LINEAR SCALING TAU CORRELATION AND TEMPORAL AWARE DEEP LEARNING-BASED FINANCIAL BIG DATA PREDICTIVE ANALYSIS FOR SME

[1]Dr P. Pushpa M.Sc, M.Phil, PhD,
Lecturer, School of Software, East China University of Technology, Nanchang, Jianxi, China.

[2]S. Sathish Kumar MCA
Lecturer, School of International Education, East China University of Technology, Nanchang, Jianxi, China.

Abstract

Small and medium-sized enterprises (SMEs) are considered a significant aspect of the national economic system that improves employment opportunities, enhances the standard of living, and paves the way for economic and sustainable development with technological innovation. In today's business environment concerning big data and deep learning, it remains paramount for SMEs to utilise big data services in a more compatible manner. Therefore, it has adopted the inception. With the era of big data in China, the swift evolution of information has affected the economic environment. As for SMEs, precise market research has based on marketing must. Therefore, the concerned party must provide, which necessitates screening enormous market information to obtain both parties' helpful information. To address this aspect, in this work, a method called Linear Scaling Tau and Temporal-aware Long Short Term Memory-based (LST-TLSTM) Financial Predictive analysis is proposed to predict the financial assistance of an organisation in an accurate and timely manner, therefore, minimising the errors in prediction for SMEs. First, using a Linear Scaling-based Preprocessing model has pre-processed the company's data or features. Second, the Kendall Tau Correlation Coefficient-based feature selection model has selected the relevant features of the processed data for further processing. Finally, utilising Temporal-aware Long Short Term, memory-based Financial Predictive analysis model has made the predictive financial analysis. Experimental results on the Chinese financial dataset have demonstrated higher prediction accuracy with a better time and error rate of the LST-TLSTM method than that of the other two state-of-the-art methods.

Keywords: Small and Medium-sized Enterprises, Big Data, Deep Learning, Linear Scaling, Kendall Tau, Correlation Coefficient, Long Short Term Memory

1. Introduction

Financial credit reporting refers to a set of data sequenced in a sequential order that into uni-variate financial credit has to split typically and multivariate financial credit following the attribute's dimension. Hence, financial credit is one of the moderately complicated data types in data analysis that is extensively appertained in finance, economics, industrial engineering, etc. The financial credit not only possesses high dimensional features like long time dimension factor, huge attribute variables, and enormous voluminous data, but circumstances like uncertainty, dynamics, and conceptual drift have completed also. These features make it unfeasible to straightly and efficiently apply the conventional materials and methods to data mining and analysis of financial aspects. Specifically, the tremendous data model in the era of big data has underscored the significance of deep learning methods.

Knowledge discovery was made properly with the obtained patterns has obtained a Multidimensional Attribute Sparse Large Data (MASLD) proposed in [1] based on big data to perform clustering analysis wherein the patterns. Here, clustering utilised the holistic distance between financial credit via sparse extensive data and multidimensional attribute large data. The clustering was performed based on the overall distance matrix and component approximate distance matrix. The proposed MASLD method reflected the overall data features significantly and enhanced the clustering effect via correlation relationship between essential elements. As a result, to be satisfactory, find the clustering accuracy involved in financial credit reporting.

Big data bestows extensive research prospects for in-depth research and apprehending market behaviour on the financial aspect. However, the issues addressed for high-frequency data or big data are comparatively less than the issues concerning low-frequency or traditional data. Hence, volatility remains a significant aspect or index of market risk and is, therefore, more important to investors, government regulators and capital markets. To this end, a support vector machine was introduced in [2] to enhance the business's short term and long term volatility. With this linear non-separable function employed with SVM, the short- and medium-term prediction accuracy for range volatility improved.

Motivated by the above two works and the state-of-the-art works presented in the literature sections, this work improves prediction accuracy, prediction time and error rate involved in financial Big Data predictive analysis, Linear Scaling Tau and Temporal-aware Long Short Term Memory-based (LST-TLSTM) method proposed.

1.1 Contributions

To summarise, the contributions of this work are listed below.

- To design an efficient method for financial Big Data predictive analysis involving huge samples.
- To handle class imbalance between features using Linear Scaling-based Preprocessing model.
- Kendall Tau Correlation Coefficient-based selection model is applied to the raw sample Chinese finance data instances to find the relevant features.
- To get better prediction accuracy with minimum time and error rate by selecting the computationally efficient and relevant features using Temporal-aware Long Short Term Memory-based Financial Predictive analysis algorithm.
- The performance of the proposed LST-TLSTM based soil quality prediction method compared with the state-of-the-art methods.
- Finally, using the Chinese Development Finance Dataset with different performance metrics for a series of experiments. The metric assessment results like prediction accuracy, prediction time, and identifying the LST-TLSTM method's performance improvement over the existing works have utilised error rate.

1.2 The organisation of the work

The remainder of our paper is structured as follows. Section 2 reviews the literature on financial predictive analysis for SMEs using optimisation, machine and deep learning. Section 3 explains the construction and theory of the proposed method, Linear Scaling Tau and Temporal-aware Long Short Term Memory-based (LST-TLSTM) Financial Predictive analysis. Section 4 discusses the results of the experiment. Section 5 has provided implications of the study. Finally, Section 6 gives conclusions.

2. Related works

Small and medium-sized enterprises (SMEs) are a significant aspect globally. In China, SMEs are the basics of economic evolution and have become paramount drivers of social progress. The Chinese government has developed several policies to aid the evolution of private companies, consisting of both credit increases and minimisation concerning taxes and fees.

In [3], proposed a systematic literature review on designing enterprise data strategy for influencing data science strategy. However, business model transformation proposed another business model transformation for small and medium enterprises for business analytics in [4]. In [5], time-series-based stock prices were proposed based on the investment portfolio and predictions. A long-short-term-memory-based deep neural network designed a new regression model with this objective. To construct an investment portfolio using an optimisation model, reducing the error significant utilised the resultant prediction.

Financial ratios were utilised via an accounting model and constructed several bankruptcy prediction methods for SMEs. In addition, a bankruptcy prediction method for SMEs utilising transactional data and payment network–based variables was utilised [6]. Moreover, offline and online tests were also performed for predictive potentiality and economic benefit of transactional data–based variables, improving accuracy. However, with high data disparity, the error rate was also increased. The multitask feature extraction model was designed in [7] to select highly significant and relevant user features effectively and precisely from an incredibly immense number of user feature fields.

With the increasing evolution of the Internet of Things (IoT) technology, many fields designed several IoT-based applications. Some fields are finance, IoT-based healthcare, managing resources, and industrial financing. As far as banks and financial organisations are concerned, IoT solutions assist in gaining real-time data, resulting in a more efficient evaluation algorithm of financial risk management.

In [8], Particle Swarm Optimization (PSO) based Backpropagation (BP) neural network was proposed for financial risk management in banks with the aid of IoT. Here, a nonlinear parallel optimisation method using Hadoop Distributed File System (HDFS) techniques was proposed both for on-balance sheet items and off-balance sheet items. This design of Big Data clusters, in a significant manner, find to reduce processing time.

