A Novel Default Risk Prediction and Feature Importance Analysis Technique for Marketplace Lending using Machine Learning

Sana Hassan Imam*, Sebastian Huhn**, Lars Hornuf***
and Rolf Drechsler****

Abstract

Marketplace lending has fundamentally changed the relationship between borrowers and lenders in financial markets. As with many other financial products that have emerged in recent years, internet-based investors may be inexperienced in marketplace lending, highlighting the importance of forecasting default rates and evaluating default features such as the loan amount, interest rates, and FICO score. Potential borrowers on marketplace lending platforms may already have been rejected by banks as too risky to lend to, which amplifies the problem of asymmetric information. This paper proposes a holistic data processing flow for the loan status classification of marketplace lending multivariate time series data by using the Bidirectional Long Short-Term Memory model (BiLSTM) to predict "non-default," "distressed," and "default" loan status, which outperforms conventional techniques. We adopt the SHapely Additive exPlanations (SHAP) and a four-step ahead model, allowing us to extract the most significant features for default risk assessment. Using our approach, lenders and regulators can identify the most relevant features to enhance the default risk assessment method over time in addition to early risk prediction.

Keywords: Default loan prediction, marketplace lending, time series classification, LSTM/BiLSTM; feature importance, SHAP

JEL classifications: D81, G32

---

* Sana Hassan Imam, Department of Informatics, University of Bremen, sanahassan@uni-bremen.de.

** Sebastian Huhn, University of Bremen/DFKI GmbH, 28359 Bremen, Germany, huhn@uni-bremen.de.

*** Lars Hornuf, Faculty of Business and Economics, Technische Universität Dresden, 01069 Dresden, Germany. lars.hornuf@tu-dresden.de.

**** Rolf Drechsler, University of Bremen/DFKI GmbH, 28359 Bremen, Germany, drechsler@uni-bremen.de.

I. Introduction

As part of the platform revolution (*Parker* et al. 2016), marketplace lending has fundamentally changed the relationship between lenders and borrowers in capital markets. Marketplace lending[1] enables borrowers who have been rejected by a bank to obtain funding from a large and diverse crowd of lenders over the internet. In addition, retail investors are now able to become lenders and diversify their portfolios with private and company loans, an activity that has previously been exclusive to banks (*Dorfleitner/Hornuf* et al. 2022). Marketplace lending has nevertheless come with its share of disadvantages since many lenders are less sophisticated than traditional banks when it comes to lending. Lenders on marketplace lending platforms often rely on easy-to-understand risk ratings rather than more sophisticated financial information (*Cumming/Hornuf* 2022). If there are different types of borrowers in the crowdlending market – for example good borrowers who have a high probability of repaying their loans and bad borrowers who have a low probability of repaying their loans – and lenders cannot distinguish between the two based upon the information that is provided on the platform, good borrowers will have to pay unreasonably high interest rates for their loans. According to *Akerlof* (1970) and *Stiglitz/Weiss* (1981), this is because lenders would not be able to charge interest rates commensurate with the quality of borrowers because they cannot distinguish between them. Moreover, lenders would be cautious about lending money to anyone, even good borrowers, because of the risk of bad borrowers defaulting. This would lead to a reduction in the supply of loans, making it difficult for good borrowers to obtain credit and potentially causing market failure. The presence of bad borrowers, or "lemons," reduces the overall volume of the loan market and creates inefficiencies. This type of asymmetric information is at the expense not only of lenders but also of marketplace lending portals, which must simultaneously maximize the deal flows of both borrowers and lenders to run a successful business (*Rochet/Tirole* 2003). If too many marketplace loans default, the portal will ultimately collapse, similar to what recently happened with the platforms Envestio, Grupeer, and Kuetzal[2].

---

[1] Because recently it is not only peers lending money to peers (peer-to-peer lending), but also institutional investors buying loans up on online marketplaces, we refer to the activity under investigation as marketplace lending (see, for example, *Mohammadi/Shafi* (2017) and *Cumming/Hornuf* (2022)).

[2] We refer interested readers to the following website for scams in marketplace lending: https://rethink-p2p.de/envestio-scam/.

