Credit Default Prediction

By

Merit Choorappadil Thomas

Submitted to

**The University of Roehampton**

In partial fulfilment of the requirements

for the degree of

**MASTER OF SCIENCE IN DATA SCIENCE**

Table of Contents

[1. Introduction 7](#_Toc927214740)

[Research Question or Problem statement 8](#_Toc522749298)

[Aims 8](#_Toc1298644285)

[Objectives 8](#_Toc198637622)

[Legal, Social, Ethical and Professional Considerations 8](#_Toc960797214)

[Background 8](#_Toc273743648)

[Report overview 8](#_Toc210207254)

[2. Literature or Technology Review 8](#_Toc1335231035)

[3. Methodology 9](#_Toc1844834773)

[4. Implementation 10](#_Toc1739317238)

[5. Results 11](#_Toc826321870)

[Evaluation 11](#_Toc1085198293)

[Related Work 12](#_Toc55229855)

[6. Conclusion 12](#_Toc253686570)

[Reflection 13](#_Toc1734103406)

[Future Work 13](#_Toc542248490)

[7. References 13](#_Toc1743583894)

[8. Appendices 14](#_Toc649705763)

**1.INTRODUCTION**

Financial institutions, often known as banking institutions are corporate bodies that offer services as middlemen for various financial and monetary transactions. The main function of financial institutions like a bank is to accept deposits and lend loans. Nowadays banks change their service strategies of lending loans by way of providing credit cards to the customer. Hence it can be considered one of the main income or beneficiary sources of these institutions. The banks must be considered many factors like credit score, and spending history before lending credit cards to their customers to reduce their risk of default. Banks have huge databases and proper understanding and analysis of these database results to improve performance and make progressive business decisions. The financial stability and repayment capacity of each customer varies depending on various socio-economic factors. Therefore, customers get noticeable and differential treatment based on their repayment and purchasing behaviour [40]. Banks need to know whether customers are good or bad by scoring their credit and behaviour. The phrase was defined by Anderson [41] by splitting it into two parts: the first is "credit," which refers to purchasing something and paying for it later, and the second is "scoring," which is similar to the process used for credit cards. Application credit scoring, where a score is used to inform a decision on a new credit application, and behavioural scoring, which is used to address existing clients after being granted a loan, are the two main categories of credit scoring [42]. A rise in using credit cards increases the chances of missing payments because the same individual has many credit cards from different banks leading to an increase in the probability of missing payments. As a result, financial institutions are under pressure to continuously improve the system which will detect early missing payment predictions in order to mitigate substantial financial losses associated with these card transaction activities. A perfect deep-learning model needs to predict these missing payments in future.

**1.1 AIM**

This paper's primary goal is to assist bank management in evaluating credit card consumers using machine learning by modelling and forecasting consumer behaviour with regard to two variables. By using the single or consecutive missed payment of credit card customers. The proposed improved model is anticipated to have the capability to effectively identify the credit default of customers to the next month by capturing meaningful patterns within consumer behaviour. The efficiency of predicting credit default is expected to be improved by leveraging the strengths of various Machine Learning techniques in combination. Preliminary investigations will be conducted to develop a credit default prediction model that solely relies on attention and transformer architecture, without employing recurrent networks for sequence processing.

**1.2 OBJECTIVES**

The project's initial development would be to incorporate feature selection and dimension reduction algorithms, for optimizing the process of learning classifiers. The problem with imbalanced data sets could be solved with the use of techniques like SMOTE. As a next step, LSTM recurrent neural networks will be employed as a dynamic pattern recognition classifier which would construct a context of consumer spending behaviour. LSTM feature extraction is internally taken place and is hidden from the observer. It is used for the estimation of customer behaviour scores [42].

Fine-tune the model to optimize its ability to predict customers' credit card payment behaviour and utilization rates. Evaluate the model's ability to accurately identify high-risk accounts by comparing its predictions against historical default data or external credit risk indicators. Evaluate the trained model using appropriate evaluation metrics such as accuracy, precision, recall, and F1 score. The developed LSTM model is compared with three classical machine learning algorithms: Support vector machine, Decision tree and logistic regression

**2.LITERATURE REVIEW**

Credit default prediction is a crucial task for the lender institution. It is in the case of particular financial institutions or banks for an early understanding of bankruptcy relies on that particular institution itself.[1]. Hence credit default prediction is an extensive research area for decades [2]. Researchers have analysed and evaluated the performance of well-known classification algorithms (E.g.: Logistic regression, Decision tree, neural network) and in addition to that they have also worked with kernel-based classification algorithms also (Support vector machine, least square SVM). The performance of ls-SVM and neural networks is compactly good as compared to logistic regression and other linear models they perform well in credit scoring [3]. Studies show that neural networks, Support vector machines and random forests are the artificial intelligence techniques appropriate to replace conventional credit evaluation models [4][5][6]. Popular mathematical techniques for behavioural score analysis was reviewed in [7]. To name a few, there are decision trees, logistic and linear regression. The author also examined the relevance of including economic circumstances in scoring systems. The conversion of consumer default probabilities into estimates of the profit or loss of a consumer organisation is the other issue examined in that study.