As the probability has referred to the credit risk of SMEs in Supply Chain Finance (SCF), the SME would not collect the amounts given for loans that are said to be derived from financing as far as the SCF platform is concerned. However, said to be overlooked the behavioural data. Therefore, a novel method called, DeepRisk was proposed in [9] to fuse enterprise demographic data and finance behavioural data for predicting the credit risk of SMEs in SCF. With this, precision and, to a greater extent, improved recall. However, another method to identify legal judgments [10] efficiently predicts credit risk and obtains relevant features within efficient legal judgments.

FinTech, combining the integration of finance and technology, has become the business catchword in recent years. However, though technology and coexists have said finance, only in the recent few years it has been fully realised, narrating the connection of modern, specifically the internet analogous in the financial services industry.

In [11], the antecedents of the utilisation of FinTech finance by businesses with the specific purpose of concentrating on the ownership and governance and propose the business executions that may alter the association with traditional stimulations for external finance. A detailed study identified weak and robust credit rationing via innovation channels and proposed inferring the negative influence in [12]. To further alleviate the problems of the prevailing methods in trafficking with the non-stationary and nonlinear characteristics of financial time series data, an ensemble method based on data denoising and wavelet transform (WT) was presented in [13] to construct a data prediction method.

The financial market consists explicitly of stocks, futures, and bonds. Several research persons dig into the financial field by constructing and designing different deep learning methods to predict volatility, trends, and cost. For example, to propose a review of the artificial intelligence method applied to stock tracing in [14]. However, another literature review and classification in finance investigate banking in [15]. Finally, to ensure better assistance for the investors in the financial data evaluation and decision-making process, a well found and proposed efficient data prediction method in [16] based on the financial data analysis integrated deep learning for further analysing financial data based on deep learning.

In [17], blockchain was integrated with a fuzzy neural network to investigate the credit risk of SMEs from the angle of supply chain finance. Also, via blockchain technology constructed, the supply chain financial system and fused supply chain financial information into blocks. Moreover, the fuzzy neural network was utilised in financial data processing and risk assessment, efficiently addressing and enhancing the risk processing level of the supply chain. Finally, in [18], a machine learning approach was designed to predict the performance of rural bank structures in Ghana.

Financial supervision plays an essential role in constructing a market economy, but financial data is a non-stationary, nonlinear, and low signal-to-noise ratio, so an effective financial detection method is needed. In [19], two machine learning techniques, namely, decision tree and random forest, are utilised to detect the company's financial data. In addition, ensure a systematic literature review on SMEs using modern analytical techniques to provide enhanced predictive results [20].

Based on the materials above and methods in this work, Linear Scaling Tau and to predict financial assistance using the Chinese Development Finance Dataset proposed Temporal-aware Long Short Term Memory-based (LST-TLSTM) Financial Predictive analysis. The following sections provide a detailed description of the LST-TLSTM method.

3. Methodology

Improving access to financing for small and mid-size enterprises (SMEs) is a topic of much interest to policymakers and academics. Moreover, upon comparison with the larger firms, the SME sector is exceedingly sensitive owing to the availability and accessibility of big data. Also, with the Big Data available in the market, SMEs are pushed in screening enormous market information to obtain helpful information. In this work, a method called Linear Scaling Tau provides timely proposed Temporal-aware Long Short Term Memory-based (LST-TLSTM) Financial Predictive analysis accurate inputs with minimum error to the concerned parties. Below has shown the block diagram of LST-TLSTM.

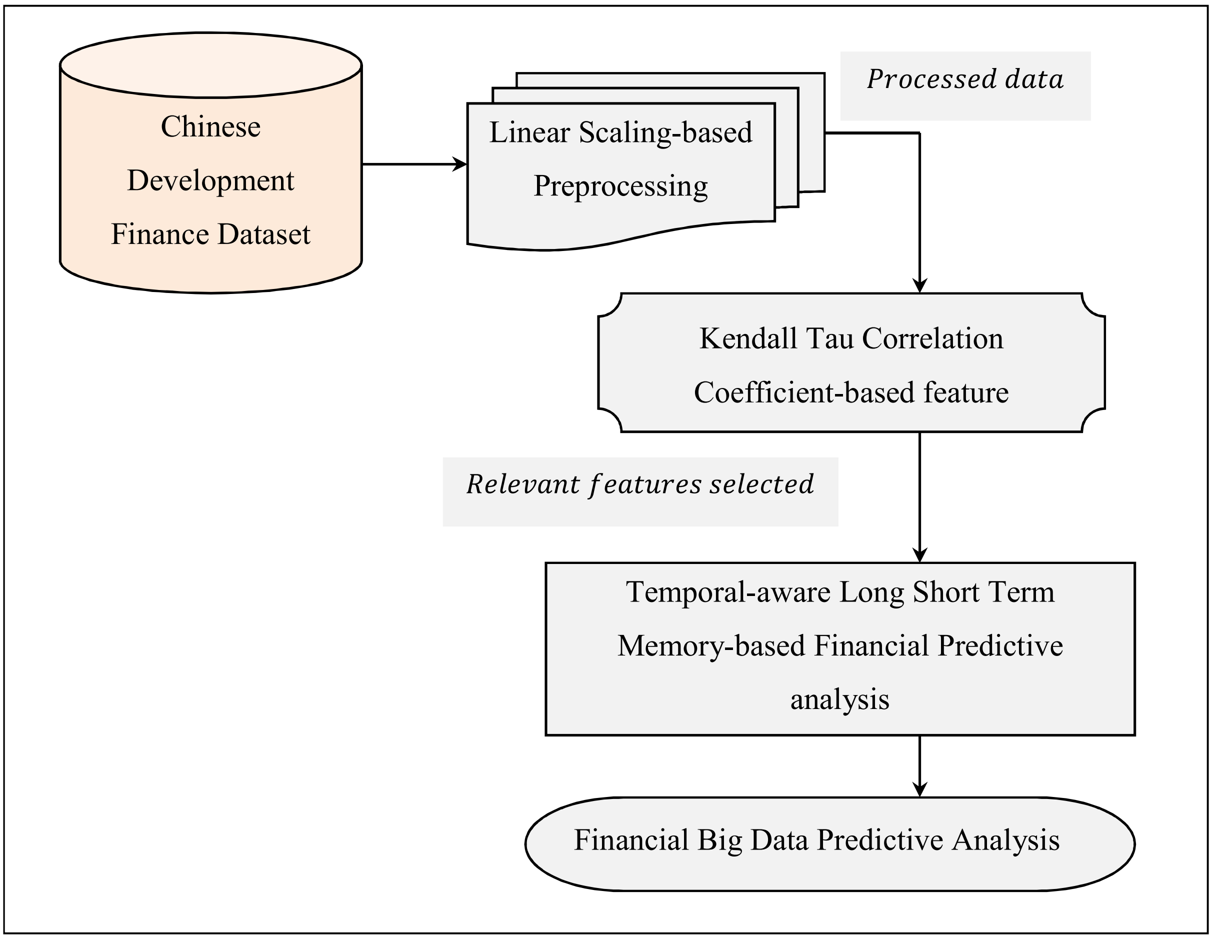

Figure 1 Block diagram of Linear Scaling Tau and Temporal-aware Long Short Term Memory-based Financial Predictive analysis

The above figure has split into three sections of the proposed LST-TLSTM method. They are pre-processing using Linear Scaling, feature selection using Kendall Tau Correlation Coefficient, and predictive financial analysis by utilising Temporal-aware Long Short Term Memory. The following sections have given a detailed description of the LST-TLSTM method.