Rapid changes in the world economy due to, for example, global pandemics or armed conflicts might quickly put pressure on borrowers and lenders alike (*Wang/Ni* 2020). As a result of the COVID-19 pandemic and the conflict in Ukraine, the global economy has suffered severe shocks. These crises and rising interest rates are likely to result in riskier and ultimately non-performing loans. Risky and distressed loans are a common feature of many financial crises. High-risk loans are problematic because they impair bank balance sheets, depress credit growth, and delay economic recovery (*Ari* et al. 2020). If marketplace lending continues to grow at a rapid rate, particularly servicing high-risk loans, this segment of the credit market could become a systematic risk (*Käfer* 2018), for example when higher central bank interest rates increase moral hazard and default rates in the credit market (*Diamond* 1984). In order to foresee systematic risks in financial markets, it is important to predict not only loans in default but also those that are currently under stress. Timely detection and resolution of risky loans is difficult but essential for economic recovery and for avoiding financial crises (*Ari* et al. 2020).

Moreover, marketplace lenders might place their capital at undue risk if lending dynamics are poorly understood and loans consequently mispriced. Distinguishing risky loans from defaulting loans and identifying the features affecting loan status can be meaningful from the lenders' and regulator's perspective alike (*Duan* 2019). Consequently, it is important to investigate the features that are most relevant in the prediction of loan default and their contribution to assessing it. Given that the existing assessments for marketplace loan defaults are inadequate and subject to unknown factors (*Xu* et al. 2021), the actual contribution of different features towards the assessment of loan default prediction is essential.

Several studies have evaluated classification methods for loan default prediction. *Xu* et al. (2021) compare different classification methods such as random forests (RF), extreme gradient boosting trees, gradient boosting models, and neural networks, and conclude that RF is superior to other classification methods. However, neural networks constitute a competitive candidate. *Zhu* et al. (2019) evaluate classification methods such as RF, support vector machines, decision trees, neural networks, and logistic regressions (LR) for marketplace lending loan default prediction and conclude that RF performs best. *Li/Chen* (2020), *Kumar* et al. (2016), and *Song* et al. (2020) test different ensemble classification methods for loan default prediction and their findings reveal that the performance of ensemble learning is better than that of individual learning. All these studies investigate which classification method has the highest accuracy but do not focus on the most essential features ensuring the highest accuracy. Previous studies are limited to the classification of only two loan statuses; i. e., "non-default" and "default" and do not consider that "distressed" might serve as an early-warning indicator for systematic stress in the financial system. By stressing

the features and prediction of the "distressed" loan status, our work provides a new perspective and fills this research gap.

To overcome the limitations of specialist knowledge, recent interest in machine learning (ML) for determining the driving factors of repayment failure has risen in the field of marketplace lending. ML techniques improve the veracity and efficiency of decision-making concerning borrower screening by platforms and risk supervision by regulators (*Xu* et al. 2021). Increasingly sophisticated ML methods have helped policymakers to analyze a large amount of data promptly (*Xu* et al. 2021). Machine learning-based classifiers, especially deep neural network-based classifiers, have outperformed the various traditional classifiers in recent research works in the domain of loan default prediction, for example, by *Duan* (2019) and *Lee* et al. (2021). These papers compare the predictive accuracy of different methods by using a certain set of features but fail to determine the best extensible features, that is, features which continue to be relevant over time, and their contribution to accurate prediction and their relevance with time. Our work examines the relevant features, ranked based on their value for predicting and lasting significance. *Xu* et al. (2021) emphasize feature importance and contribution. However, their study considers only a specific data set from China and is limited to a specific kind of data having the feature of borrower video verification provided by the platform. Our approach is novel in terms of identifying the importance of the extensible features and addressing the distressed status prediction. The combination of these two techniques has never been used before.

To address the problem of feature importance along with time dependency, BiLSTM-based multivariate time series (MTS) classification has been applied in this work. The multivariate data set is converted into MTS to assess the extensibility of the features concerning time. To the best of our knowledge, this idea is novel, and the LSTM/BiLSTM model is implemented for the very first time in marketplace lending loan default classification. A SHapely Additive exPlanations (SHAP) model for deep neural networks (DNNs) is implemented in this article for the first time to extract the importance and extensibility of the features from this marketplace lending MTS data set. *Wang/Ni* (2020) address the sequential marketplace loan data as univariate and MTS, and find LSTM to be the most promising candidate for the default rate aggregation problem for marketplace lending sequential data, but only for the prediction of default rate as an aggregation problem. Similarly, *Liang/Cai* (2020) applies the LSTM/Gated Recurrent Unit network-based model for the loan default rate prediction for the marketplace lending sequential data. *Wang/Ni* (2020) and *Liang/Cai* (2020) conclude that LSTM is one of the most promising candidates for loan default prediction. Both of these papers performed the "default rate prediction" as an aggregated monthly rate. To the best of our knowledge, this is the first loan default prediction for individual loan classification using LSTM and BiLSTM models,

while the SHAP model for DNNs is used for the first time for feature ranking of the MTS marketplace lending data set.

The remainder of this article is structured as follows: in Section II, we describe marketplace lending in general and the LendingClub platform in particular. In Section III, we provide a review of the literature. Then, we describe the model in Section IV, and Section V outlines the empirical results and presents robustness checks. Section VI concludes.

II. Institutional Background

This section presents an overview of the marketplace lending process on LendingClub. Marketplace lending has become an established form of lending to individuals or companies without the direct involvement of a commercial bank. Marketplace lending utilizes online platforms that match potential borrowers with lenders. Due to lower intermediation costs as a result of digitalization and the use of big data analytics, the platforms promise lower interest rates for borrowers and higher returns for lenders compared to commercial banks[3]. The lending process begins when potential borrowers submit their loan application on the marketplace lending platform, which acts exclusively as an intermediary between the potential borrower and the lender but does not extend the loan as such. The platform determines the loan interest according to the creditworthiness of the potential borrower and the loan amount. The platform usually classifies the loan based on a risk assessment into seven classes ranging from A to G (*Zhu* et al. 2019). If enough lenders are willing to fund the loan, the online platform informs a partner bank, which then legally extends a loan that is immediately sold to lenders. The partner bank also wires the loan amount to the borrower's account. The marketplace lending platform manages the process as an intermediary and collects its fees from both the lender and the borrower. The borrower repays the loan to the marketplace lending platform, which subsequently returns it to the lenders. Figure 1 provides an overview of how the marketplace lending process works.

Until recently, LendingClub was one of the largest marketplace lending platforms in the world. The platform is representative of the marketplace lending business model worldwide and has provided loan data to researchers[4] that has been used in previous empirical studies (e.g., *Tang* 2019). LendingClub was founded in 2006 and brokered loans over the internet, allowing borrowers to obtain unsecured personal loans ranging from USD 1,000 to USD 40,000. The

---

[3] We refer interested readers to the following website for details: https://bit.ly/3TGwzTM.

[4] We refer interested readers to the following website for LendingClub details: https://bit.ly/3cNa4M8.
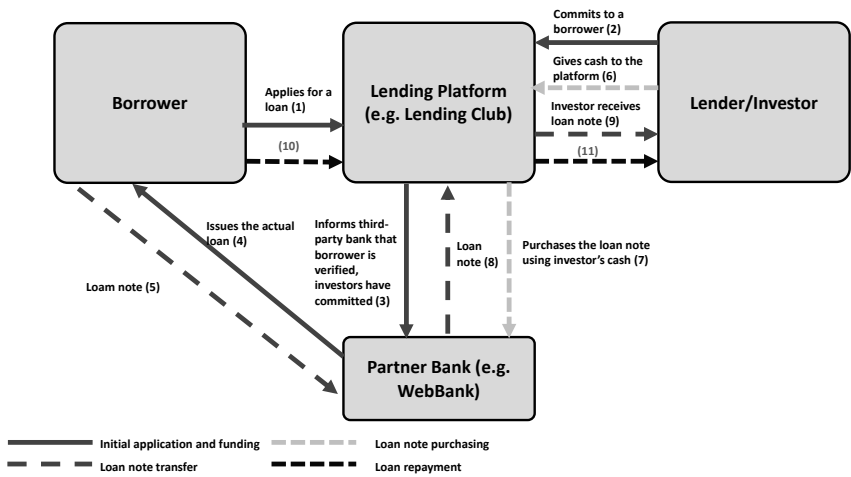
Source: Bakker (2015).

Figure 1: Marketplace Lending Process

usual loan period in our sample was three years. Lenders could browse the loan listings on the LendingClub website and select loans to invest in based on the information provided about the borrower, loan size, loan quality, and loan purpose. After investing, lenders earned an annuity payment from the interest on these loans. LendingClub made money by charging borrowers an origination fee and investors a service fee. Over time, LendingClub experienced increasing difficulty in attracting lenders, in 2020, acquired Radius Bank, and finally announced that it will ultimately shut down its marketplace lending platform. However, to the best of our knowledge, LendingClub continues to operate the marketplace model alongside an underwriting business. According to Lending-Club, the marketplace model primarily helps to gain size in a bull market, while the underwriting business serves as a hedge in a bear market[5].

III. Literature Review

1. Machine Learning

This section presents an overview of the ML techniques implemented in this article: recurrent neural networks, LSTM, BiLSTM, SHAP, and principal component analysis.

---

[5] See for example: bit.ly/3KqWBZ9.

a) Recurrent Neural Network

The conventional feed-forward artificial neural network (ANN) comprises an input layer, an output layer, and several hidden layers. The input layer accepts the input, the hidden layers process it, and the output layer produces the results using an activation function (*Abiodun* et al., 2019). Typically, ANNs demonstrate only limited performance on sequential data such as time series because the connections within a network are set between layers only. The recurrent neural network (RNN) is an extension of the ANN with the ability to manage variable-length sequential input. In RNNs, the connection is between the layers and also with the neurons within the layers, which allows for remembering the past. The disadvantage of RNNs is the vanishing and exploding gradient problem. LSTM solves this problem by introducing another cell state along with input and output gates using an additional gate (the forget gate) (*Sherstinsky* 2020).

b) Long Short-Term Memory Models

LSTM-based models are an extension of RNNs, usually implemented for time series classification or regression problems. LSTM models are designed to preserve and learn the long-term dependencies of the inputs. Furthermore, they have the ability to remember the information over an extended period, with the ability to decide between preserving or ignoring the memory information. An LSTM model captures important input features and preserves this information over a long period of time. The decision of keeping or forgetting the information is based on the weights assigned during the training process and through the gates. An LSTM model typically consists of three gates: the forget, input, and output gates.

The forget gate makes the decision of keeping or eliminating the existing information, the input gate specifies the duration for which the new information will be kept in the memory, and the output gate controls whether the preserved value in the cell contributes to the output. These models are able to address the vanishing gradient problem with the help of the forget gate. Figure 2 shows the LSTM structure unit. A drop layer follows the LSTM model to avoid the over-fitting problem, and a dense layer is followed by an output layer. The details are discussed in the subsection IV.7. *Hyper-parameters Selection*.
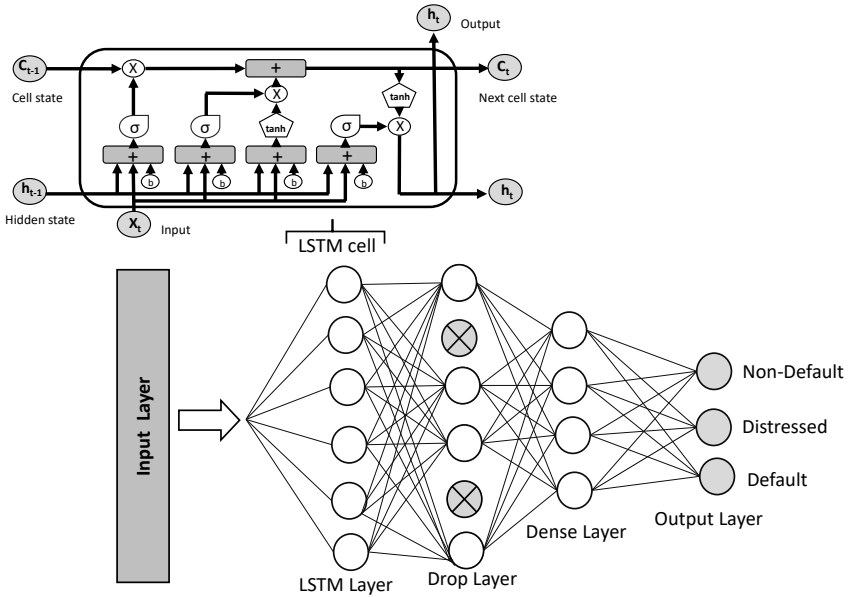
Figure 2: Structure of LSTM

The architecture of the LSTM can be represented by the following six formulas:

$$Input_{gate}\,(i_t) = \sigma(W_i X_t + R_i h_{t-1} + b_i)$$

$$forget_{gate}\,(f_t) = \sigma(W_f X_t + R_f h_{t-1} + b_f)$$

$$Cell_{state}\,(\tilde{C}_t) = tanh(W_c X_t + R_c h_{t-1} + b_c)$$

$$Output_{gate}\,(O_t) = \sigma(W_o X_t + R_o h_{t-1} + b_o)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

$$h_t = O_t \odot tanh(C_t)$$

where σ represents the gate activation function, which is Sigmoid, *tanh* is a hyperbolic tangent function, $\odot$ represents element-wise multiplication. W_i, W_f, W_c and W_o are input weight matrices. R_i, R_f, R_c and R_o are the recurrent weight matrices. X_t is the input, h_t is the current block output and h_{t-1} is the previous block

output, while C_t and C_{t-1} are the memory block of the current and previous LSTM unit, respectively.b_i, b_f, b_c, and b_o are the bias vectors of the input, forget gate, cell state, and output (*Liang/Cai* 2020).

c) Deep Bidirectional LSTMs

Deep BiLSTMs are an extension of the LSTM models in which two models are applied to the input data. At first, a model is applied to the input sequence, i.e., the forward layer. In the second step, the reverse form of the input sequence is fed into the LSTM model, i.e., the backward layer. Applying the model twice leads to improved learning of long-term dependencies and will consequently improve the accuracy of the model (*Siami-Namini* et al. 2019). The architecture used in this article for the implementation of BiLSTM is shown in Figure 3, where three BiLSTM models are implemented, followed by a dense layer and an output layer. Details of all the hyper-parameters and their selection are discussed in the subsection IV.7. *Hyper-parameters Selection*.
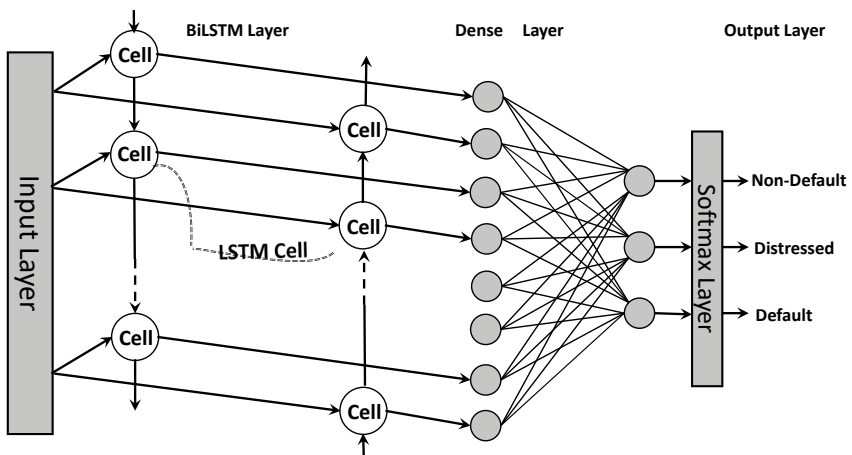
Figure 3: Architecture of BiLSTM

d) SHAP for Feature Importance

SHAP is a state-of-the-art technique in the ML domain for deriving an explainability basis for ML models. SHAP values are determined whenever a certain understanding of complex prediction models, such as a decision tree or a deep neural network, is required (*Chromik* 2020). The contribution of every one of the individual features, as used for the prediction, is covered by

SHAP[6]. SHAP offers a local explanation and a global explanation of the model. More precisely, the "Deep Explainer" of the SHAP model implemented in this article is designed specifically for the model explanation of the deep neural network. The one-step-ahead approach for checking feature importance and extensibility is described in subsection V.1. *Feature Importance by SHAP*.

e) Principal Component Analysis for Feature Extraction

Principal components analysis (PCA) is a well-known approach for the dimensionality reduction of big data. In this method, eigenvalues of the covariance matrix are calculated to explore the data's principal components (PCs). PCs are the eigenvectors and are designed such that the first PC accounts for the largest possible variance in the data set. PCA finds orthogonal linear combinations of the original variables, of which a smaller number of variables explain most of the variability among the original variables (*Bartholomew* 2010). The importance of each variable/feature is reflected by the magnitude of the corresponding values in the eigenvectors.

2. Marketplace Lending

This section presents the marketplace lending literature, particularly those articles dealing with loan default rates and ML techniques. In a seminal study, *Lin* et al. (2013) explore the lending process on Prosper and discover that the online friendship networks of borrowers serve as signals of credit quality to lenders. Friendship networks increase the chances that a loan will be funded, decrease the interest rates, and correlate with lower default rates. *Iyer* et al. (2016) also studied data from Prosper and investigated the impact of soft factors such as characteristics of the listing text or friend endorsements on loan performance. They find that lenders are 45 % more accurate in predicting an individual's default probability compared to the borrower's credit score. However, lenders do not only consider soft factors when making lending decisions. *Herzenstein* et al. (2011) use Prosper data to examine the role of verifiable hard factors in funding decisions. They find that more trustworthy or successful borrowers have higher chances of funding success but poorer loan performance. *Serrano-Cinca* et al. (2015) conducted a study on the performance of LendingClub loans. Their findings show that the credit grade assigned by the platform is the most significant indicator of default. However, the researchers discovered that incorporating the borrower's debt level and other factors into the analysis can improve the accuracy of the prediction.

---

[6] We refer interested readers to the following website for the details of the SHAP model: https://shap.readthedocs.io/en/latest/index.html.

Studies have also looked at the impact of borrowers' physical appearance, gender, age, and race in marketplace lending, with varying results. While some researchers find that female borrowers have higher chances of funding success and pay lower interest rates (*Duarte* et al. 2012; *Ravina* 2019; *Pope/Sydnor* 2011), others find no evidence to support that female borrowers have a higher probability of funding success (*Barasinska/Schäfer* 2014). The role of the loan description has also been studied, with researchers finding that a well-written description with shorter sentences increases the chances of funding success (*Lin* et al. 2013; *Dorfleitner*, *Priberny* et al. 2016). The design of the marketplace lending platform has also been shown to impact the likelihood of funding, with posted price regimes resulting in a higher likelihood of funding, but also higher default rates (*Wei/Lin* 2017).

The studies discussed so far have primarily used the methods of classic inferential statistics to predict funding success and loan defaults. Predicting default risk using ML techniques has been of research interest to finance and computer science scholars alike. Several feature-selection algorithms are proposed to minimize the uncertainty of the loan default risk, and recently developed classifiers are applied to classify the default loan with higher accuracy. The respective research can be categorized into three different areas based on their contribution to the default risk prediction for marketplace lending data:

Number of loan status classes:[7] Binary classification is performed in most papers to predict loan default. Two loan status classes, "non-default" and "default", are frequently used in the literature for the classification of individual borrowers. *Duan* (2019) and *Jin/Zhu* (2015) consider "late payment for less than 120 days" as a risky or distressed class.

Feature Selection: For feature selection and importance assessment, several techniques are adopted. *Xu* et al. (2021) identify RF as the best classifier, while the neural network also has comparable performance. In this case, SHAP has been used for deriving the feature importance graphs of various samples. *Mokhtari* et al. (2019) conclude that the SHAP deep explainer has proven itself as effective for explaining the time series feature importance. This explanation model is further applied to reveal the specific features triggering each observation class.

Classifier Types: Teply/Polena (2020) show that logistic regression, artificial neural networks, and linear discriminant analysis are the three most effective algorithms when processing LendingClub data. *Aleksandrova* (2021) compares the deep learning ensemble classifiers and the traditional classifiers. The results clearly show that the deep learning models result in a notable increase in the

---

[7] Please note that "classes" refers to loan statuses; i. e., "non-default," "default," and "distressed."

specificity score of the evaluation matrix, i.e., more than 89 % of the defaulted loans are correctly classified. Ensemble classifiers outperform single classifiers, and XGBoost is declared the best classifier. *Dzik-Walczak/Heba* (2021) use an ensembled model that is a combination of logistic regression and neural network by averaging the probabilities obtained from both models and achieves higher accuracy than any other available model. *Duan* (2019) invokes DNNs for effective risk assessment and prediction, concluding that the performance of the proposed Multi-Layer Perceptron (MLPs) structure outperforms conventional logistic regression approaches. *Kim/Cho* (2019) propose a deep dense convolutional network for repayment prediction, which is automatically extracted for important low and high-level features simultaneously.

Liang/Cai (2020) suggest an aggregated default rate prediction using LSTM time series, stating that other research has failed to extract time series features from data. The same problem is highlighted in detail by the authors of *Wang/Ni* (2020), who state that the LSTM approach reveals its great potential by outperforming traditionally utilized time series models. *Wang/Ni* (2020) convert the marketplace lending data into univariate as well as MTS but, specifically for the prediction of default rate, as an aggregation problem.

Marketplace data is converted into sequential time series data, and default rate forecasting as a regression problem is implemented for this MTS data set (*Wang/Ni*, 2020). In this article, we have performed the default risk prediction for this MTS data set as a classification problem, which has the potential to outperform the different existing classification techniques. The "distressed" class is introduced in this study because it is critical to identify the high stress related to this risky class with the potential of default. Since around a third of the loans in our sample have a term of 60 months, even a maximum delay of 120 days in loan repayment can result in a relatively high present value compared to the loan amount. Along with the classification of default and distressed loan statuses, this article identifies those sets of features that can extend over time. For assessing the feature's extensibility over the time span, the problem is converted into an MTS and both LSTM and BiLSTM are invoked to predict and classify the loan defaults into three classes: "non-default," "distressed," and "default." The proposed approach applies the SHAP for DNNs model to extract the most important features of the marketplace lending MTS data set. By training the BiLSTM and SHAP models with the one-step-ahead approach and repeating the process four steps ahead, features that remain significant over time are extracted.

IV. Model

Figure 4 describes the workflow of the proposed method in detail, which allows for evaluating feature extensibility and importance in loan default predic-
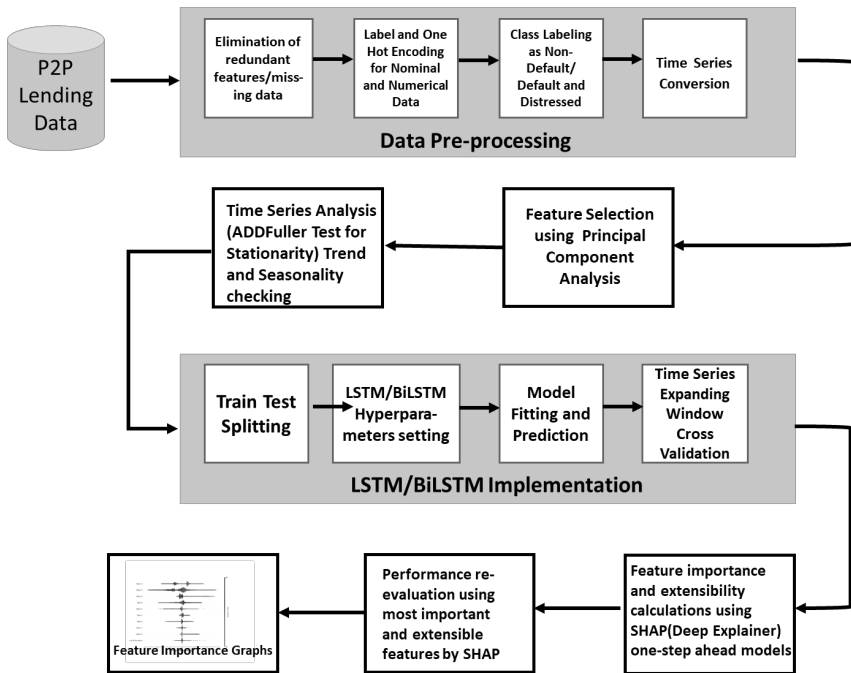
Figure 4: Proposed Methodology

tion. The proposed algorithm starts with data pre-processing and converting the marketplace lending data into time series data by adding the time-stamped index. The unit of observation in the data is the loan. The time dimension is the loan term. Various events can occur over the course of the loan term period. The loan pays back the annuities, the loan is in distress, or the loan is in default or has expired. The periodicity of the time series is quarterly, as the data from LendingClub are only available in this periodicity. In the next step, we apply an Augmented Dickey-Fuller (ADF) test to check the stationarity of the time series and perform the feature extraction by using PCA. The deep neural network models LSTM and BiLSTM are implemented for the default loan prediction. On the basis of the trained model, a SHAP model is invoked for extracting the most important features from this time series to visualize the features' extensibility using the one-step-ahead approach. The details of each block are discussed in the following subsections.

1. Data Collection and Pre-processing

The analysis is conducted on the data from LendingClub. The data spans the years 2016 to 2020 and comprises 2,009,689 observations with 150 variables, representing a total loan amount of USD 31.45 billion. At first, the attributes containing the string values are identified and all strings are converted to numerical or categorical values based on the data requirements. Analogously, the duplicate and meaningless attributes are deleted. The missing values are filled in for the remaining attributes by applying mode interpolation. Table A1 shows the eliminated attributes based on their types and reasons for elimination. Class labeling is performed after the elimination and conversion of the data are completed. The marketplace lending loan default prediction problem is considered as a binary problem, but an important aspect has been ignored in previous research. If loan repayments are delayed for at least 15 days, the loan status switches to "distressed." The following classes are assigned based on the loan status of the borrower: "non-default," "distressed," and "default." Three classes and their descriptions are given in Table 1 and Table 2.

2. Features Elimination

Marketplace lending data sets generally span a high dimensional space, clearly contributing to a high level of complexity as soon as classification techniques are applied. This results in a significant decrease in accuracy in marketplace loan default prediction. Thus, it is crucial to reduce the number of dimensions of marketplace lending data. Feature selection is a frequently used technique for dimensionality reduction and accuracy improvement by extracting the most vital set of features. Further techniques have been adopted in previous studies, such as *Zhu* et al. (2019) including recursive feature elimination and the Pearson coefficient correlation method. In this article, the features are reduced from 99 to 22 using the PCA method and specific domain knowledge. PCA calculates the principal components having the largest eigenvalues (as described in the subsection III.1.e) *Principal Component Analysis for Feature Extraction*) by detecting the list of features with their largest contributions towards the principal components and by using the domain knowledge of the Gini coefficient test score (of each feature calculated by *Dzik-Walczak/Heba* (2021)), the most important features are selected. The following features have furthermore been deliberately selected for the prediction of loan default and evaluating the feature importance and extensibility. "loan amount (€)," "installment (€)," "grade," "application type," "funded amount (€)," "open account," "total payment (€)," "total received interest (€)," "term 36 months," "fico range high," "annual income (€)," "employment length of borrower," "delinquency in 2 years," "home ownership mortgage," "purpose car," "dti," "purpose renewable energy," "purpose small

business," "total received principal (€)," "outstanding principal by investor (€)," and "interest rate." