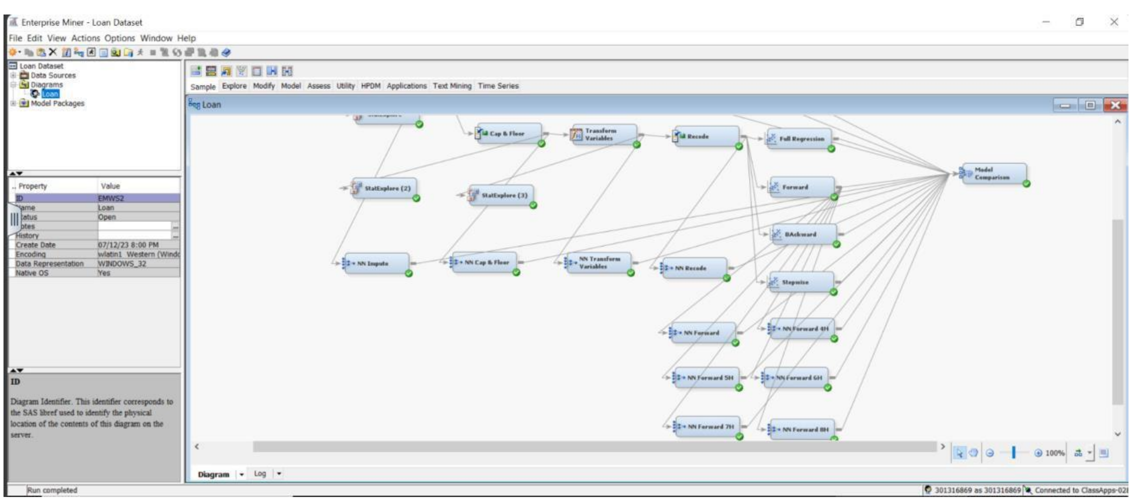
Results - Node: BAckward Diagram: Loan

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Loan Status	Loan Status	AIC	Akaike's Information Criterion	326.8356	.	.
Loan Status	Loan Status	ASE	Average Squared Error	0.136056	0.167397	.
Loan Status	Loan Status	AVERR	Average Error Function	0.432836	0.5201	.
Loan Status	Loan Status	DFE	Degrees of Freedom for Error	361	.	.
Loan Status	Loan Status	DFM	Model Degrees of Freedom	5	.	.
Loan Status	Loan Status	DFT	Total Degrees of Freedom	366	.	.
Loan Status	Loan Status	DIV	Divisor for ASE	732	496	.
Loan Status	Loan Status	ERR	Error Function	316.8356	257.9695	.
Loan Status	Loan Status	FPE	Final Prediction Error	0.139824	.	.
Loan Status	Loan Status	MAX	Maximum Absolute Error	0.95781	0.980912	.
Loan Status	Loan Status	MSE	Mean Square Error	0.13794	0.167397	.
Loan Status	Loan Status	NOBS	Sum of Frequencies	366	248	.
Loan Status	Loan Status	NW	Number of Estimate Weights	5	.	.
Loan Status	Loan Status	RASE	Root Average Sum of Squares	0.368857	0.409142	.
Loan Status	Loan Status	RFPE	Root Final Prediction Error	0.373931	.	.
Loan Status	Loan Status	RMSE	Root Mean Squared Error	0.371403	0.409142	.
Loan Status	Loan Status	SBC	Schwarz's Bayesian Criterion	346.3488	.	.
Loan Status	Loan Status	SSE	Sum of Squared Errors	99.59269	83.0291	.
Loan Status	Loan Status	SUMW	Sum of Case Weights Times Freq	732	496	.
Loan Status	Loan Status	MISC	Misclassification Rate	0.177596	0.209677	.

NEURAL NETWORKS

Neural network models were run on the various stages of data cleanup to enable us to verify the best model. The neural network model was run on the impute, Cap and floor, transform, and recode nodes with the image shown below:



NEURAL NETWORK IMPUTE (NN IMPUTE)

This was connected to the impute node and we have an ASE score of 0.179212.

Results - Node: NN Impute Diagram: Loan

File Edit View Window

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Loan Status	Loan Status	DFT	Total Degrees of Freedom	366	.	.
Loan Status	Loan Status	DFE	Degrees of Freedom for Error	281	.	.
Loan Status	Loan Status	DFM	Model Degrees of Freedom	85	.	.
Loan Status	Loan Status	NW	Number of Estimated Weights	85	.	.
Loan Status	Loan Status	AIC	Akaike's Information Criterion	503.1242	.	.
Loan Status	Loan Status	SBC	Schwarz's Bayesian Criterion	834.848	.	.
Loan Status	Loan Status	ASE	Average Squared Error	0.142603	0.179212	.
Loan Status	Loan Status	MAX	Maximum Absolute Error	0.910949	0.916121	.
Loan Status	Loan Status	DIV	Divisor for ASE	732	496	.
Loan Status	Loan Status	NOBS	Sum of Frequencies	366	248	.
Loan Status	Loan Status	RASE	Root Average Squared Error	0.377628	0.423335	.
Loan Status	Loan Status	SSE	Sum of Squared Errors	104.3856	88.88922	.
Loan Status	Loan Status	SUMW	Sum of Case Weights Times Freq	732	496	.
Loan Status	Loan Status	FPE	Final Prediction Error	0.228876	.	.
Loan Status	Loan Status	MSE	Mean Squared Error	0.185739	0.179212	.
Loan Status	Loan Status	RFPE	Root Final Prediction Error	0.47841	.	.
Loan Status	Loan Status	RMSE	Root Mean Squared Error	0.430975	0.423335	.
Loan Status	Loan Status	AVERR	Average Error Function	0.455088	0.543931	.
Loan Status	Loan Status	ERR	Error Function	333.1242	269.7896	.
Loan Status	Loan Status	MISC	Misclassification Rate	0.180328	0.225806	.
Loan Status	Loan Status	WRONG	Number of Wrong Classifications	66	56	.