Linear Scaling-based Preprocessing model

The tremendous heightening in the scale of data has been noticed in recent years as a crucial element as far as Big Data for SMEs is concerned. Hence, big data addressing is challenging and time-consuming and necessitates a substantial computational framework to make specific successful predictive analyses. Moreover, the imbalanced class nature of the raw dataset creates massive havoc compromising the entire predictive model involving financial Big Data for SMEs.

In this work, a Linear Scaling-based Preprocessing model is applied to the raw financial big data to handle a class imbalance effectively. Figure 2 shows the block diagram of the Linear Scaling-based Preprocessing model.

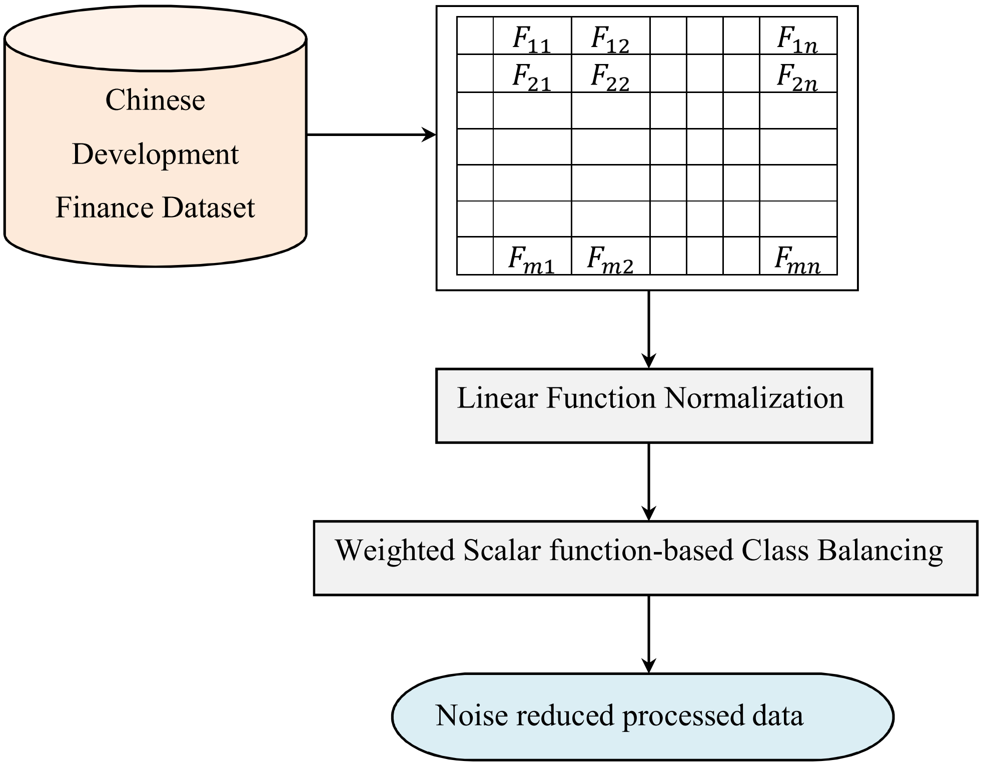

Figure 2 Block diagram of Linear Scaling-based Preprocessing model

As shown in the above figure, the Linear Scaling-based Preprocessing model is presented, consisting of the raw data, the shift from relational data into SME network, and the handling of class imbalance among the model features or variables. For example, in the real world financial dataset, most of the credit samples are creditworthy or negative class, and the remaining credit samples are non-creditworthy or default cases or positive class. In other words, the financial dataset involving both temporal and geographic coverage reveals a class imbalance distribution.

In small business financial predictive analysis data modelling, misclassification of non-creditworthy projects is extensively higher than misclassifying creditworthy projects as non-creditworthy. Therefore, we removed samples from both the features and the data space during the data pre-processing by obtaining new negative class samples via linear function normalisation. This class imbalance update finally in the minority samples improves the prediction accuracy even with big data.

The first step here remains in constructing a feature evaluation matrix. Based on the input data object set 'F', the dataset consists of 'n' data. All the data comprises 'm' features, then the expression form or the feature matrix 'FM' (i.e., raw financial data) of the class object is mathematically stated below.

$$FM = \begin{pmatrix} F_{11} & F_{12} & \dots & F_{1n} \\ F_{21} & F_{22} & \dots & F_{2n} \\ \dots & \dots & \dots & \dots \\ F_{m1} & F_{m2} & \dots & F_{mn} \end{pmatrix} \tag{1}$$

In practical applications, the dimension of each feature in the Global Chinese Development Finance dataset is distinct. Therefore, to eliminate the corresponding grouping influence, discarded the influence of dimension on distinct features. The data of target objects or projects for sanctioning by the institutions should be analogous to estimate each feature's contribution to the overall grouping. Hence, the above feature evaluation matrix is normalised based on the dimension of the feature. Our work performs this by employing Linear Function Normalization (i.e., relational data in an SME network).

$$FMNorm_{ij} = \frac{F_{ij} - Min\,[F_{ij}]}{Max\,[F_{ij}] - Min\,[F_{ij}]} \tag{2}$$

From the above equation (2), '$FMNorm_{ij}$' represents the feature matrix evaluated after normalisation for the corresponding features, 'F_{ij}' from the feature matrix before normalisation. 'F_{ij}', '$Min\,[F_{ij}]$', '$Max\,[F_{ij}]$' representing the minimum and maximum values of distinct features in the original dataset. Finally, the feature matrix after normalisation. '$FMNorm_{ij}$' is further modelled as an optimal feature matrix (i.e., class imbalance handling) by employing the Weighted Scalar function, '$FMNorm_{ij}, 0 < FMNorm_{ij} \leq 1$'.

$$WS = \sum_{i,j=1}^{m,n} \frac{FMNorm_{ij}}{SUM\left(FMNorm_{ij}\right)} \quad (3)$$

From the above equation (3), the weighted scalar function 'WS' is obtained based on the feature matrix evaluated after normalisation. '$FMNorm_{ij}$' and the sum of its respective values '$SUM\left(FMNorm_{ij}\right)$'. If the resultant value of the optimal feature matrix is '1', then it is considered the optimal feature evaluation model for handling class imbalance issues. Given below has been formulated mathematically. Finally, given below is the optimal class balanced pre-processed data.

$$PD = FMNorm_{opt} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \dots & \dots & \dots & \dots \\ 1 & 1 & \dots & 1 \end{pmatrix} \quad (4)$$

With the above results as in (4), the noise reduced pre-processed data or processed is obtained. Given below is the pseudo-code representation of Linear Scaling-based Preprocessing.