The later study of this same author states how Markov chain stochastic processes may be utilised to simulate the changes in customers' scores for their behaviour and delinquent status over time. It goes through how to utilise segmentation, mover-stayer models, and second-and third-order models to make these models fit better. The use of dynamic modelling in analysing the risk associated with consumer credit is covered in this article. It examines the methods and goals of profit scoring, behavioural scoring, and customer scoring [8].[9] Louzada F gives a thorough literature review linking theory and application of binary classification approaches for financial analysis of credit scores. The overall findings highlight the use and significance of the key credit rating procedures as well as significant historical shifts in the scientific perspective. Usually, we are evaluating a fixed time period to calculate a particular behaviour or pattern of the transaction that we get from the transaction history of an individual and values are selected arbitrarily and causing the forecast to be unstable.[10] the study analyses data from the Irish market spanning seven years to examine the differing effects of changing the performance period and outcome period. The findings of this study show that when compared to other historical payment periods of various durations, a 12-month performance period gives an easier prediction task. Findings demonstrate that when the result window is increased beyond six months, a logistic regression classifier's performance drastically declines.

**2.1 Credit cards holder’s behavioural scoring**

From the above details we can understand that most of the behaviour scoring is done with customers’ loan behaviour and behaviour scoring using credit cards is very less investigated. [11] In this study, a method for deriving symbolic rules from neural networks is presented. We provide an example of how the method effectively derived rules from a data set obtained from an examination of the service industries in the United Kingdom. [12] Examines the methodological facets of neural network studies in commercial applications. They stress the necessity for a more robust statistical comparison of neural networks and other conventional methods. They specifically provide a number of metrics for gauging a categorization technique's effectiveness. They use a comparison of neural networks and discriminant analysis to demonstrate their points. Hence the studies based on the dataset are taken place. credit scoring and behavioural scoring have grown to be crucial credit risk management duties. [13] The suggested study aims to investigate the performance of behavioural scoring utilising three widely discussed data mining techniques: support vector machine (SVM), backpropagation neural networks (BPN), and linear discriminant analysis (LDA).

Behavioural scoring tasks are carried out on one bank credit card dataset in Taiwan to show the efficacy of behavioural scoring utilising the aforementioned methods. The findings show that BPN performs better than other strategies in terms of overall scoring accuracy and is thus a useful substitute when applying behavioural scoring tasks. [14] In the study, the issue of timing the time to default in credit behavioural scoring is discussed. This offers up the prospect of including a dynamic component into behavioural scoring models, which would enable making judgements about limits, collection and recovery techniques, retention and attrition, as well as giving information about the profitability, price, or term structure of the loan. In this research, we compare the modelling and output of neural networks with survival analysis. The neural network architecture is created so that the output is similar to the output of the survival analysis.[15] In a thesis, the author uses different machine learning algorithms to find out the robustness among all those in the default prediction of credit cards.[16] Models for the duration of debt default may be created using survival analysis. In this article, we describe how survival analysis was used to simulate credit card account default using a sizable data set. We investigate the claim that the probability of default (PD) is influenced over time by overall economic conditions. It is difficult to easily include these macroeconomic variables (MVs) in logistic regression models. However, the inclusion of them as time-varying factors can be supported by survival analysis. The study takes into account many MVs, including the interest rate and unemployment rate. We demonstrate that adding these indicators to the model enhances model fit and has an impact on PD, which results in a slight improvement in default predictions on an independent test set.[17]A survey by Tony Bellotti and Crook provides discrete-time survival models of borrower default for credit cards that take into account macroeconomic variables over the course of the card's lifespan as well as information on cardholder conduct. We discover that when applied to an out-of-sample data set, dynamic models that incorporate these behavioural and macroeconomic factors significantly enhance model fit, which results in improved default forecasts at both the account and portfolio levels. We demonstrate the application of these models to stress test credit card portfolios under scenarios of severe economic situations. Dynamic data on alterations in the behaviour of a customer's other accounts with the same bank aids in the long-term prediction of the behaviour of the credit card account.

According to these findings, relationship banking has a lot to offer banks in terms of prospective advantages. For instance, the lender's access to information may be utilised to reduce the credit risk associated with a credit card account [18]. Wang L studies have Mail-in questionnaires employed in the study, and they were distributed to credit card customers who were currently using or had previously used either revolving credit or small payment plans. Regression functions revealed that in comparison to attitude and personality factors, demographic and credit card characteristics had less explanatory power. Particularly, we discovered that views about credit cards, money, and debt were strongly correlated with both revolving credit use and small-scale instalment use. Risk aversion correctly predicted the usage of small instalment loans, but it had no relationship with the use of revolving credit. The use of revolving credit was significantly correlated with personality traits of self-control, self-esteem, self-efficacy, deferring gratification, internal locus of control, and impulsiveness; however, the use of small instalment loans was significantly correlated with sensation seeking, impulsiveness, and deferring gratification. We also discovered that some credit card features quickly produced an "illusion of income" that encouraged people to use credit cards for debt [19]. A credit assessment of the clients is becoming more and more important. Big data has made it normal practise to assess a customer's creditworthiness using information found online. Credit managers are unable to assess certain consumers' genuine credit situations because their credit information is either insufficient or missing. This issue may, however, be efficiently resolved with the use of social data, particularly behavioural data and credit rating systems. For the purpose of evaluating credit, this study employed Weibo to collect user behaviour information from Chinese users. Analytic hierarchy process (AHP) and fuzzy comprehensive assessment methods are the two techniques utilised to determine the credit scores of Weibo users [20].