NEURAL NETWORK CAP AND FLOOR (NN CAP AND FLOOR)

This was connected to the cap and floor node, and we have an ASE score of 0.181055.

Results - Node: NN Cap & Floor Diagram: Loan

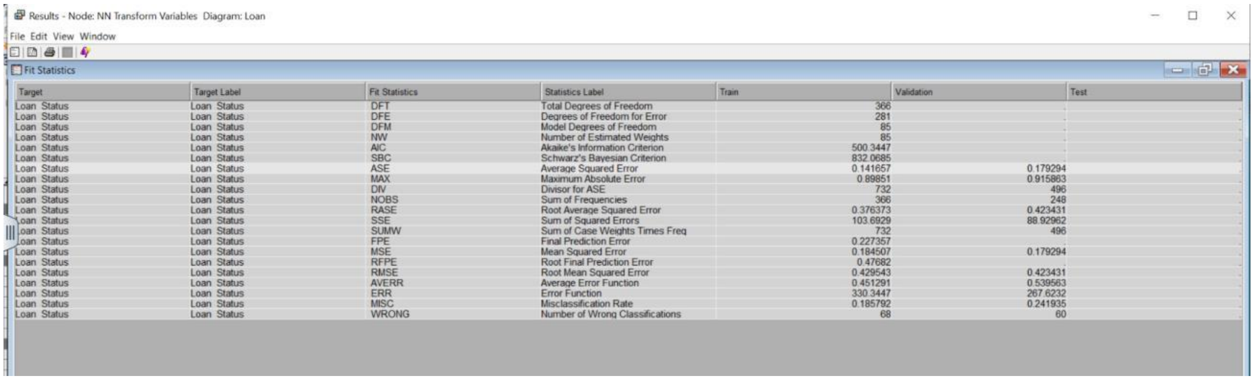
File Edit View Window

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Loan Status	Loan Status	DFT	Total Degrees of Freedom	366	.	.
Loan Status	Loan Status	DFE	Degrees of Freedom for Error	281	.	.
Loan Status	Loan Status	DFM	Model Degrees of Freedom	85	.	.
Loan Status	Loan Status	NW	Number of Estimated Weights	85	.	.
Loan Status	Loan Status	AIC	Akaike's Information Criterion	501.5501	.	.
Loan Status	Loan Status	SBC	Schwarz's Bayesian Criterion	833.2739	.	.
Loan Status	Loan Status	ASE	Average Squared Error	0.142425	0.181055	.
Loan Status	Loan Status	MAX	Maximum Absolute Error	0.909258	0.915495	.
Loan Status	Loan Status	DIV	Divisor for ASE	732	496	.
Loan Status	Loan Status	NOBS	Sum of Frequencies	366	248	.
Loan Status	Loan Status	RASE	Root Average Squared Error	0.377393	0.425506	.
Loan Status	Loan Status	SSE	Sum of Squared Errors	104.2554	89.80329	.
Loan Status	Loan Status	SUMW	Sum of Case Weights Times Freq	732	496	.
Loan Status	Loan Status	FPE	Final Prediction Error	0.22859	.	.
Loan Status	Loan Status	MSE	Mean Squared Error	0.185508	0.181055	.
Loan Status	Loan Status	RFPE	Root Final Prediction Error	0.478111	.	.
Loan Status	Loan Status	RMSE	Root Mean Squared Error	0.430706	0.425506	.
Loan Status	Loan Status	AVERR	Average Error Function	0.452937	0.547512	.
Loan Status	Loan Status	ERR	Error Function	331.5501	271.566	.
Loan Status	Loan Status	MISC	Misclassification Rate	0.185792	0.241935	.
Loan Status	Loan Status	WRONG	Number of Wrong Classifications	68	60	.

NEURAL NETWORK NN TRANSFORM

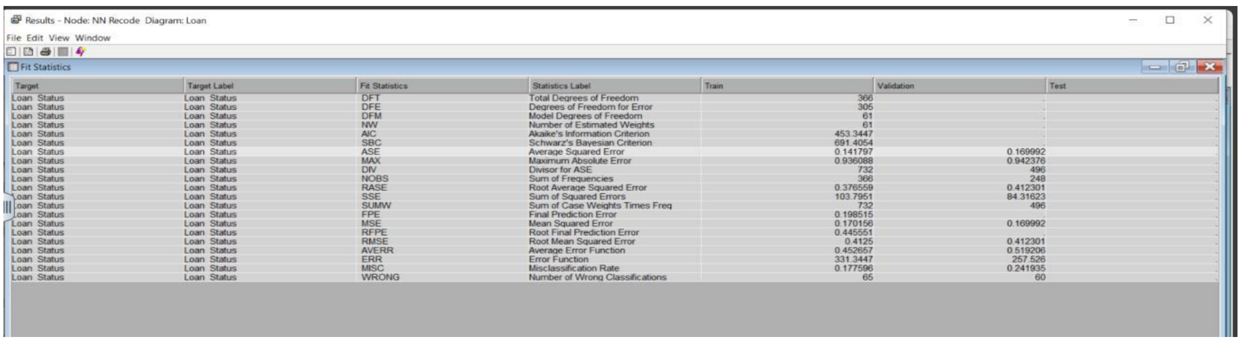
This was connected to the transform node, and we have an ASE score of 0.179294.



Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Loan Status	Loan Status	DFT	Total Degrees of Freedom	366	.	.
Loan Status	Loan Status	DFE	Degrees of Freedom for Error	281	.	.
Loan Status	Loan Status	DFM	Model Degrees of Freedom	85	.	.
Loan Status	Loan Status	NW	Number of Estimated Weights	85	.	.
Loan Status	Loan Status	AIC	Akaike's Information Criterion	500.3447	.	.
Loan Status	Loan Status	SBC	Schwarz's Bayesian Criterion	832.0685	.	.
Loan Status	Loan Status	ASE	Average Squared Error	0.141657	0.179294	.
Loan Status	Loan Status	MAX	Maximum Absolute Error	0.89851	0.915863	.
Loan Status	Loan Status	DIV	Divisor for ASE	732	496	.
Loan Status	Loan Status	NOBS	Sum of Frequencies	366	248	.
Loan Status	Loan Status	RASE	Root Average Squared Error	0.376373	0.423431	.
Loan Status	Loan Status	SSE	Sum of Squared Errors	103.6929	88.92962	.
Loan Status	Loan Status	SUMW	Sum of Case Weights Times Freq	732	496	.
Loan Status	Loan Status	FPE	Final Prediction Error	0.227357	.	.
Loan Status	Loan Status	MSE	Mean Squared Error	0.184507	0.179294	.
Loan Status	Loan Status	RFPE	Root Final Prediction Error	0.47682	.	.
Loan Status	Loan Status	RMSE	Root Mean Squared Error	0.429543	0.423431	.
Loan Status	Loan Status	AVERR	Average Error Function	0.451291	0.539563	.
Loan Status	Loan Status	ERR	Error Function	330.3447	267.6232	.
Loan Status	Loan Status	MISC	Misclassification Rate	0.185792	0.241935	.
Loan Status	Loan Status	WRONG	Number of Wrong Classifications	68	60	.

NEURAL NETWORK NN RECODE

This was connected to the transform node, and we have an ASE score of 0.169992.

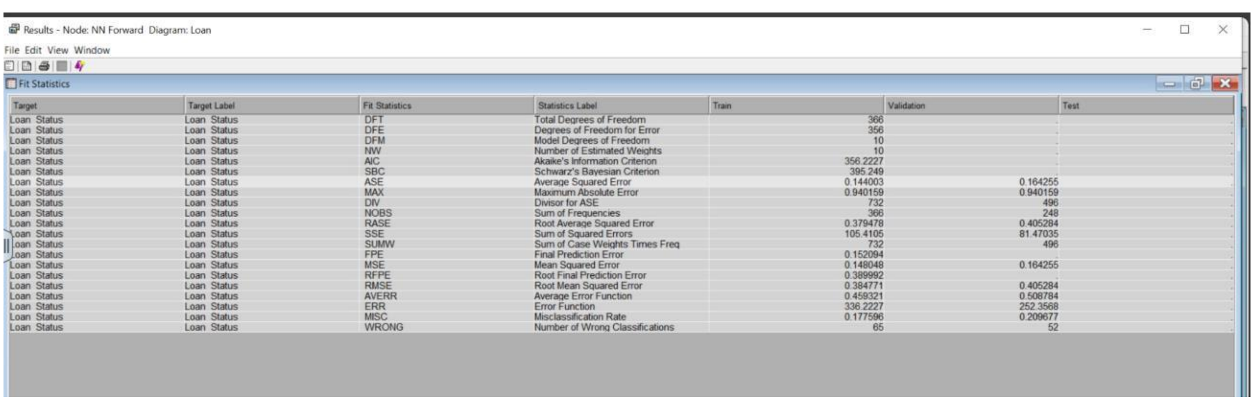


Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Loan Status	Loan Status	DFT	Total Degrees of Freedom	366	.	.
Loan Status	Loan Status	DFE	Degrees of Freedom for Error	305	.	.
Loan Status	Loan Status	DFM	Model Degrees of Freedom	61	.	.
Loan Status	Loan Status	NW	Number of Estimated Weights	61	.	.
Loan Status	Loan Status	AIC	Akaike's Information Criterion	453.3447	.	.
Loan Status	Loan Status	SBC	Schwarz's Bayesian Criterion	691.4054	.	.
Loan Status	Loan Status	ASE	Average Squared Error	0.141797	0.169992	.
Loan Status	Loan Status	MAX	Maximum Absolute Error	0.936088	0.942376	.
Loan Status	Loan Status	DIV	Divisor for ASE	732	496	.
Loan Status	Loan Status	NOBS	Sum of Frequencies	366	248	.
Loan Status	Loan Status	RASE	Root Average Squared Error	0.376559	0.412301	.
Loan Status	Loan Status	SSE	Sum of Squared Errors	103.7951	84.31623	.
Loan Status	Loan Status	SUMW	Sum of Case Weights Times Freq	732	496	.
Loan Status	Loan Status	FPE	Final Prediction Error	0.198515	.	.
Loan Status	Loan Status	MSE	Mean Squared Error	0.170156	0.169992	.
Loan Status	Loan Status	RFPE	Root Final Prediction Error	0.445551	.	.
Loan Status	Loan Status	RMSE	Root Mean Squared Error	0.4125	0.412301	.
Loan Status	Loan Status	AVERR	Average Error Function	0.452657	0.519206	.
Loan Status	Loan Status	ERR	Error Function	331.3447	257.526	.
Loan Status	Loan Status	MISC	Misclassification Rate	0.177596	0.241935	.
Loan Status	Loan Status	WRONG	Number of Wrong Classifications	65	60	.