Input: Financial Dataset 'DS', Features '$F = F_1, F_2, \dots, F_n$'
Output: noise-reduced optimal class balanced pre-processed data 'PD'
1: **Initialise** data ''n'', features 'm' 2: **Begin** 3: **For** each Financial Dataset 'DS' with Features 'F' 4: Obtain feature matrix 'FM' of the class object as in (1) 5: Convert relational data into SME network using Linear Function as in (2) 6: **Evaluate** Weighted Scalar function for each feature as in (3) 7: **Return** optimal class balanced pre-processed data 'PD' as in (4) 8: **End for** 9: **End**

Algorithm 1 Linear Scaling-based Preprocessing

As given in the Linear Scaling-based Preprocessing, the objective remains to pre-process the data by handling class imbalance so that a more significant extent reduces the predictive error. With this objective, two distinct functions, linear and weighted scalar functions, are applied to the raw feature matrix data. By applying these two functions, the class imbalanced learning performance is investigated separately for each class or feature, improving the prediction error rate.

3.1 Kendall Tau Correlation Coefficient-based feature selection model

Upon successful pre-processing of raw data performed using linear scaling function, the next step in predictive Big Data analysis for SME is the feature selection. Selecting pertinent features is paramount for enhancing prediction accuracy and minimising computation time. Feature selection is considered the dimensionality reduction model that minimises dataset dimensionality by eliminating irrelevant and redundant features. Moreover, it enhances the predictive model's interpretability specificity and minimises its cost. Based on mutual information, the correlation between labels and features is measured to identify the redundant features and discard the same from further processing using Kendall Tau Correlation Coefficient. Figure 3 shows the block diagram of the Kendall Tau Correlation Coefficient-based feature selection model.

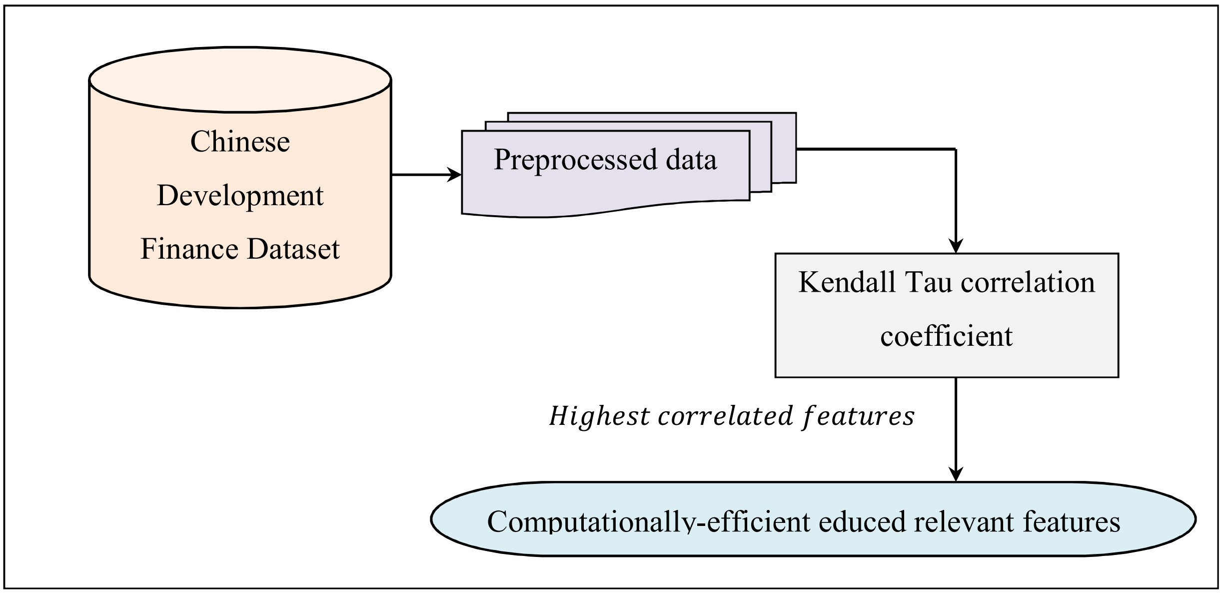

Figure 3 Block diagram of Kendall Tau Correlation Coefficient-based feature selection

As shown in the above figure, with the pre-processed data provided as input, the Kendall Tau Correlation Coefficient measures the correlation between two feature vectors and 'τ' (tau) represent. Here, by employing the Kendall Tau correlation coefficient, the correlation between labels and features is first measured, identifying the redundancy between features for the pre-processed data evaluated followed the highest correlation. Therefore, the total number of projects below has based on the sample rank correlation and total financial value.

Table-1 Sample rank correlation based on the total number of projects and total financial value

Pre-processed data	S_1	S_2	S_3	S_4	S_5	$Ranking$
Rank R_i by (total number of projects)	152	224	303	425	250	$S_4 > S_3 > S_5 > S_2 > S_1$
Rank R_j by (total financial value)	25000	3850	3125	51455	38325	$S_4 > S_5 > S_1 > S_2 > S_3$

Let us consider two different features (i.e., the total number of projects 'F_i' and total financial value, 'F_j') from the input data set with the pre-processed data to perform correlation. Given below, stated mathematically.

$$\tau = \frac{F_i - F_j}{\sqrt{(A_3 - A_1)(A_3 - A_2)}} \quad (5)$$

$$A_1 = \frac{1}{2}\sum_{i=1}^{m} X_i(X_i - 1) \quad (6)$$

$$A_2 = \frac{1}{2}\sum_{j=1}^{n} Y_j\left(Y_j - 1\right) \quad (7)$$

$$A_3 = \frac{1}{2}S(S - 1) \quad (8)$$

From the above equations (5), (6), (7) and (8), 'A_1' is obtained by combining the similar features in 'F' into a small set represented by 'm' and 'X_i' represents the number of features in the '$i - th$' set. In a similar manner, 'A_2' is evaluated based on the set 'Y'. The smaller set has combined the similar features in 'Y', represented by 'n' and 'Y_j' denotes the number of features in the '$j - th$' set. With the presence of multiple features in the Chinese development finance data set, and the feature 'F_i' with the highest correlation is mathematically formulated as given below.

$$Res(\tau) = KCC(F_i - F_j) = MAX\ [\tau] \tag{9}$$

The above Kendall Correlation Coefficient ranges from '-1' to '1'. When '$Res(\tau)$' is equal to '1' (i.e., relevant features), which means that the resultant value possesses consistent rank correlation; on the other hand, when '$Res(\tau)$' is '-1' (i.e., irrelevant features). Below is the pseudo-code representation of Kendall Tau Correlation Coefficient-based feature selection.

