IOCRF 2025 Conference

The 1st International Online Conference on Risk and Financial Management

17–18 June 2025 | Online

PREDICTING EDUCATION LOAN REPAYMENT: A SEM-ANN INTEGRATIVE MODELLING APPROACH

Rakshith Bhandary

Manipal School of Commerce and Economics, Manipal Academy of Higher Education, Manipal, Karnataka, India - 575001

INTRODUCTION & AIM

Education loan complements human capital development, and with successful recovery it becomes self-sustainable. Recovery can be enhanced if default can be predicted accurately that also optimizes the capital reserve requirement. In addition, Education loan default prediction helps to filter applicants, enhance recovery and optimize provisioning for losses. **Mortgage** and **income contingent** education loans are the two types of loan schemes prevalent across the globe. Mortgage type loans are usually associated with high initial repayment burdens (Chapman & Liu, 2013; Usher, 2005). Hence, Many countries have partially adopted the income driven loans: US from 1994, Thailand from 2006, Korea from 2009, Brazil from 2016, Netherlands from 2016, Japan from 2017, Canada from 2017 and Colombia from 2022 (Chapman, 2022).

Age (AG) of the borrower influences loan repayment negatively (Ofori et al., 2014; Agarwal et al., 2011; Acquah & Addo, 2011; Oladeebo & Oladeebo, 2008; Bandyopadhyay, 2016; Steiner & Teszler, 2005).

Income (IN) positively affects loan repayment performance (Kohansal & Mansoori, 2009; Flint, 1997; Lochner & Monge-Naranjo, 2016; Bii et al., 2023; Boonchai, 2022).

Loan interest rates negatively influence repayment performance (Kohansal & Mansoori, 2009). On the contrary, **Loan instalments (EM)** positively influenced repayment behaviour (Bandyopadhyay, 2016).

Size of the family negatively influences loan repayment performance (Haque et al., 2011; Peter & Peter, 2006; Ojiako & Ogbukwa, 2012; Bandyopadhyay, 2016). Hence, we can say that number of **dependents (DE)** negatively influence the loan repayment process.

In the study by Idres et al. (2018) the recipients had the intention to repay their loans as they were aware of their responsibilities especially those recipients who had high **Integrity (IT).** The office of the education loans in Thailand relies heavily on students' **Integrity** for loan repayments (Savatsomboon, 2004).

Credit information sharing and reporting (CR) can discipline borrowers into exerting high effort in projects (Padilla & Pagano, 2000; Vercammen, 1995) and repaying loans (Klein, 1992).

Delay of gratification (DG) predicts debt behaviour (Livingstone & Lunt, 1992; O'Guinn & Faber, 1989; Norvilitis et al., 2006). self-reported delay of gratification was related to more positive financial attitudes, repayment intention and lower levels of debt (Norvilitis & Mendes-Da-Silva, 2013; Strayhorn Jr, 2002).

Financial efficacy (FE) was significantly related to tolerant attitudes towards debt (Norvilitis et al., 2006). Students who prioritize financial management seriously and practice financial discipline, display better financial well-being and intend to repay their debt regularly (Chan et al., 2012).

Education loans provide better access to education, better career choices, and better lifelong decisions (Lamkin, 2004; Millett, 2003). Therefore, **Debt Utility (DU)** influences repayment.

Hence, this study aims to evaluate the attitudinal factors of educational loan repayment by integrating willingness and ability of the borrower.

METHOD

Theory of planned behaviour and **human capital theory** is the basis for postulating the hypotheses and constructing the conceptual model.

The **willingness variables (DG, FE, DU, CR and IT)** are tested using **PLS-SEM** and the significant variables are **integrated** with **ability variables (AG, IN, EM, DE)** of the **ANN** and **DNN** model.

This study follows a multi-method approach to test the antecedent attitudinal variables of education loan repayment intention. Literature search finds variables to frame hypothesis that is tested quantitatively with partial least square-structural equation modelling (PLS-SEM) and prediction accuracy is calculated using artificial neural network (ANN) and deep neural network (DNN) modelling in multiple stages.

Proportional stratified random sampling was employed for the study with **Survey method** using electronic medium with **7-point Likert scale questions.** Content validity, pre pilot study using 30 and pilot study using 100 samples were performed before proceeding for the final data collection of **406** education loan borrowers in India. **Data cleaning** was performed as per the methods prescribed by Osborne and Overbay (2008). This study preferred PLS-SEM over CB-SEM because it was exploratory in nature and prediction was the main objective. R studio programming language was used to run PLS-SEM.

The ANN and DNN model are tested for hypotheses and model fit before integration using age, income , instalment, and dependants as input variables and default as output variable.The default status is coded as 1 for presence of default and 0 for absence of default. Here, default is the standard definition of NPA given by RBI. If the borrower does not pay the EMI for more than 3 months, it is considered as a default.