NEURAL NETWORK FORWARD (NN FORWARD)

The neural network model was connected to the forward regression model. This gives an ASE score of 0.164255.

We do further analysis with the best model with ASE score by changing the hidden variables. This model has 3 hidden variables.



Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Loan Status	Loan Status	DFT	Total Degrees of Freedom	366	.	.
Loan Status	Loan Status	DFE	Degrees of Freedom for Error	356	.	.
Loan Status	Loan Status	DFM	Model Degrees of Freedom	10	.	.
Loan Status	Loan Status	NW	Number of Estimated Weights	10	.	.
Loan Status	Loan Status	AIC	Akaike's Information Criterion	356.2227	.	.
Loan Status	Loan Status	SBC	Schwarz's Bayesian Criterion	395.249	.	.
Loan Status	Loan Status	ASE	Average Squared Error	0.144003	0.164255	.
Loan Status	Loan Status	MAX	Maximum Absolute Error	0.940159	0.940159	.
Loan Status	Loan Status	DIV	Divisor for ASE	732	496	.
Loan Status	Loan Status	NOBS	Sum of Frequencies	366	248	.
Loan Status	Loan Status	RASE	Root Average Squared Error	0.379478	0.405284	.
Loan Status	Loan Status	SSE	Sum of Squared Errors	105.4105	81.47035	.
Loan Status	Loan Status	SUMW	Sum of Case Weights Times Freq	732	496	.
Loan Status	Loan Status	FPE	Final Prediction Error	0.152094	.	.
Loan Status	Loan Status	MSE	Mean Squared Error	0.148048	0.164255	.
Loan Status	Loan Status	RFPE	Root Final Prediction Error	0.389992	.	.
Loan Status	Loan Status	RMSE	Root Mean Squared Error	0.384771	0.405284	.
Loan Status	Loan Status	AVERR	Average Error Function	0.459321	0.508784	.
Loan Status	Loan Status	ERR	Error Function	336.2227	252.3568	.
Loan Status	Loan Status	MISC	Misclassification Rate	0.177596	0.209677	.
Loan Status	Loan Status	WRONG	Number of Wrong Classifications	65	52	.

NEURAL NETWORK NN FORWARD 2H

The neural network model was connected to the forward regression model. This model has the number of hidden variables changed to two (2). The ASE score is 0.164178.

Results - Node: NN Forward 2H Diagram: Loan

ile Edit View Window

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Loan Status	Loan Status	DFT	Total Degrees of Freedom	366	.	.
Loan Status	Loan Status	DFE	Degrees of Freedom for Error	359	.	.
Loan Status	Loan Status	DFM	Model Degrees of Freedom	7	.	.
Loan Status	Loan Status	NW	Number of Estimated Weights	7	.	.
Loan Status	Loan Status	AIC	Akaike's Information Criterion	350.1271	.	.
Loan Status	Loan Status	SBC	Schwarz's Bayesian Criterion	377.4455	.	.
Loan Status	Loan Status	ASE	Average Squared Error	0.14399	0.164178	.
Loan Status	Loan Status	MAX	Maximum Absolute Error	0.930977	0.930977	.
Loan Status	Loan Status	DIV	Divisor for ASE	732	496	.
Loan Status	Loan Status	NOBS	Sum of Frequencies	366	248	.
Loan Status	Loan Status	RASE	Root Average Squared Error	0.37946	0.405189	.
Loan Status	Loan Status	SSE	Sum of Squared Errors	105.4005	81.4325	.
Loan Status	Loan Status	SUMW	Sum of Case Weights Times Freq	732	496	.
Loan Status	Loan Status	FPE	Final Prediction Error	0.149605	.	.
Loan Status	Loan Status	MSE	Mean Squared Error	0.146797	0.164178	.
Loan Status	Loan Status	RFPE	Root Final Prediction Error	0.386788	.	.
Loan Status	Loan Status	RMSE	Root Mean Squared Error	0.383141	0.405189	.
Loan Status	Loan Status	AVERR	Average Error Function	0.45919	0.508184	.
Loan Status	Loan Status	ERR	Error Function	336.1271	252.0592	.
Loan Status	Loan Status	MISC	Misclassification Rate	0.177596	0.209677	.
Loan Status	Loan Status	WRONG	Number of Wrong Classifications	65	52	.

NEURAL NETWORK NN FORWARD 4H

The neural network model was connected to the forward regression model. This model has the number of hidden variables changed to four (4). The ASE score is 0.164367.

Results - Node: NN Forward 4H Diagram: Loan

File Edit View Window

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Loan Status	Loan Status	DFT	Total Degrees of Freedom	366	.	.
Loan Status	Loan Status	DFE	Degrees of Freedom for Error	353	.	.
Loan Status	Loan Status	DFM	Model Degrees of Freedom	13	.	.
Loan Status	Loan Status	NW	Number of Estimated Weights	13	.	.
Loan Status	Loan Status	AIC	Akaike's Information Criterion	362.0177	.	.
Loan Status	Loan Status	SBC	Schwarz's Bayesian Criterion	412.7519	.	.
Loan Status	Loan Status	ASE	Average Squared Error	0.143941	0.164367	.
Loan Status	Loan Status	MAX	Maximum Absolute Error	0.931435	0.931435	.
Loan Status	Loan Status	DIV	Divisor for ASE	732	496	.
Loan Status	Loan Status	NOBS	Sum of Frequencies	366	248	.
Loan Status	Loan Status	RASE	Root Average Squared Error	0.379395	0.405421	.
Loan Status	Loan Status	SSE	Sum of Squared Errors	105.3647	81.52583	.
Loan Status	Loan Status	SUMW	Sum of Case Weights Times Freq	732	496	.
Loan Status	Loan Status	FPE	Final Prediction Error	0.154543	.	.
Loan Status	Loan Status	MSE	Mean Squared Error	0.149242	0.164367	.
Loan Status	Loan Status	RFPE	Root Final Prediction Error	0.393119	.	.
Loan Status	Loan Status	RMSE	Root Mean Squared Error	0.386318	0.405421	.
Loan Status	Loan Status	AVERR	Average Error Function	0.459041	0.508778	.
Loan Status	Loan Status	ERR	Error Function	336.0177	252.3538	.
Loan Status	Loan Status	MISC	Misclassification Rate	0.177596	0.209677	.
Loan Status	Loan Status	WRONG	Number of Wrong Classifications	65	52	.

NEURAL NETWORK NN FORWARD 5H

The neural network model was connected to the forward regression model. This model has the number of hidden variables changed to five (5). The ASE score is 0.164537.

Results - Node: NN Forward 5H Diagram: Loan

File Edit View Window

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Loan_Status	Loan_Status	DFT	Total Degrees of Freedom	366	.	.
Loan_Status	Loan_Status	DFE	Degrees of Freedom for Error	350	.	.
Loan_Status	Loan_Status	DFM	Model Degrees of Freedom	16	.	.
Loan_Status	Loan_Status	NW	Number of Estimated Weights	16	.	.
Loan_Status	Loan_Status	AIC	Akaike's Information Criterion	368.0162	.	.
Loan_Status	Loan_Status	SBC	Schwarz's Bayesian Criterion	430.4583	.	.
Loan_Status	Loan_Status	ASE	Average Squared Error	0.143933	0.164537	.
Loan_Status	Loan_Status	MAX	Maximum Absolute Error	0.923645	0.923645	.
Loan_Status	Loan_Status	DIV	Divisor for ASE	732	496	.
Loan_Status	Loan_Status	NOBS	Sum of Frequencies	366	248	.
Loan_Status	Loan_Status	RASE	Root Average Squared Error	0.379385	0.405632	.
Loan_Status	Loan_Status	SSE	Sum of Squared Errors	105.3589	81.61045	.
Loan_Status	Loan_Status	SUMW	Sum of Case Weights Times Freq	732	496	.
Loan_Status	Loan_Status	FPE	Final Prediction Error	0.157093	.	.
Loan_Status	Loan_Status	MSE	Mean Squared Error	0.150513	0.164537	.
Loan_Status	Loan_Status	RFPE	Root Final Prediction Error	0.396349	.	.
Loan_Status	Loan_Status	RMSE	Root Mean Squared Error	0.38796	0.405632	.
Loan_Status	Loan_Status	AVERR	Average Error Function	0.459039	0.509127	.
Loan_Status	Loan_Status	ERR	Error Function	336.0162	252.5271	.
Loan_Status	Loan_Status	MISC	Misclassification Rate	0.177596	0.209677	.
Loan_Status	Loan_Status	WRONG	Number of Wrong Classifications	65	52	.

NEURAL NETWORK NN FORWARD 6H

The neural network model was connected to the forward regression model. This model has the number of hidden variables changed to six (6). The ASE score is 0.164031.

Results - Node: NN Forward 6H Diagram: Loan

File Edit View Window

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Loan_Status	Loan_Status	DFT	Total Degrees of Freedom	366	.	.
Loan_Status	Loan_Status	DFE	Degrees of Freedom for Error	347	.	.
Loan_Status	Loan_Status	DFM	Model Degrees of Freedom	19	.	.
Loan_Status	Loan_Status	NW	Number of Estimated Weights	19	.	.
Loan_Status	Loan_Status	AIC	Akaike's Information Criterion	374.4152	.	.
Loan_Status	Loan_Status	SBC	Schwarz's Bayesian Criterion	448.5652	.	.
Loan_Status	Loan_Status	ASE	Average Squared Error	0.14411	0.164031	.
Loan_Status	Loan_Status	MAX	Maximum Absolute Error	0.936048	0.936048	.
Loan_Status	Loan_Status	DIV	Divisor for ASE	732	496	.
Loan_Status	Loan_Status	NOBS	Sum of Frequencies	366	248	.
Loan_Status	Loan_Status	RASE	Root Average Squared Error	0.379618	0.405008	.
Loan_Status	Loan_Status	SSE	Sum of Squared Errors	105.4886	81.35942	.
Loan_Status	Loan_Status	SUMW	Sum of Case Weights Times Freq	732	496	.
Loan_Status	Loan_Status	FPE	Final Prediction Error	0.159892	.	.
Loan_Status	Loan_Status	MSE	Mean Squared Error	0.152001	0.164031	.
Loan_Status	Loan_Status	RFPE	Root Final Prediction Error	0.399865	.	.
Loan_Status	Loan_Status	RMSE	Root Mean Squared Error	0.389873	0.405008	.
Loan_Status	Loan_Status	AVERR	Average Error Function	0.459584	0.507918	.
Loan_Status	Loan_Status	ERR	Error Function	336.4152	251.9272	.
Loan_Status	Loan_Status	MISC	Misclassification Rate	0.177596	0.209677	.
Loan_Status	Loan_Status	WRONG	Number of Wrong Classifications	65	52	.