Input: Financial Dataset 'DS', Features '$F = F_1, F_2, \ldots, F_n$'
Output: Computationally relevant feature selection
Begin
For each financial Dataset 'DS' with Features 'F' and pre-processed data 'PD'
Measure correlation between labels and features as in (6)
Estimate Kendall Correlation Coefficient between features as in (7)
If '$Res(\tau) = 1$'
Ensures consistent rank correlation
Features are relevant 'RF'
End if
If '$Res(\tau) = -1$'
Ensures inconsistent rank correlation
Features are irrelevant 'IRF'
End if
End for
End

Algorithm 2 Kendall Tau Correlation Coefficient-based feature selection (project ID, recipient, implementing agency, implementation start and completion time, geographical location, recommendation for aggregates, funding agency, receiving agency, project implementation score, loan detail score)

As given in the above Kendall Tau Correlation Coefficient-based feature selection algorithm, the objective remains to select the relevant features computationally efficiently. With this objective, first, the features in consideration are obtained for which the Kendall Correlation

Coefficient is applied and evaluated. Next, with the highest correlated features remain the features of highly relevant and selected features, whereas the lesser correlated features remain of less relevance and hence discarded from further processing. Therefore by selecting the highly correlated features, computationally efficient relevant features are selected.

3.2 Temporal-aware Long Short Term Memory-based Financial Predictive analysis model

Finally, with the pre-processed data and relevant features selected, predictive analysis for financial assistance to SMEs are performed by employing LSTMs are a type of network that preserve the memory of previous prediction (i.e., financial pre-processed data of the selected features) and are potential enough for learning long term dependencies (i.e., performing future prediction). In our work, the selected relevant features applied Temporal-aware Lon Short Term, the Memory-based Financial Predictive analysis model.

The advantage of Temporal-aware LSTM remains in initially predicting the value and then storing it to be utilised in addition to the next value of input for further processing based on the temporal factors. Therefore, the training process of the Temporal-aware Long Short Term, Memory-based Financial Predictive analysis model can better learn the temporal relationship between financial data and also maintain stronger robustness. Figure 4 shows the structure of the Temporal-aware Long Short Term, Memory-based Financial Predictive analysis model.

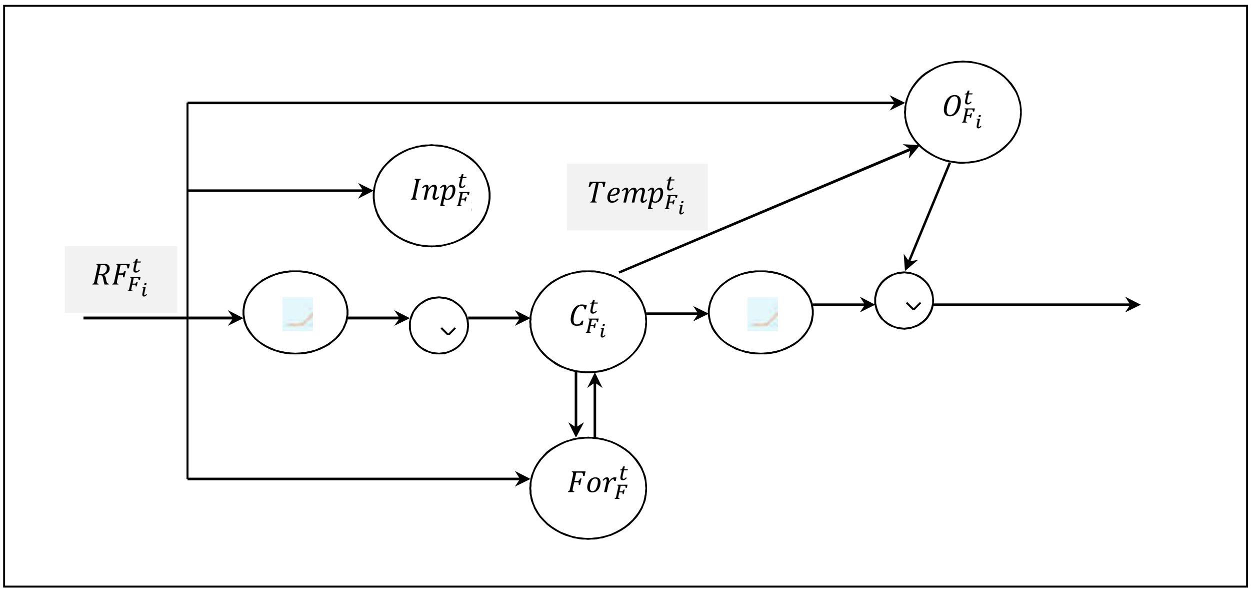

Figure 4 Block diagram of Temporal-aware Long Short Term Memory-based Financial Predictive analysis model

As illustrated in the above figure, Temporal-aware Long Short Term Memory-based Financial Predictive analysis performs the entire function by utilising 4 distinct stages initial prediction '$Inp_{F_i}^t$', ignoring, forgetting '$For_{F_i}^t$' obtaining cell state '$C_{F_i}^t$' and output '$O_{F_i}^t$' respectively. Given a project 'F_i' The temporal value is mathematically formulated as given below to be allocated for financial assistance by the institution.

$$Temp_{F_i}^t = \emptyset \left(RF_{F_i}^t, e_{F_i,F_j}^t \right) \tag{10}$$

From the above equation (10), the temporal value '$Temp$' for the corresponding feature (i.e., project) 'F_i' at the time 't' is measured based on the relevant features '$RF_{F_i}^t$' at the time 't' and the edges or features 'e_{F_i,F_j}^t' at the time 't' respectively.

$$Inp_{F_i}^t = \sigma \left(W_{\varphi Inp} \left[H_{F_i}^{t-1}, Temp_{F_i}^t \right] \right) \tag{11}$$

$$For_{F_i}^t = \sigma \left(W_{\varphi For} \left[H_{F_i}^{t-1}, Temp_{F_i}^t \right] \right) \tag{12}$$

$$C_{F_i}^t = For_{F_i}^t \odot C_{F_i}^{t-1} + Inp_{F_i}^t \odot tanh \left(W_{\varphi C} \left[H_{F_i}^{t-1}, Temp_{F_i}^t \right] \right) \tag{13}$$

$$O_{F_i}^t = \sigma \left(W_{\varphi O} \left[H_{F_i}^{t-1}, Temp_{F_i}^t \right] \right) \tag{14}$$

$$H_{F_i}^t = O_{F_i}^t \odot \tanh\left(C_{F_i}^t \right) \tag{15}$$

From the above equations (11), (12), (13), (14) and (15), '$\odot$' represent the element-wise multiplication, '$W_{\varphi Inp}$', '$W_{\varphi For}$', '$W_{\varphi C}$' and '$W_{\varphi O}$' denoting the weight of the input gate, forget gate, cell state and output gate for the previous temporal '$H_{F_i}^{t-1}$' and the current temporal '$H_{F_i}^t$' value respectively. Finally, output in the output unit '$O_{F_i}^t$' provide the predicted results. Below is the pseudo-code representation of Temporal-aware Long Short Term Memory-based Financial Predictive analysis.