The **most significant variables** of PLS-SEM model are integrated with ANN and DNN model in the later stages to calculate prediction accuracy.

`The thumb rule for ANN sampling is **fifty responses** for **each input** variable. Hence 406 samples for 7 variables is sufficient for the task.

Train test split ratio was 80:20. **Feature scaling** and **cross validation** was performed. Scaling was done to prevent any feature from dominating the model and cross validation using 10 folds was performed to avoid overfitting. The ANN model was coded and executed using R programming language.

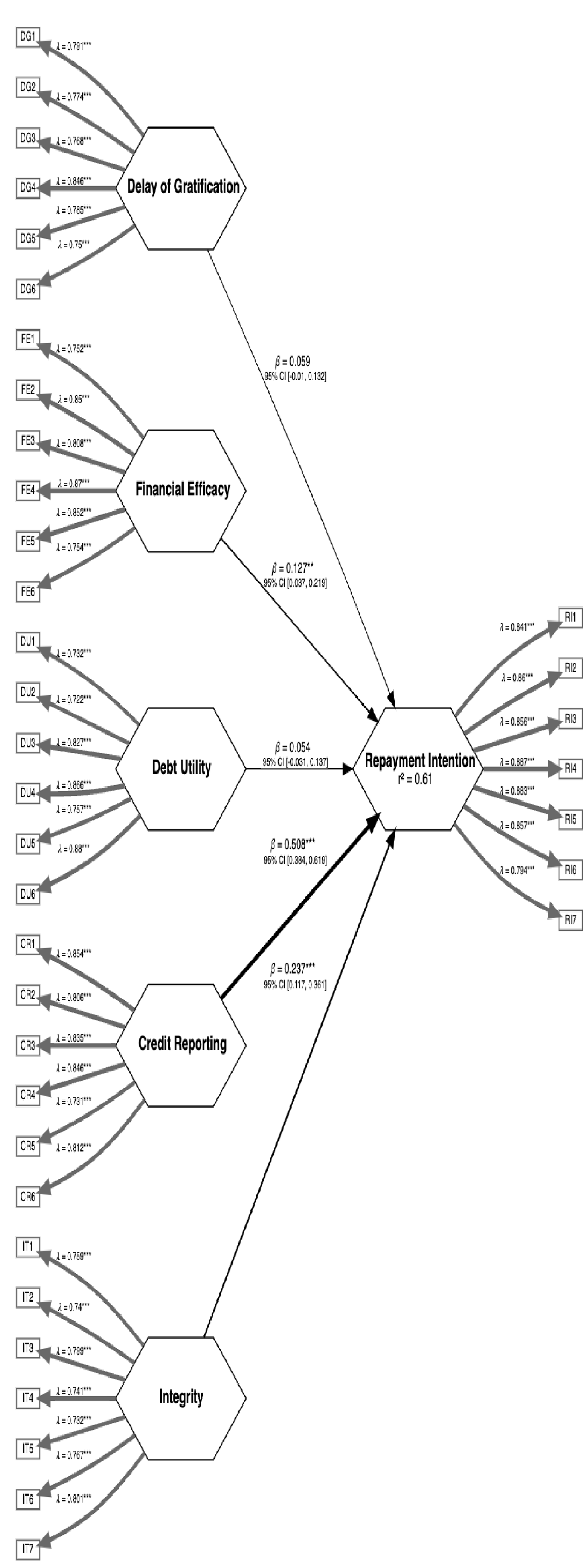

Fig 1: PLS-SEM model

RESULTS & DISCUSSION

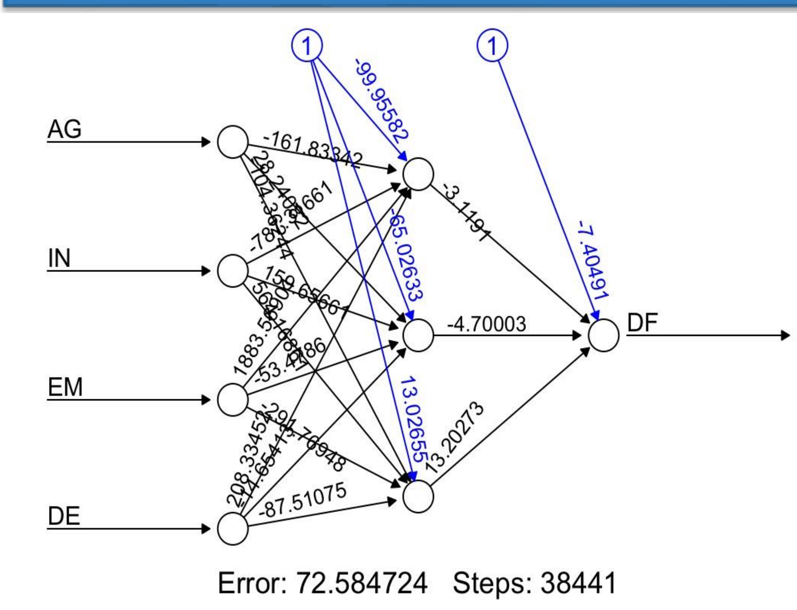

Fig 2: ANN before integration

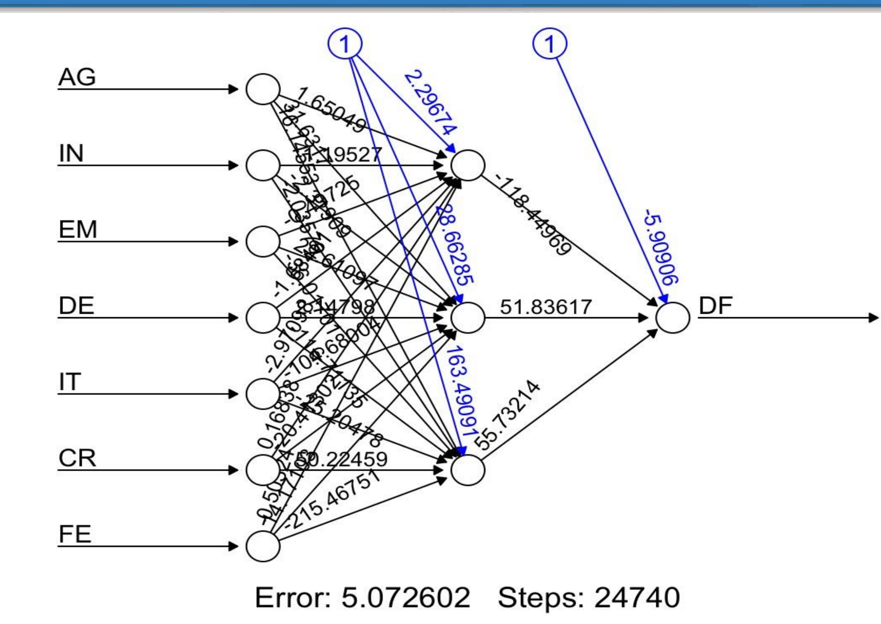

Fig 3: ANN after SEM integration

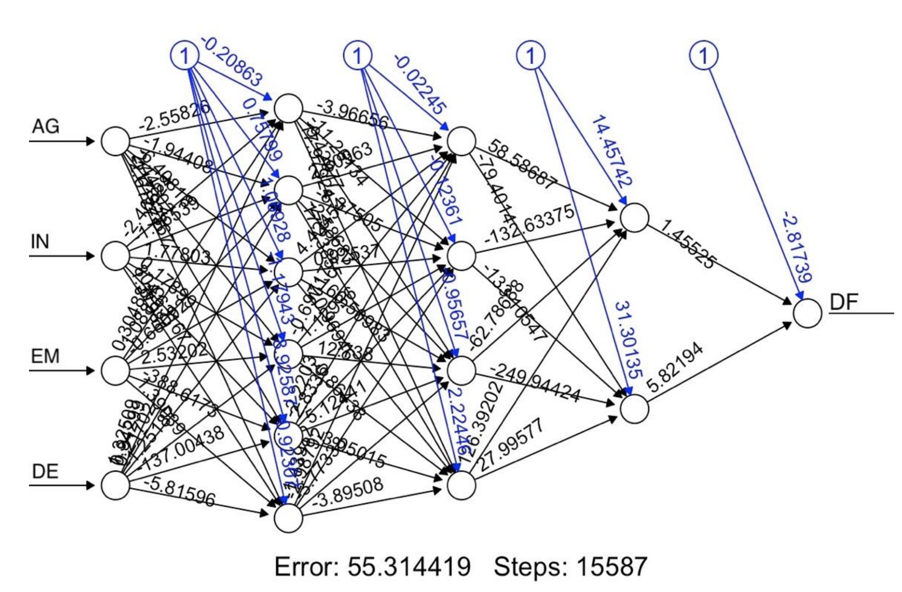

Fig 4: DNN before integration

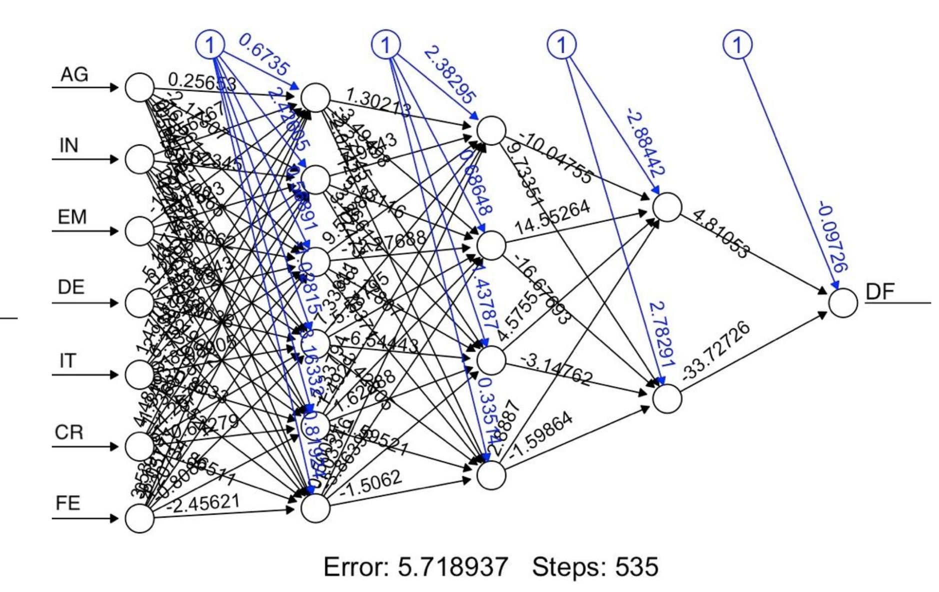

Fig 5: DNN after SEM integration

PLS-SEM Model

The outer loadings for each item is above 0.7, alpha and composite reliability for each construct is above 0.7, and the average variance for each construct is above the threshold values of 0.5

The diagonal values is greater than row values to the left and column values below to meet the Fornell Larcker criteria of discriminant validity. The HTMT values are much below the threshold of 0.8. The R^2 value of **Repayment intention** construct is **0.61.** Collinearity was not an issue in our study since the **variance inflation factors (VIF)** values of all items were **lower than 3**. Bootstrapping using **5000 resamples** was done to assess the significance level of the path coefficients**.** Path value coefficients range from -1 to +1. **IT (β =0.237***), CR (β =0.508***), and FE (β =0.127**)** are the **three most significant variables** of PLS-SEM model integrated with the ANN and DNN model. The findings of this study are aligned with prior studies in the literature.

ANN before integration			ANN after integration			DNN before integration			DNN before integration		
Actual	Prediction		Actual	Prediction		Actual	Prediction		Actual	Prediction	
	0	1		0	1		0	1		0	1
0	39	3	0	40	2	0	42	0	0	42	0
1	6	23	1	4	25	1	7	22	1	4	25
Accuracy	87.32 %		Accuracy	91.55 %		Accuracy	90.14 %		Accuracy	94.37%	
95% CI	(0.773, 0.9404)		95% CI	(0.8251, 0.9684)		95% CI	(0.8074,0.9594)		95% CI	(0.862,0.9844)	
P-Value	0.00000609		P-Value	0.00000001		P-Value	0.00000000		P-Value	0.00000000	
Kappa	0.7334		Kappa	0.8232		Kappa	0.7881		Kappa	0.8809	
Sensitivity	86.67 %		Sensitivity	90.91 %		Sensitivity	85.71 %		Sensitivity	91.30 %	
Specificity	88.46 %		Specificity	92.59 %		Specificity	100 %		Specificity	100 %	
Balanced Accuracy	87.56 %		Balanced Accuracy	91.75 %		Balanced Accuracy	92.86 %		Balanced Accuracy	95.65 %	

Table 1: **Confusion matrix of the ANN & DNN Model**

CONCLUSION

The prediction accuracy of **ANN** model **before SEM integration** was **87.32%** and **after SEM-ANN integration** it increased to **91.55%**, and from **90.14%** to **94.37% after deep neural network (DNN) integration**. Therefore, multi-stage SEM-ANN-DNN integration improves the default prediction accuracy. Improvement in prediction accuracy helps financial institutions to plan their loan recovery and calculate the optimum capital reserve requirement for provisioning towards non-performing assets.

Most of the methods employed to sanction credit and to predict credit defaults prioritize the **borrower's ability to repay (Repayment Capacity)** with **little or no emphasis** on the **willingness to repay (Behavioural Aspects)**. This study highlights the importance of **integrating both** for credit decisions with a proper methodology to **quantify** it, rather than going by the **intuition** of the lender.

The study is aligned to the goals of **SDG 4** to **increase the higher education accessibility, SDG 8** to **improve resource efficiency** in **consumption** and **production,** and **SDG 12** to make the **education loan product self-sustainable.**

FUTURE WORK / REFERENCES

The study provides a new pathway to conduct further research in exploring the behavioural variables for credit default related studies. **Behavioural aspects** can be **explored for different retail loan products** in future research, and it can be **integrated** with ability to evaluate and appraise the customer's loan application. Cost effectiveness of the models can be studied further by doing a **cost benefit analysis** for **traditional v/s modern methods.**