NEURAL NETWORK FORWARD 7H (NN FORWARD 7H)

The neural network model was connected to the forward regression model. This model has the number of hidden variables changed to seven (7). The ase score is 0.164019.

Results - Node: NN Forward 7H Diagram: Loan

File Edit View Window

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Loan_Status	Loan_Status	DFT	Total Degrees of Freedom	366	.	.
Loan_Status	Loan_Status	DFE	Degrees of Freedom for Error	344	.	.
Loan_Status	Loan_Status	DFM	Model Degrees of Freedom	22	.	.
Loan_Status	Loan_Status	NW	Number of Estimated Weights	22	.	.
Loan_Status	Loan_Status	AIC	Akaike's Information Criterion	380.4488	.	.
Loan_Status	Loan_Status	SBC	Schwarz's Bayesian Criterion	466.3068	.	.
Loan_Status	Loan_Status	ASE	Average Squared Error	0.144124	0.164019	.
Loan_Status	Loan_Status	MAX	Maximum Absolute Error	0.936348	0.936348	.
Loan_Status	Loan_Status	DIV	Divisor for ASE	732	496	.
Loan_Status	Loan_Status	NOBS	Sum of Frequencies	366	248	.
Loan_Status	Loan_Status	RASE	Root Average Squared Error	0.379637	0.404993	.
Loan_Status	Loan_Status	SSE	Sum of Squared Errors	105.4991	81.3536	.
Loan_Status	Loan_Status	SUMW	Sum of Case Weights Times Freq	732	496	.
Loan_Status	Loan_Status	FPE	Final Prediction Error	0.162559	.	.
Loan_Status	Loan_Status	MSE	Mean Squared Error	0.153342	0.164019	.
Loan_Status	Loan_Status	RFPE	Root Final Prediction Error	0.403186	.	.
Loan_Status	Loan_Status	RMSE	Root Mean Squared Error	0.391589	0.404993	.
Loan_Status	Loan_Status	AVERR	Average Error Function	0.45963	0.507895	.
Loan_Status	Loan_Status	ERR	Error Function	336.4488	251.916	.
Loan_Status	Loan_Status	MISC	Misclassification Rate	0.177596	0.209677	.
Loan_Status	Loan_Status	WRONG	Number of Wrong Classifications	65	52	.

NEURAL NETWORK FORWARD 8H (NN FORWARD 8H)

The neural network model was connected to the forward regression model. This model has the number of hidden variables changed to eight (8). The score is 0.164035.

Results - Node: NN Forward 8H Diagram: Loan

File Edit View Window

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Loan Status	Loan Status	DFT	Total Degrees of Freedom	366	.	.
Loan Status	Loan Status	DFE	Degrees of Freedom for Error	341	.	.
Loan Status	Loan Status	DFM	Model Degrees of Freedom	25	.	.
Loan Status	Loan Status	NW	Number of Estimated Weights	25	.	.
Loan Status	Loan Status	AIC	Akaike's Information Criterion	386.4176	.	.
Loan Status	Loan Status	SBC	Schwarz's Bayesian Criterion	483.9834	.	.
Loan Status	Loan Status	ASE	Average Squared Error	0.14411	0.164035	.
Loan Status	Loan Status	MAX	Maximum Absolute Error	0.936427	0.936427	.
Loan Status	Loan Status	DIV	Divisor for ASE	732	496	.
Loan Status	Loan Status	NOBS	Sum of Frequencies	366	248	.
Loan Status	Loan Status	RASE	Root Average Squared Error	0.379618	0.405012	.
Loan Status	Loan Status	SSE	Sum of Squared Errors	105.4884	81.36142	.
Loan Status	Loan Status	SUMW	Sum of Case Weights Times Freq	732	496	.
Loan Status	Loan Status	FPE	Final Prediction Error	0.16524	.	.
Loan Status	Loan Status	MSE	Mean Squared Error	0.154675	0.164035	.
Loan Status	Loan Status	RFPE	Root Final Prediction Error	0.406498	.	.
Loan Status	Loan Status	RMSE	Root Mean Squared Error	0.393288	0.405012	.
Loan Status	Loan Status	AVERR	Average Error Function	0.459587	0.507946	.
Loan Status	Loan Status	ERR	Error Function	336.4176	251.941	.
Loan Status	Loan Status	MISC	Misclassification Rate	0.177596	0.209677	.
Loan Status	Loan Status	WRONG	Number of Wrong Classifications	65	52	.

ODDS RATIO

Using Chi-Sq **IMP_Credit_History** and **Property_Area** recorded the highest Chi-Sq value making them highly statistically significant to our Backward regression model.

- People with **no credit history** are 98.8% less likely to get credit loans when compared to people with **credit history**.
- Customers in **Rural areas** are 47% less likely to get credit loans when compared to people in **Urban area**.
- Customers in **semiurban areas** are 86.6% more likely to get credit loans when compared to people in **Urban areas**.

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)
Intercept		1	1.2342	1.9911	0.38	0.5354		3.436
Education	Graduate	1	0.2429	0.1760	1.91	0.1675		1.275
IMP_Credit_History	0	1	-2.2314	0.2985	55.89	<.0001		0.107
IMP_Gender	Female	1	0.0862	0.2084	0.17	0.6793		1.090
IMP_Self_Employed	No	1	-0.2397	0.2347	1.04	0.3070		0.787
LOG_REP_CoapplicantIncome		1	0.0757	0.0406	3.47	0.0627	0.1594	1.079
LOG_REP_IMP_Dependents		1	0.3477	0.3139	1.23	0.2681	0.0997	1.416
LOG_REP_IMP_LoanAmount		1	-0.4049	0.3254	1.55	0.2133	-0.1052	0.667
M_Credit_History	0	1	0.2636	0.2360	1.25	0.2640		1.302
M_Dependents	0	1	-0.3462	0.6057	0.33	0.5675		0.707
M_Gender	0	1	0.0885	0.5313	0.03	0.8677		1.093
M_LoanAmount	0	1	-0.2674	0.5044	0.28	0.5960		0.765
M_Loan_Amount_Term	0	1	-0.4091	0.5247	0.61	0.4355		0.664
M_Self_Employed	0	1	0.0852	0.3142	0.07	0.7862		1.089
Married	No	1	-0.3233	0.1756	3.39	0.0657		0.724
Property_Area	Rural	1	-0.6308	0.2068	9.31	0.0023		0.532
Property_Area	Semiurban	1	0.6274	0.2175	8.32	0.0039		1.873
REP_IMP_Loan_Amount_Term	360	1	0.1554	0.2153	0.52	0.4704		1.168

Odds Ratio Estimates

Effect		Point Estimate
Education	Graduate vs Not Graduate	1.625
IMP_Credit_History	0 vs 1	0.012
IMP_Gender	Female vs Male	1.188

Odds Ratio Estimates

Effect		Point Estimate
Education	Graduate vs Not Graduate	1.625
IMP_Credit_History	0 vs 1	0.012
IMP_Gender	Female vs Male	1.188
IMP_Self_Employed	No vs Yes	0.619
LOG_REP_CoapplicantIncome		1.079
LOG_REP_IMP_Dependents		1.416
LOG_REP_IMP_LoanAmount		0.667
M_Credit_History	0 vs 1	1.694
M_Dependents	0 vs 1	0.500
M_Gender	0 vs 1	1.194
M_LoanAmount	0 vs 1	0.586
M_Loan_Amount_Term	0 vs 1	0.441
M_Self_Employed	0 vs 1	1.186
Married	No vs Yes	0.524
Property_Area	Rural vs Urban	0.530
Property_Area	Semiurban vs Urban	1.866
REP_IMP_Loan_Amount_Term	360 vs 999	1.365

MODEL COMPARISON

Results - Node: Model Comparison Diagram: Loan

File Edit View Window

Fit Statistics

Selected Model	Predecessor Node	Model Node	Target Variable	Target Label	Model Description	Selection Criterion: Valid: Roc Index ▼	Valid: Average Squared Error	Train: Sum of Frequencies	Train: Misclassification Rate	Train: Maximum Absolute Error	Train: Sum of Squared Errors
Y	Neural5	Neural5	Loan Sta...	Loan Sta...	NN Recode	0.73	0.169992	366	0.177596	0.936088	103.7951
	Neural2	Neural2	Loan Sta...	Loan Sta...	NN Transform Variables	0.727	0.179294	366	0.185792	0.89851	103.6929
	Reg2	Reg2	Loan Sta...	Loan Sta...	BAckward	0.723	0.167397	366	0.177596	0.95781	99.59269
	Reg4	Reg4	Loan Sta...	Loan Sta...	Full Regression	0.71	0.172466	366	0.180328	0.966576	96.59542
	Tree	Tree	Loan Sta...	Loan Sta...	Max Tree	0.704	0.18532	366	0.166667	0.96	95.33516
	Neural	Neural	Loan Sta...	Loan Sta...	NN Cap & Floor	0.7	0.181055	366	0.185792	0.909258	104.2554
	Neural3	Neural3	Loan Sta...	Loan Sta...	NN Impute	0.699	0.179212	366	0.180328	0.910949	104.3856
	Neural9	Neural9	Loan Sta...	Loan Sta...	NN Forward 7H	0.677	0.164019	366	0.177596	0.936348	105.4991
	Neural10	Neural10	Loan Sta...	Loan Sta...	NN Forward 6H	0.677	0.164031	366	0.177596	0.936048	105.4886
	Neural8	Neural8	Loan Sta...	Loan Sta...	NN Forward 8H	0.677	0.164035	366	0.177596	0.936427	105.4884
	Neural4	Neural4	Loan Sta...	Loan Sta...	NN Forward	0.677	0.164255	366	0.177596	0.940159	105.4105
	Neural7	Neural7	Loan Sta...	Loan Sta...	NN Forward 4H	0.677	0.164367	366	0.177596	0.931435	105.3647
	Neural6	Neural6	Loan Sta...	Loan Sta...	NN Forward 5H	0.677	0.164537	366	0.177596	0.923645	105.3589
	Reg	Reg	Loan Sta...	Loan Sta...	Forward	0.677	0.164548	366	0.177596	0.929824	105.3545
	Reg3	Reg3	Loan Sta...	Loan Sta...	Stepwise	0.677	0.164548	366	0.177596	0.929824	105.3545
	Tree2	Tree2	Loan Sta...	Loan Sta...	ASE Tree	0.677	0.164548	366	0.177596	0.929825	105.3545
	Tree3	Tree3	Loan Sta...	Loan Sta...	Misclassification Tree	0.677	0.164548	366	0.177596	0.929825	105.3545

Model Node	Train: Average Squared Error	Train: Root Average Squared Error	Train: Divisor for ASE	Train: Total Degrees of Freedom	Valid: Sum of Frequencies	Valid: Misclassification Rate	Valid: Maximum Absolute Error	Valid: Sum of Squared Errors	Valid: Root Average Squared Error	Valid: Divisor for VASE	Train: Degrees of Freedom for Error	Train: Model Degrees of Freedom
Neural5	0.141797	0.376559	732	366	248	0.241935	0.942376	84.31623	0.412301	496	305	6
Neural2	0.141657	0.376373	732	366	248	0.241935	0.915863	88.92962	0.423431	496	281	8
Reg2	0.136056	0.368857	732	366	248	0.209677	0.980912	83.0291	0.409142	496	361	
Reg4	0.131961	0.363264	732	366	248	0.241935	0.988345	85.54308	0.41529	496	347	1
Tree	0.130239	0.360887	732	366	248	0.241935	0.96	91.91865	0.430488	496	.	
Neural	0.142425	0.377393	732	366	248	0.241935	0.915495	89.80329	0.425506	496	281	8
Neural3	0.142603	0.377628	732	366	248	0.225806	0.916121	88.88922	0.423335	496	281	8
Neural9	0.144124	0.379637	732	366	248	0.209677	0.936348	81.3536	0.404993	496	344	2
Neural10	0.14411	0.379618	732	366	248	0.209677	0.936048	81.35942	0.405008	496	347	1
Neural8	0.14411	0.379618	732	366	248	0.209677	0.936427	81.36142	0.405012	496	341	2
Neural4	0.144003	0.379478	732	366	248	0.209677	0.940159	81.47035	0.405284	496	356	1
Neural7	0.143941	0.379395	732	366	248	0.209677	0.931435	81.52583	0.405421	496	353	1
Neural6	0.143933	0.379385	732	366	248	0.209677	0.923645	81.61045	0.405632	496	350	1
Reg	0.143927	0.379377	732	366	248	0.209677	0.929824	81.61603	0.405646	496	364	
Reg3	0.143927	0.379377	732	366	248	0.209677	0.929824	81.61603	0.405646	496	364	
Tree2	0.143927	0.379377	732	366	248	0.209677	0.929825	81.61603	0.405646	496	.	
Tree3	0.143927	0.379377	732	366	248	0.209677	0.929825	81.61603	0.405646	496	.	

The best model to predict payment is the neural network recode with the best ROC score of 0.73.

CONCLUSION

- We performed exploratory data analysis on the dataset's features.
- The best **model is Neural5 (NN Recode)** with an **ROC index of 0.73**.
- Due to government regulation, we cannot use the best models (neural networks) because we cannot deduce or explain why the loan will be rejected. **Backward regression** is the

third-best model with an **ROC index of 0.723** will be used as this can be explained**.**

RECOMMENDATIONS BASED ON THE ODDS RATION AND CHI-SQ

Odds Ratio Estimates

Effect		Point Estimate
Education	Graduate vs Not Graduate	1.625
IMP_Credit_History	0 vs 1	0.012
IMP_Gender	Female vs Male	1.188
IMP_Self_Employed	No vs Yes	0.619
LOG_REP_CoapplicantIncome		1.079
LOG_REP_IMP_Dependents		1.416
LOG_REP_IMP_LoanAmount		0.667
M_Credit_History	0 vs 1	1.694
M_Dependents	0 vs 1	0.500
M_Gender	0 vs 1	1.194
M_LoanAmount	0 vs 1	0.586
M_Loan_Amount_Term	0 vs 1	0.441
M_Self_Employed	0 vs 1	1.186
Married	No vs Yes	0.524
Property_Area	Rural vs Urban	0.530
Property_Area	Semiurban vs Urban	1.866
REP_IMP_Loan_Amount_Term	360 vs 999	1.365

- Using Chi-Sq **IMP_Credit_History** and **Property_Area** recorded the highest Chi-Sq value making them highly statistically significant to our Backward regression model.
- People with **no credit history** are 98.8% less likely to get credit loans when compared to people with **credit history**.
- Customers in **Rural areas** is 47% less likely to get credit loans when compared to people in **Urban area**.
- Customers in semiurban areas are 86.6% more likely to get credit loans when compared to people in Urban areas.

Customers without a credit history and those who reside in rural areas should also be considered by the business in case they qualify for a credit loan.

This would facilitate the expansion of credit availability for qualified clients and help boost business interest in credit loans.

Also, we should try to check other factors that can lead to a lack of credit history especially in a country with so many immigrants. We can use other factors like work experience and salary to expand the pool of customers who can receive loans.