**2.2 Machine learning approaches in behavioural scoring**

In this era, the usage of credit card users is increased in day by day and hence the transaction dataset is also getting increased accordingly, handling these big data using conventional statistics and mathematical models has some limitations up to some extent. In the case of behavioural scoring, we need to hold more on the amount of transactional data itself. Researchers have not given behavioural scoring the same consideration as they have for application scoring. The vast amount of data needed for behavioural scoring studies can be partly blamed for the bias in application scoring research. Furthermore, the publishing of work in this field is sometimes prohibited by the economic sensitivities connected to such a sizable pool of consumer data [21]. Bastani K's studies present a two-stage scoring system. To choose the best loans for investment, the current scoring methodologies only include probabilities of default (PD) prediction, also known as credit scoring, or profitability prediction, also known as profit scoring.[26] Wang c proposed research on lending institutions that are updating their business models as a result of the cutting-edge technologies linked with Big Data, data availability, and computer capacity. In this study, we develop binary classifiers based on deep learning and machine learning models using actual data to forecast the likelihood of loan default. We find that models based on multilayer artificial neural networks are less stable than models based on trees. This raises various issues with the extensive application of deep learning systems in enterprises. The evaluation of each loan's default risk has been a significant responsibility given the rapidly growing peer-to-peer lending market in China. We leverage the online operation behaviour data of borrowers and present a consumer credit scoring technique based on the attention mechanism LSTM. We then use an attention mechanism LSTM network to forecast the likelihood of user default, which is inspired by the study in natural language processing. The approach is tested on a real dataset, and the findings demonstrate that, in comparison to the conventional artificial feature extraction method and the conventional LSTM model, the suggested solution may significantly improve prediction accuracy [26

**2.2 TECHNOLOGICAL LITERATURE**

Following are the technological developments that happened in the area of Bidirectional LSTM during recent years.

 Bidirectional LSTMs are capable of capturing both forward and backward dependencies in sequential data, making them ideal for modelling intricate spending patterns. Time-series financial data, such as credit card transactions, have demonstrated exceptional performance in capturing long-term dependencies using Bidirectional LSTMs.

The bidirectional LSTM exhibits resilience to noisy data, making it a suitable choice for dealing with scenarios where there are missing transactions or incomplete card usage patterns. In order to compensate seamlessly for missing data, Bidirectional LSTM captures the underlying dynamics of sequences effectively [44].

In comparison to other sequential data modelling approaches, Bidirectional LSTM has demonstrated improved accuracy in modelling credit card usage patterns. Its capacity to capture information from sequences both in the forward and backward directions contributes to its superior performance [45].

The banking industry relies heavily on real-time analysis of credit card usage to meet its risk management objectives. The computational efficiency of Bidirectional LSTM, relative to traditional neural networks, enables its utilization in real-time credit card usage analysis by banks.

Bidirectional LSTM facilitates the incorporation of external data sources, such as demographics, geographic locations, and transaction amounts, to improve the modelling and prediction of customers' credit card usage. This capability enhances the overall accuracy and predictive power of the model.

**3.METHODOLOGY**

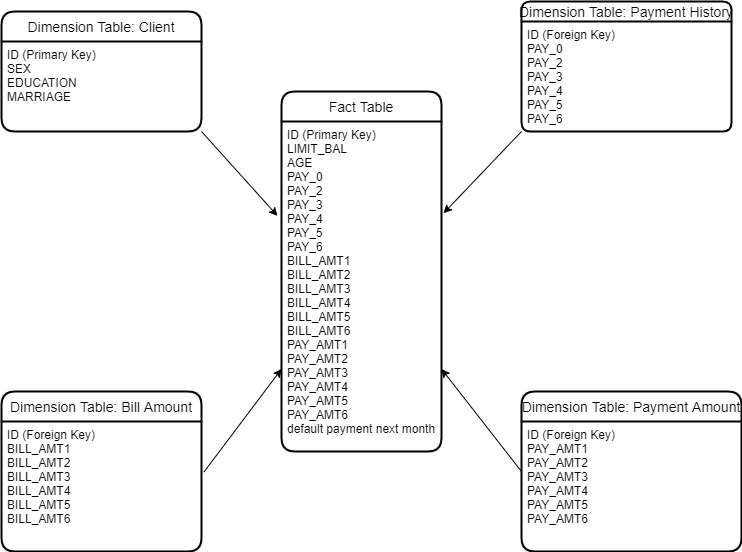
**3.1 Data collection**

The dataset collected for this credit default prediction using deep learning task is an open source or public non transactional credit cards dataset. It shows customers default payment in Taiwan [00]. Usually, banks do not disclose transactional databases in original form, mostly there are appeared to be processed form. Hence, we are using this dataset because this is the only publicly available dataset can convert to temporal form.

The size of the dataset is 30,0000 records and having 23 variables.it is enough for test accuracy, precision, recall, F1-Score.The number of non-default payment is 23,364 and default payment is 6636.There is no missing value in the dataset. The attributes of this datasets are:

1. X1: Amount of the given credit (NT dollar): it includes both the individual consumer credit and his/her family (supplementary) credit.
2. X2: Gender (1 = male; 2 = female).
3. X3: Education (1 = graduate school; 2 = university; 3 = high school; 4 = others).
4. X4: Marital status (1 = married; 2 = single; 3 = others).
5. X5: Age (year).
6. **X6 - X11**: History of past payment. We tracked the past monthly payment records (from April to September, 2005) as follows:
7. X6 = the repayment status in September, 2005;
8. X7 = the repayment status in August, 2005; . . .;
9. X11 = the repayment status in April, 2005. The measurement scale for the repayment status is: -1 = pay duly; 1 = payment delay for one month; 2 = payment delay for two months; . . .; 8 = payment delay for eight months; 9 = payment delay for nine months and above.
10. **X12-X17**: Amount of bill statement (NT dollar).
11. X12 = amount of bill statement in September, 2005;
12. X13 = amount of bill statement in August, 2005; . . .;
13. X17 = amount of bill statement in April, 2005.
14. **X18-X23**: Amount of previous payment (NT dollar).
15. X18 = amount paid in September, 2005;
16. X19 = amount paid in August, 2005; . . .;
17. X23 = amount paid in April, 2005.

The snow flake schema of the dataset with attributes are displayed in the fig.1



**Fig.1 Snowflake Diagram**

**3.2 Preprocessing**

**3.2.1 Label Encoding**

Label encoding is preprocessing technique use for converting categorical values into numerical one. It makes data to be ready for training process. Reduces the memory usage as compared to the one-hot encoding. Which replaces a single categorical column with single numerical column.

**3.2.2 Feature Scaling**

Since there are no any missed values, we can apply feature scaling to check whether all features are on same scale. Which should be beneficial for machine learning algorithms. Normally we are using Min-max Scaling or standardization technique. In min-max scaling the values range between 0 and 1.

Min-max scaling, also known as normalization, is a common data preprocessing technique used in machine learning and statistics. It is applied to rescale the values of a numerical feature within a specific range, typically between 0 and 1.

The formula as:

Scaled Value= (Max Value−Min Value)/ (Original Value−Min Value)

- "Original Value" denotes the actual value of the data point.

- "Min Value" refers to the feature's lowest value throughout the whole dataset.

- "Max Value" refers to the feature's highest possible value throughout the whole dataset.

**3.2.3 Handle missing Value**

There is no any missing value to handle.

**3.3 Data Splitting**

Splitting the data into training and testing sets is known as data splitting. The model is trained using a training set, and its performance on testing data is assessed.

**\*3.4 Feature Selection**

For the job of predicting credit defaults, choose the features that are most pertinent. The choice of features can improve model performance by lowering noise. In case of conventional algorithm like support vector machine, logistic regression and decision tree algorithm we need feature extraction as main step but in the case of deep learning we don’t need to accomplish this step because feature is taken internally.

**3.5. Model Selection**

The machine learning models for classification cannot be used for high dimensional datasets which results in bad scoring quality Classification problems like the Support vector machine, Decision tree and Logistic regression are good for classification tasks but they take much more time while the training process [28]. The next step we have to do with feature extraction from the given dataset which leads to the loss of information hence considering all these factors neural network is much better there no need for any specific feature task is to be done here [29].

Here, we choose some machine learning models for predicting credit defaults. Here to compare conventional machine learning algorithm with bidirectional LSTM which is deep learning model. Conventional model like Support vector machines (SVM), decision trees, random forests, logistic regression algorithms are compared against with bidirectional LSTM and its accuracy and performance evaluation is analysed.

**3.5.1 Support Vector Machine**

An innovative machine learning technique called an SVM is also used to solve problems with credit rating and order. The optimal separation of the input data into two classes (good and poor credit) is made using binary classification using SVMs. By modifying the design of a linear classifier, Cortes and Vapnik [46] first developed SVMs. The function that is utilised to translate the data into a higher dimensional space is the main difference between the SVM model and the linear one. The following kernel functions were proposed to do this: linear, polynomial, radial basis, and sigmoid. A linear model is then utilised to implement the non-linear classes from the mapped high-dimensional feature space of two classes of non-linear data produced by an SVM.

SVMs are efficient in high-dimensional spaces and are thus appropriate for situations with many characteristics. SVMs strive to maximise the margin between classes, which aids in the creation of a decision boundary that generalises well to untried data. This makes SVMs robust against overfitting. SVMs are less likely to overfit because to this feature, especially when utilising a soft margin. By using the kernel method, which implicitly translates the input to a higher-dimensional space where it could be linearly separable, SVMs can handle both linearly separable and non-linearly separable data. Small to medium-sized datasets are appropriate hence works well for this algorithm.

**3.5.2 Logistic Regression**

For the creation of credit scoring models, logistic regression (LOGR) has been regarded as the industry standard up to this point [46]. This widely used statistical method is well-liked for resolving classification and regression issues. A binary outcome variable, often denoted by 0 or 1 (good and bad loans), is modelled using LOGR. Atiya and Palos [47] provide the LOGR formula.

Due to its suitability for binary and multi-class classification tasks, logistic regression is employed in classification since it models the chance of an instance belonging to a certain class. Making decisions is aided by the logistic (sigmoid) function's guarantee that the anticipated probability fall between 0 and 1.

**3.5.3 Decision Tree**

According to Bierman’s [48] proposal, a random forest (RF) is a novel decision tree (DT) technique that consists of a lot of trees made from n subsets of the core dataset, each of which is a tree made up of randomly chosen variables, hence the name "random forest." The final decision class is based on a voting process, where the RF chooses the most well-liked class determined by the trees as the final output class once all the DTs have been formed and trained.

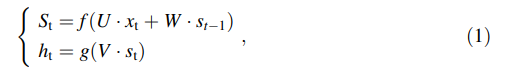
Due to its interpretability and capacity for handling complicated decision boundaries, decision tree algorithms are employed for categorization. They produced a tree-like structure by iteratively dividing the data depending on the feature values. This helps make decision-making processes understandable, which makes it useful for both straightforward and hard categorization jobs.

**3.5.4 Bidirectional LSTM**

The recurrent neural network is a powerful model for sequence data and good for labelling input and output labelling if the labelling is unknown for classification problems. The combination of a Recurrent neural network with long short-term memory is considered to be more fruitful [27]. The performance of bidirectional LSTM is good compared to unidirectional LSTM and conventional recurrent neural networks. Bidirectional neural network process input data in both forward and backward directions and hence can process both current and future data [30].

Directed cycles are created via connections between cells in a recurrent neural network. A hidden state is present in each cell and is updated with each iteration using its previous values. Such a structure serves as a memory and establishes an internal network state.

The RNN equations are as follows:



x = input vector, s = hidden vector of RNN layer values, h = output vector of RNN layer values, U = weight matrix from the input layer to hidden layer, V = weight matrix from hidden layer to the output layer, W = weight matrix from previous time point to current time point of hidden layer, and g and f = activation functions for output and hidden layers. The function of one RNN cell is shown in Fig. 1. We feed the cell the time series signal X one element at a time. Te vector X may be output from another vector or an input vector. From the preceding layer, RNN cell. Te RNN cell keeps its current state. Eq. (1) is used to determine the state st and output ht at each iteration t. RNNs' design enables them to [31]: Learning long-term reliance, meanwhile, has its challenges. It is challenging to understand long-term dependencies because RNN is prone to disappearing gradients during training [32, 33]. An LSTM based on RNN has been presented by Hochreiter and Schmidhuber [34] as a solution to this issue. Similar to RNNs, the experience of the network's inputs always influences how well LSTM predictions perform.

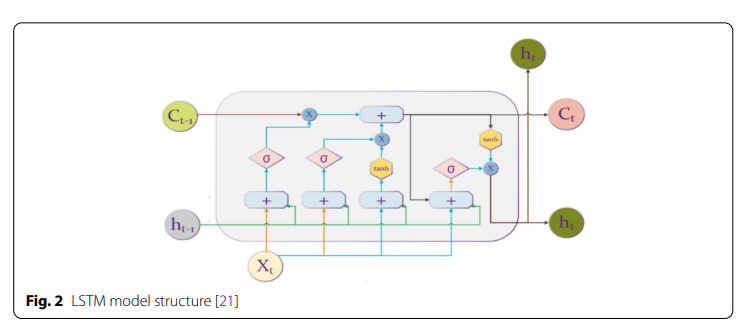


Figure 2 depicts the structure of a single LSTM cell. An input gate, which regulates how many cell states must be stored, is one of the three gates that make up an LSTM cell, as shown in Figure 2. A forget gate, which regulates how much information has to be deleted, and an output gate, which regulates how many cell states are transferred to the following cell must [36, 37]. Internal states may be found in two of these gates. As can be observed, the LSTM cell calculates the candidate vector Ct1 and output vector ht1's subsequent values using the vectors' prior values on each iteration t. Activation functions are used to post-process each gate's output. The activation function's form is



The output gate of an LSTM cell uses the hyperbolic tangent function by default [35]. The range of values for the smooth antisymmetric hyperbolic tangent function, tanh, is [ 1, 1]. The tanh function's output is represented by:



Tanh's key benefit is that it generates zero-centred output, which facilitates the back-propagation process. The specific steps of an LSTM cell are described as follows:

LSTM should choose which knowledge to forget in the first stage. For this reason, the forget gate ft is used to process the data from the prior memory state.



Input gates it determines which data has to be updated on the second step, and the tanh layer updates the candidate vector C t:



The following step involves updating memory states Ct in order to combine the two elements above:

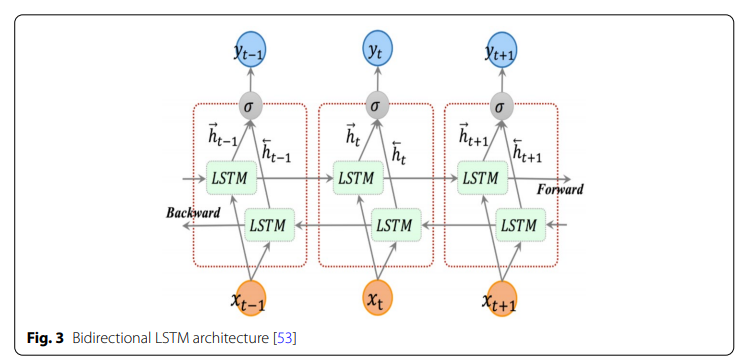
Finally,



As a result, each LSTM layer is [38] characterised by:

1. Matrix Wf and bf, vector, which are the forget gate's parameters
2. The parameters of the input gate are the matrix WC and the vector bC, and
3. The parameters of the output gate are the matrix Wo and the vector bo.

Bidirectional LSTM neural networks were suggested in the research [39] as a way to improve the functionality and learning rate of LSTM neural networks. Bidirectional LSTMs are an extension of conventional LSTMs that can enhance model performance on sequence classification issues, according to Schuster and Paliwal [39]. Bidirectional LSTMs train two LSTMs rather than one on the input sequence in instances where all time steps of the input sequence are known. First on the original input sequence, then on a reversed replica of the original input sequence. This can provide the network extra context and lead to a quicker and even more thorough learning process for the problem.



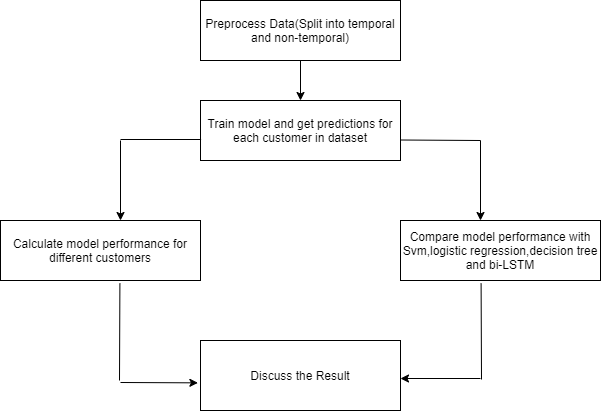
According to fig-3, The forward layer output sequence, h, is computed iteratively using inputs in a positive sequence from time t = 0 to time t = T, whereas the backward layer output sequence,, is calculated using inputs in a negative sequence.

h is computed using the inputs from time t = T to t = 0 in reverse order. The conventional LSTM updating equations, Eqs. (2–7), are used to compute the forward and backward layer outputs. The output vector, Yt, that the Bidirectional LSTM layer produces has each element computed using the equation below:



**PROPOSED MODEL**

The proposed model frame can be visualized as follows in fig-4



**Fig.4 Model frame**

The proposed model having two different types of data from a same dataset. They are temporal and non-temporal where temporal data which shows customer behaviour in time and which reshaped into three-dimensional array of shape and each customers have three temporal features.

1) Each previous month's payment was late.

2) Amount of bills due at the conclusion of each previous month.

3) Amount of each monthly payment.

Non temporal having some overlaps to statis features. Static data having both numerical and categorical. In this dataset context data can be considered a specific type of static features. Therefore for each customers there have been eight non-temporal features.

1) Amount of credit granted

2) Gender

3) Education—graduate school

4) Education—university

5) Education—high school

6) Education—others

7) Marital status

8) Age

The design of a neural network is often determined in consideration of the training data.

To develop an efficient model with a small number of parameters (weights), it is crucial to employ spatial organisation and order of the input data. RNNs and LSTMs are frequently employed in neural networks for temporal input. However, the LSTM network is not appropriate for combining temporal and non-temporal input. Feeding non-temporal data into the dense layers of the LSTM's top layer is one option, however in this instance, non-temporal characteristics are only utilised in the final stage of the model.

The optional query input for the attention layer serves as the temporal input's context. To provide this layer with a context for customers' excellent or bad payment behaviour, we leverage non-temporal data as a query input. Therefore, such a layer is capable of differentiating client financial behaviour while taking into consideration their level of education and marital status, as well as their gender and age.Bidirectional LSTMs make up the first two layers, followed by attention. The output of the Attention layer and the nontemporal client data are combined in the two final layers. One neuron makes up the single layer.

There are various phases in the suggested framework. In the beginning, the dataset is prepared and packaged for usage by the bidirectional LSTM classifier. The fvefold validation approach is then used to provide predictions for each client in the dataset. The performance measurements are calculated for several client categories that are of financial importance to the banking institutions (banks are particularly interested in consumers with a dismal payment history). The model is compared against benchmark models using a variety of performance metrics to illustrate its performance. In the final part, the findings are addressed.

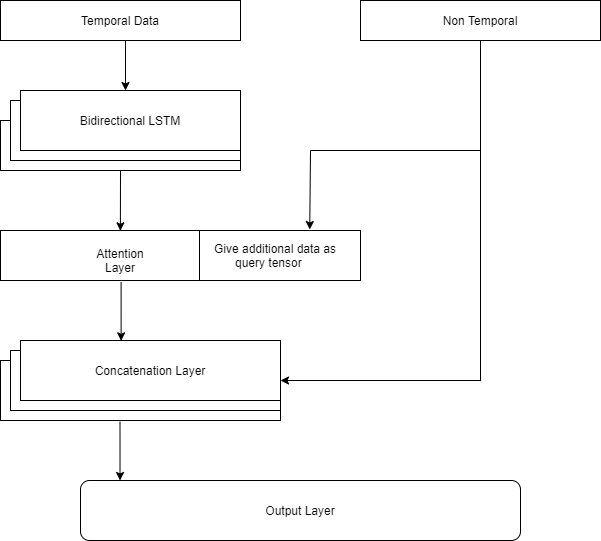


Figure 3: The general pipeline of LSTM and Bi-LSTM for Default prediction

**PERFORMANCE EVALUATION**

It is essential to evaluate the performance of credit default prediction models in order to determine how well they forecast consumer behaviour and spot possible defaulters. The effectiveness of such models may be evaluated using a number of assessment indicators. These measurements are some of the most popular.

***Accuracy***: By comparing the proportion of accurate forecasts to all other predictions produced, accuracy assesses how accurate the model's predictions are overall. However, accuracy by itself could not give a whole picture, particularly if the data are unbalanced (i.e., the proportion of defaults to non-defaults is substantially smaller).

***Precision***: Out of all occurrences anticipated to default, precision represents the percentage of accurately predicted defaults. It focuses on how trustworthy the optimistic forecasts are. A low false positive rate is shown by high accuracy, which means that the majority of anticipated defaults are accurate.

***Recall (Sensitivity or True Positive Rate):*** Recall calculates the percentage of expected defaults that really occur that were right. It emphasises the model's capacity to locate genuine positives. A low false negative rate is shown by a high recall, which means that the majority of real defaults are found.

***F1 Score:*** The harmonic mean of recall and accuracy is the F1 score. It offers a fair assessment of both measurements and is particularly helpful in cases where the data is unbalanced. A high F1 score suggests that accuracy and recall are well-balanced.

***Receiver operating characteristic area under the curve (AUC-ROC):*** The trade-off between true positive rate (TPR) and false positive rate (FPR) at different classification thresholds is represented by the ROC curve. The model's capacity to differentiate defaulters from non-defaulters across various thresholds is measured by AUC-ROC. Better model performance is indicated by higher AUC-ROC values.

***Confusion matrix:*** This gives an explanation of model performance and displays true positives, true negatives, false positives, and false negatives. It may be used to determine a number of performance parameters, including F1 score, recall, accuracy, and precision. It cannot be used for regression problems.

# **Teamwork link**

**https://universityofroehampton15.teamwork.com/app/home/projects**

**GitHub link**

**https://github.com/Meritc/Data-Science**

# **REFERENCE**

# [1] Loan default prediction using a credit rating-specific and multi-objective ensemble learning scheme

# [2] Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review

# [3] Benchmarking state-of-the-art classification algorithms for credit scoring

# [4] Atiya AF, Parlos AG (2000) New results on recurrent network training: unifying the algorithms and accelerating convergence. IEEE Trans Neural Netw 11(3):697–709. https://doi.org/10.1109/ 72.846741

# [5] ellotti T, Crook J (2009) Credit scoring with macroeconomic variables using survival analysis. J Oper Res Soc 60(12):1699–1707. <https://doi.org/10.1057/jors.2008.130>

# [6] . Bhatia S, Sharma P, Burman R, Hazari S, Hande R (2017) Credit scoring using machine learning techniques. Int J Comput Appl 161:1–4

# [7] Thomas LC (2000) A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. Int J Forecast 16(2):149–172. <https://doi.org/10.1016/s0169-2070(00)00034-0>

# [8]Thomas LC, Ho J, Scherer W (2001) Time will tell: behavioural scoring and the dynamics of consumer credit assessment. IMA J Manag Math 12(1):89–103. [https://doi.org/10.1093/imaman/12.1. 89](https://doi.org/10.1093/imaman/12.1.%2089)

# [9] Louzada F, Ara A, Fernandes GB (2016) Classification methods applied to credit scoring: systematic review and overall comparison. Surv Oper Res Manag Sci 21(2):117–134. https://doi. org/10.1016/j.sorms.2016.10.001

# [10] Kennedy K, Namee BM, Delany SJ, O’Sullivan M, Watson N (2013) A window of opportunity: assessing behavioural scoring. Expert Syst Appl 40(4):1372–1380. [https://doi.org/10.1016/j. eswa.2012.08.052](https://doi.org/10.1016/j.%20eswa.2012.08.052)

# [11] Setiono R, Thong JYL, Yap C-S (1998) Symbolic rule extraction from neural networks. Inf Manag 34(2):91–101. https://doi.org/ 10.1016/s0378-7206(98)00048-2

# [12] Sharda R, Wilson RL (1996) Neural network experiments in business failures predication: a review of predictive performance issues. Int J Comput Intell Organ. [https://doi.org/10.1109/hicss. 1993.284245](https://doi.org/10.1109/hicss.%201993.284245)

# [13] Hsieh H-I, Lee T-P, Lee T-S (2010) Data mining in building behavioral scoring models. In: 2010 international conference on computational intelligence and software engineering, pp 1–4. <https://doi.org/10.1109/cise.2010.5677005>

# [14]Sarlija N, Bensic M, Zekic-Susac M (2009) Comparison procedure of predicting the time to default in behavioural scoring. Expert Syst Appl 36(5):8778–8788. [https://doi.org/10.1016/j. eswa.2008.11.042](https://doi.org/10.1016/j.%20eswa.2008.11.042)

# [15] Gui L (2019) Application of machine learning algorithms in predicting credit card default payment. University of California

# [16] Bellotti T, Crook J (2009) Credit scoring with macroeconomic variables using survival analysis. J Oper Res Soc 60(12):1699–1707. <https://doi.org/10.1057/jors.2008.130>

# [17] Bellotti T, Crook J (2013) Forecasting and stress testing credit card default using dynamic models. Int J Forecast 29(4):563–574. <https://doi.org/10.1016/j.ijforecast.2013.04.003>

# [18] Agarwal S, Chomsisengphet S, Liu C, Song C, Souleles NS (2018) Benefits of relationship banking: Evidence from consumer credit markets. J Monetary Econ 96:16–32

# [19] Wang L, Lu W, Malhotra NK (2011) Demographics, attitude, personality and credit card features correlate with credit card debt: a view from China. J Econ Psychol 32(1):179–193. https:// doi.org/10.1016/j.joep.2010.11.006

# [20] Yu J, Yao J, Chen J (2019) Credit scoring with AHP and fuzzy comprehensive evaluation based on behavioural data from weibo platform. Tehn Vjes Tech Gaz 26(2):462–470. [https://doi.org/10. 17559/tv-20181217180231](https://doi.org/10.%2017559/tv-20181217180231)

# [21] Kennedy K, Namee BM, Delany SJ, O’Sullivan M, Watson N (2013) A window of opportunity: assessing behavioural scoring. Expert Syst Appl 40(4):1372–1380. [https://doi.org/10.1016/j. eswa.2012.08.052](https://doi.org/10.1016/j.%20eswa.2012.08.052)

# [22] Pereira S (2019) Modelling credit card customer behaviour. Work Project presented as a partial requirement for Degree of Master of Statistics and Information Management, with a specialization in Information Analysis and Management

# [23]Pereira S (2019) Modelling credit card customer behaviour. Work Project presented as a partial requirement for Degree of Master of Statistics and Information Management, with a specialization in Information Analysis and Management

# [24] Bastani K, Asgari E, Namavari H (2019) Wide and deep learning for peer-to-peer lending. Expert Syst Appl 134:209–224. https:// doi.org/10.1016/j.eswa.2019.05.042

# [25] Addo P, Guegan D, Hassani B (2018) Credit risk analysis using machine and deep learning models. Risks 6(2):38. https://doi.org/ 10.3390/risks6020038

# [26]Wang C, Han D, Liu Q, Luo S (2018) A Deep learning approach for credit scoring of peer-to-peer lending using attention Neural Computing and Applications (2022) 34:5839–5866 5865 123 mechanism LSTM. IEEE Access 7:2161–2168. https://doi.org/10. 1109/access.2018.2887138 72.

[27] Graves A, Mohamed A, Hinton G (2013) Speech recognition with deep recurrent neural networks. In: ICASSP, IEEE international conference on acoustics, speech and signal processing—Proceedings, vol 38, pp 6645–6649. [https://doi.org/10.1109/ICASSP. 2013.6638947](https://doi.org/10.1109/ICASSP.%202013.6638947)

[28] Wang C, Han D, Liu Q, Luo S (2018) A Deep learning approach for credit scoring of peer-to-peer lending using attention Neural Computing and Applications (2022) 34:5839–5866 5865 123 mechanism LSTM. IEEE Access 7:2161–2168. [https://doi.org/10. 1109/access.2018.2887138](https://doi.org/10.%201109/access.2018.2887138)

[29] A deep learning model for behavioural credit scoring in banks. Maher Ala’raj1,Maysam F. Abbod2,Munir Majdalawieh1,Luay Juma

[30] Bidirectional LSTM Networks for Improved Phoneme Classification and Recognition, A[lex Graves](https://link.springer.com/chapter/10.1007/11550907_126#auth-Alex-Graves), [Santiago Fernández](https://link.springer.com/chapter/10.1007/11550907_126#auth-Santiago-Fern_ndez) &[Jürgen Schmidhuber](https://link.springer.com/chapter/10.1007/11550907_126#auth-J_rgen-Schmidhuber)

[31] . Wang C, Han D, Liu Q, Luo S. A deep learning approach for credit scoring of peer-to-peer lending using attention mechanism LSTM. IEEE Access. 2019;7:2161–8. <https://doi.org/10.1109/access.2018.2887138>.

[32]Bengio Y, Frasconi P, Simard P. The problem of learning long-term dependencies in recurrent networks. In: IEEE international conference on neural networks. 1993. p. 1183–8. 47.

[33]Srinivasan K, Cherukuri AK, Vincent DR, Garg A, Chen BY. Chen, an efcient implementation of artifcial neural networks with K-fold cross-validation for process optimization. J Internet Technol. 2019;20:1213–25. https://doi.org/10.3966/16079 2642019072004020. 48.

[34]Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9:1735–80. https://doi.org/10.1162/neco. 1997.9.8.1735.

[35] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9:1735–80. https://doi.org/10.1162/neco. 1997.9.8.1735.

[36] Lai CY, Chen RC, Caraka RE. Prediction average stock price market using LSTM. 2019.

[37]Toharudin T, Pontoh R, Caraka R, Zahroh S, Lee Y, Chen R. Employing long short-term memory and facebook prophet model in air temperature forecasting. Commun Stat Simul Comput. 2021. https://doi.org/10.1080/03610918.2020.18543 02.

[38]Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9:1735–80. https://doi.org/10.1162/neco. 1997.9.8.1735

# [39] Schuster M, Paliwal K. Bidirectional recurrent neural networks. IEEE Trans Signal Proces. 1997;45:2673–81. https://doi.org/ 10.1109/78.650093.

[40] Dyché J. The CRM handbook: a business guide to customer relationship management. Boston: Addison-Wesley Longman Publishing; 2001

[41] Anderson R. The credit scoring toolkit: theory and practice for retail credit risk management and decision automation. Oxford: Oxford University Press; 2007

[42] Modelling customer's credit card behaviour using bidirectional LSTM neural networks. Maher Ala’raj1\*, Maysam F. Abbod2 and Munir Majdalawieh1

[43] Liu Y. New issues in credit scoring application. Institut für Wirtschaftsinformatik, Abteilung Wirtschaftsinformatik II, Georg-August-Universität, Göttingen. 2001.

[44] Saha, Swapnil Sayan, Sandeep Singh Sandha, and Mani Srivastava. "Deep convolutional bidirectional lstm for complex activity recognition with missing data." Human Activity Recognition Challenge (2021): 39-53.

[45]Laitala, Juho, et al. "Robust ECG R-peak detection using LSTM." Proceedings of the 35th annual ACM symposium on applied computing. 2020

[46] Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20(3):273–97. <https://doi.org/10.1007/BF00994018>

[47]Atiya AF, Parlos AG. New results on recurrent network training: unifying the algorithms and accelerating convergence. IEEE Trans Neural Netw. 2000;11(3):697–709. <https://doi.org/10.1109/72.846741>

[48]Breiman L. Random forests. Mach Learn. 2001;45(1):5–32. https://doi.org/10.1023/A:1010933404324.