Input: Financial Dataset 'DS', Features '$F = F_1, F_2, \dots, F_n$'
Output: Accurate and precise financial predictive analysis
1: **Begin** 2: **For** each Financial Dataset 'DS' with pre-processed data 'PD' and relevant features 'RF' 3: Obtain temporal value as in equation (10) 4: **For** each temporal values 5: Estimate input gate as in equation (11) 6: Estimate forget gate and cell state as in equations (12) and (13) 7: Estimate the output gate as in equation (14) 8: **Return** predicted results 9: **End for** 10: **End for** 11: **End**

Algorithm 3 Temporal-aware Long Short Term Memory-based Financial Predictive analysis

As given in the above algorithm, the objective remains in obtaining accurate financial prediction for assistance provided for various projects from the funding agencies. As the disbursal of finance differs from time to time for various projects from the funding agencies, it initially obtained a temporal value. Then, with the obtained temporal value and relevant features as input provided in the input unit, a more fine-grained processing unit is utilised to store in the corresponding cell state, update the pre-processed context data transferred between hidden layers, and control the flow of pre-processed financial data information utilised a gate control unit. Finally, in the output layer, with the relevant features, the conditionality checking for features, project implementation score and loan detail score is made and accordingly provided as output, ensuring accurate and precise financial predictive analysis.

4. Results and Discussion

The proposed Linear Scaling Tau and Temporal-aware Long Short Term Memory-based (LST-TLSTM) Financial Predictive analysis is implemented in the Python environment using Global Chinese Development Finance Dataset. The proposed method is related to the existing predictive models, Multidimensional Attribute Sparse Large Data (MASLD) [1], and support vector machine [2] for predictive financial analysis. The methods are validated using the Global Chinese Development Finance Dataset for a fair comparison. This analysis made the best classifier for predictive financial analysis. The entire dataset is split into training and test data for each validation to improve accuracy. This process is repeated 'n' times or 'n' simulation runs. In this work, '$n = 10$' simulation runs are used to analyse the developed method.

4.1 Performance metrics

Different performance metrics are utilised in the proposed method to predict financial analysis using a Chinese dataset for SMEs. In addition, prediction accuracy had carried out in an experimental evaluation, prediction time, and error rate concerning the number of data points.

Prediction accuracy: The first and foremost parameter involved in predictive financial analysis with big data concern is the financial prediction accuracy or simply prediction accuracy. It has assessed the analysis of prediction accuracy deduces how far the financial aspect for the related parties (i.e., individual or financial distributing agency). Given below has been formulated mathematically.

$$P_{acc} = \sum_{i=1}^{n} \frac{S_{AP}}{S_i} * 100 \quad (16)$$

From the above equation (16), the prediction accuracy, 'P_{acc}' is evaluated based on the samples 'S_i' involved in the simulation and the number of samples accurately predicted, 'S_{AP}'. In terms of percentage measured.

Prediction time: The second predominant metric involved in the predictive financial analysis concerning big data is the prediction time. The prediction time refers to the time consumed in predicting the financial aspects for SMEs and is given below mathematically formulated.

$$P_{time} = \sum_{i=1}^{n} S_i * Time\ [FPA] \quad (17)$$

From the above equation (17), the prediction time, 'P_{time}' is measured based on the samples involved in simulation, 'S_i' and the time consumed in financial predictive analysis '$Time\ [FPA]$'. In terms of milliseconds (ms) measured.

Error rate: While predicting financial analysis for individuals or a project, a significant error occurs by wrongly predicting the project implementation score or loan detail score. Given below, mathematically expressed the error rate in our work.

$$Err = \frac{S_{WP}}{S_C} * 100 \quad (18)$$

From the above equation (18), error 'Err' is measured based on the samples wrongly predicted with the score, 'S_{WP}' to the sample's actual scores. 'S_C'. In terms of percentage (%) measured.

4.2 Performance analysis

The proposed Linear Scaling Tau and the Chinese financial dataset have implemented the Temporal-aware Long Short Term Memory-based (LST-TLSTM) Financial Predictive analysis method. First, the proposed method performs financial prediction according to the projection implementation and loan detail scores. Next, using Linear Scaling-based pre-processing to address class balancing performed the financial data pre-processing for different projects by several funding agencies. Followed by which, Kendall Tau Correlation Coefficient-based feature selection algorithm is applied to the processed data, therefore selecting computationally efficient relevant features. Then the selected features are provided as input to the Temporal-aware Long Short Term Memory-based Financial Predictive analysis algorithm to predict financial assistance provided to the public for the specific project. Finally, evaluate the different performance metrics. The proposed LST-TLSTM method provides comparatively better outcomes in performance metrics for the Chinese financial dataset than the existing methods of MASLD [1] and supports vector machine [4] in terms of prediction accuracy, prediction time and error rate performance results.

4.2.1 Performance results of prediction time

This section provides the performance analysis of prediction time based on the graphical representation. To illustrate, the graphical representation of prediction time has given below in Figure-5.

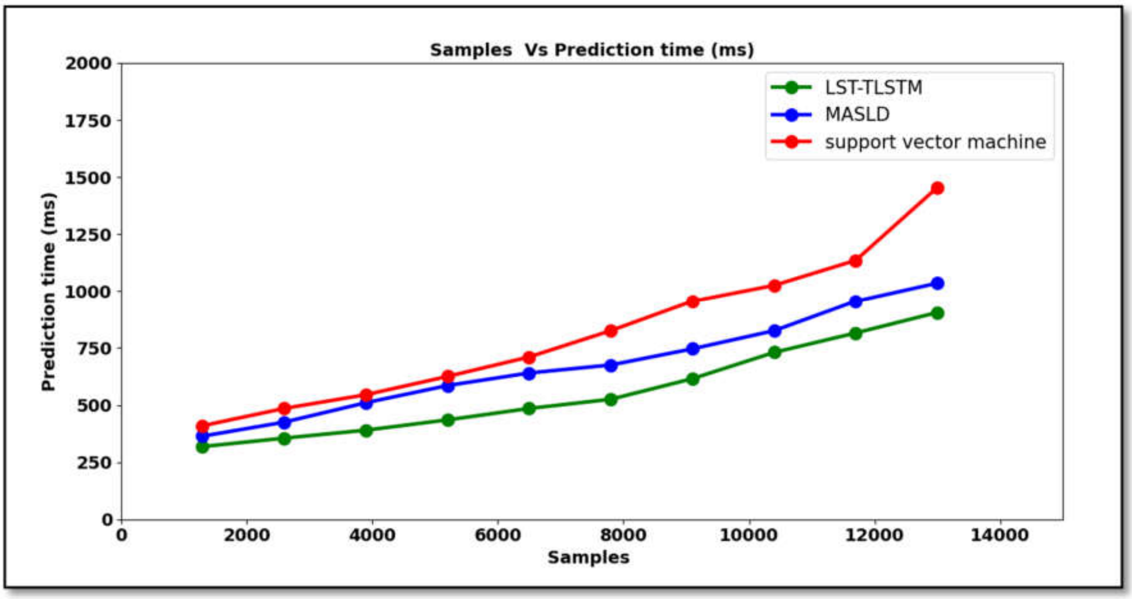

Figure 5 Graphical representation of prediction time

As shown in the above figure, increasing the number of samples involved in the Chinese development financial simulation dataset causes an increase in the predicted samples, increasing the overall prediction time using all the three methods considered for comparison. However, with simulations performed using 1300 samples, the prediction time was 318.5ms using the LST-TLSTM method, 364 ms using [1], and 409.5 ms using [2], respectively. From the results, the prediction time was comparatively lesser using the LST-TLSTM method compared to three other methods. The prediction time improvement was due to the Kendall Tau Correlation Coefficient applied to the processed data for selecting relevant features subject to the highest correlation. As a result, the prediction time using the LST-TLSTM method was reduced by 18% compared to [1] and 31% [2].

4.2.2 Performance results of prediction accuracy

This section provides the performance analysis of prediction accuracy based on the graphical representation. Figure-6 shows the graphical representation of prediction accuracy.

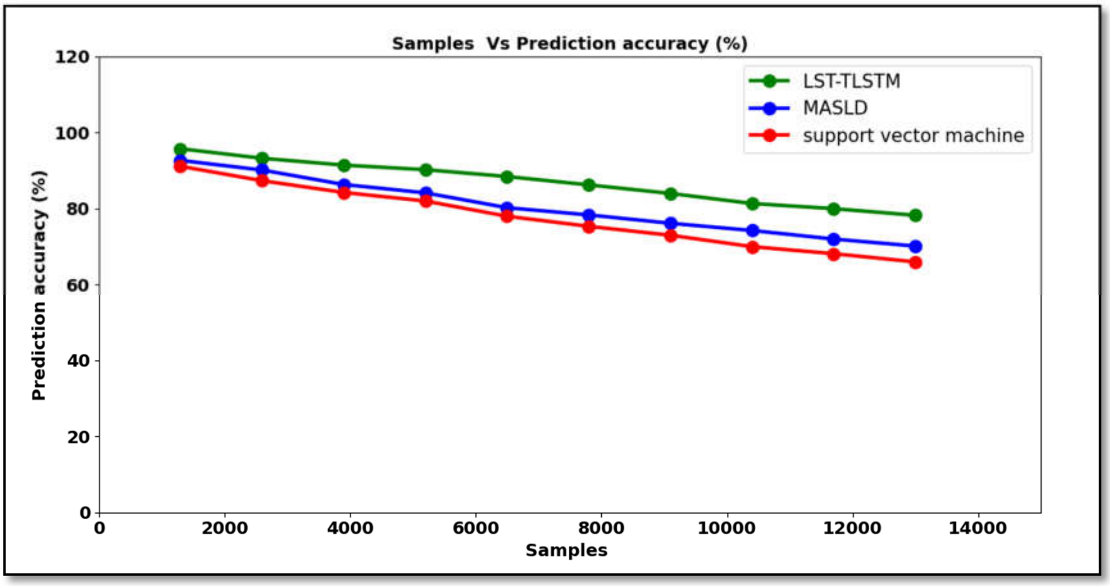

Figure-6 Graphical representation of prediction accuracy

From the above figure, an increase in the number of samples provided as input results in an increase in the samples involved in pre-processing. It, in turn, minimises the prediction accuracy results using all three methods. To be more specific, increasing the samples involved in simulation results minimises the accuracy. However, comparative analysis for 1300 samples shows 1245 samples correctly predicted using LST-TLSTM, 1205 samples using [1] and 1185 samples using [2], therefore resulting in the accuracy of 95.75%, 92.69% [1] and 91.15% [2] respectively. From this result, the prediction accuracy using the LST-TLSTM method was comparatively better than the state of the art methods. The improvement was due to the Linear function and Weighted Scalar function to address class imbalance for the raw Chinese financial data. With this linear and scalar function, class balancing ensured obtaining class balanced processed data. As a result, it improves prediction accuracy using the LST-TLSTM method by 8% compared to [1] and 13% compared to [2].

4.2.3 Performance results of error rate

Figure-7 below illustrates the error rate results of the proposed LST-TLSTM method and the existing two state-of-the-art methods using the Chinese Development Finance dataset.

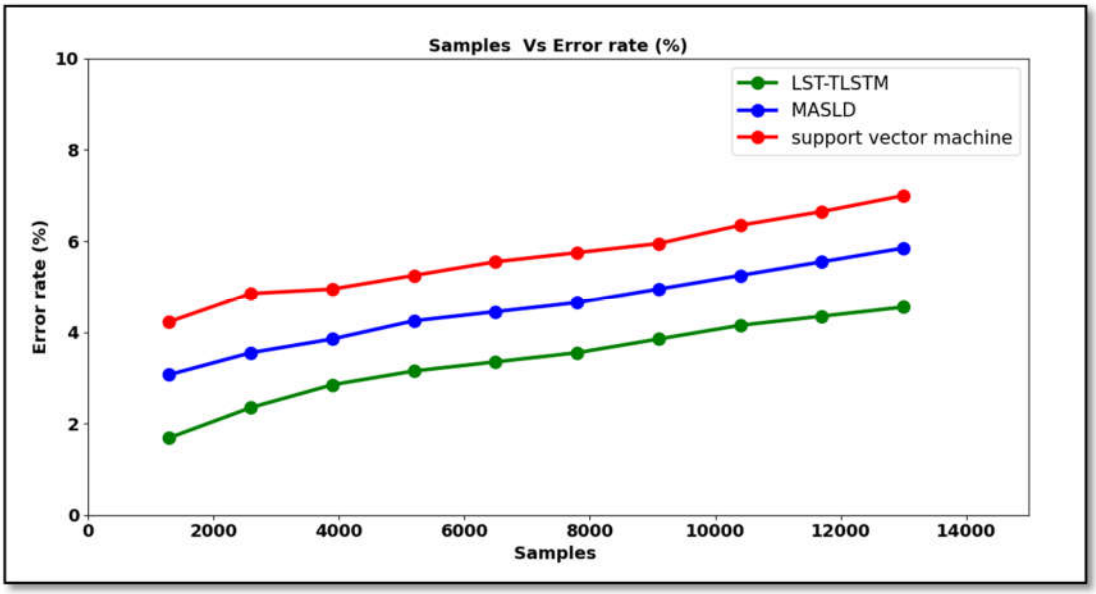

Figure 7 Graphical representation of error rate

From the above figure, the horizontal axis represents samples provided as input at different time intervals for distinct project IDS recommended for aggregates according to the funding agency type. The vertical axis denotes the error rate in terms of percentage. Therefore, the error rate is directly proportional to the projects or samples provided as inferred input in Figure-7. To be more specific, increasing the samples involved in financial assistance causes an increase in the amount of data to be pre-processed, which increases the error rate. However, simulation results found 1.69% error using LST-TLSTM method, 3.07% using [1] and 4.23% using [2]. From this result, the error rate was comparatively lesser using the LST-TLSTM method compared with the state-of-the-art methods. The minimisation of the error rate was due to the application of Temporal-aware Long Short Term Memory-based Financial Predictive analysis for classification purposes with the aid of two distinct features project implementation score and loan detail score, respectively. Moreover, with the temporal factor taken into consideration, the error rate using the LST-TLSTM method was said to be reduced by 27% compared to [1] and 41% compared to [2], respectively.

5. Conclusion

Accurate financial prediction remains the basis for significantly developing the business environment and economic growth. Hence, early accurate and robust prediction of financial aspects provided to individuals by funding agencies is a must as improper prediction results in a severe negative influence on business environments, even resulting in informal credit choice. This work proposes an effective financial prediction method based on Kendall Tau Correlation Coefficient-based feature selection algorithm and Temporal-aware Long Short Term Memory-based Financial Predictive analysis algorithm. The different processes involved in the design are pre-processing, relevant feature selection and robust classification. First, the pre-processing data stage is accomplished by handling imbalanced classes, followed by the normalisation process. Next, the relevant and computationally efficient features are selected using the Kendall Correlation Coefficient with the processed data. Then, it efficiently selects the computationally efficient and relevant features from all the features. Finally, input to the Temporal-aware Long Short Term has provided the selected features, Memory-based Financial Predictive analysis model. Here, a temporal matrix to classify project implementation utilises the loan detail scores. Using the Chinese Development Finance dataset in the Python environment has implemented the proposed LST-TLSTM method. The simulation results validated that the LST-TLSTM method provides better performance metrics like prediction accuracy, prediction time, and error rate than the numerous state-of-the-art methods.

References

[1] JingqiSun, YuLia, QiangLi, YingjiLi, YanshuJia, DongmeiXia, “Fine Clustering Analysis of Internet Financial Credit Investigation Based on Big Data”, Big Data Research, Elsevier, Feb 2022 [Multidimensional Attribute Sparse Large Data (MASLD)]

[2] Rongjun Yang, Lin Yu, Yuanjun Zhao, Hongxin Yu, Guiping Xu, Yiting Wu, Zhengkai Liu, “Big data analytics for financial Market volatility forecast based on support vector machine”, International Journal of Information Management, Elsevier, Feb 2020 (support vector machine)

[3] Rajesh Chidananda Reddy, Biplab Bhattacharjee, Debasisha Mishra, Anandadeep Mandal, “A systematic literature review towards a conceptual framework for enablers and barriers of an

enterprise data science strategy", Information Systems and e-Business Management, Springer, Jan 2022

[4] Efpraxia D. Zamani, Anastasia Griva, Kieran Conboy, "Using Business Analytics for SME Business Model Transformation under Pandemic Time Pressure", Information Systems Frontiers, Springer, Mar 2022

[5] Akhter Mohiuddin Rather, "LSTM-based Deep Learning Model for Stock Prediction and Predictive Optimization Model", EURO Journal on Decision Processes, Elsevier, Oct 2021

[6] Gang Kou, Yong Xu, Yi Peng, Feng Shen, Yang Chen, Kun Chang, Shaomin Kou, "Bankruptcy prediction for SMEs using transactional data and two-stage multiobjective feature selection", Decision Support Systems, Elsevier, Oct 2020

[7] Qixiang Shao, Runlong Yu, Hongke Zhao, Chunli Liu_, Mengyi Zhang, Hongmei Song, and Qi Liu, "Toward Intelligent Financial Advisors for Identifying Potential Clients: A Multitask Perspective", Big Data Mining and Analytics, Mar 2022

[8] Hangjun Zhou, Guang Sun, Sha Fu, Jing Liu, Xingxing Zhou, Jieyu Zhou, "A Big Data Mining Approach of PSO-Based BP Neural Network for Financial Risk Management With IoT", IEEE Access, Nov 2019

[9] Wen Zhang, Shaoshan Yan, Jian Li, Xin Tian, Taketoshi Yoshida, "Credit risk prediction of SMEs in supply chain finance by fusing demographic and behavioral data", Transportation Research Part E, Elsevier, Jan 2022

[10] Chang Yina, Cuiqing Jianga, Hemant K. Jainb, Zhao Wanga, "Evaluating the credit risk of SMEs using legal judgments", Decision Support Systems, Elsevier, Jul 2020

[11] Dong Xiang, Yuming Zhang and Andrew C. Worthington, "Determinants of the Use of Fintech Finance Among Chinese Small and Medium-Sized Enterprises", IEEE Transactions on Engineering Management, Jul 2020

[12] Jian Yu, Jiasha Fu, "Credit rationing, innovation, and productivity: Evidence from small- and medium-sized enterprises in China", Economic Modelling, Elsevier, Feb 2021

[13] Qi Tang, Ruchen Shi, Tongmei Fan, Yidan M, and Jingyan Huang, "Prediction of Financial Time Series Based on LSTM Using Wavelet Transform and Singular Spectrum Analysis", Mathematical Problems in Engineering, Hindawi, Jun 2021

[14] Fernando G. D. C. Ferreira, Amir H. Gandomi, Rodrigo T. N. Cardoso, "Artificial Intelligence Applied to Stock Market Trading: A Review", IEEE Access, Feb 2021

[15] Jian Huang, Junyi Chai and Stella Cho, "Deep learning in finance and banking: A literature review and classification", Frontiers of Business Research in China, Springer, Oct 2020

[16] Meiyi Liang, "Optimisation of Quantitative Financial Data Analysis System Based on Deep Learning", Complexity, Wiley, Mar 2021

[17] Yijie Wang, "Research on Supply Chain Financial Risk Assessment Based on Blockchain and Fuzzy Neural Networks", Wireless Communications and Mobile Computing, Wiley, Feb 2021

[18] Emmanuel Awoin, Peter Appiahene, Frank Gyasi, Abdulai Sabtiwu, "Predicting the Performance of Rural Banks in Ghana Using Machine Learning Approach", Advances in Fuzzy Systems, Hindawi, Feb 2020

[19] Xinchun Liu, "Empirical Analysis of Financial Statement Fraud of Listed Companies Based on Logistic Regression and Random Forest Algorithm", Journal of Mathematics, Hindawi, Dec 2021

[20] Francesco Ciampi, Alessandro Giannozzi, Giacomo Marzi, Edward I. Altman, "Rethinking SME default prediction: a systematic literature review and future perspectives", Scientometrics, Springer, Jan 2021

Data Availability:

Not Applicable.

Data Availability:

The datasets generated and analysed during the current study are available from the corresponding author on responsible request.

Funding statement

Not available

Conflict of interest disclosure

Not applicable.

Ethics approval statement

The manuscript in part or in full has not been submitted or published anywhere.

Patient consent statement

Not Applicable

Permission to reproduce material from other sources

Not Applicable

Clinical trial registration

Not Applicable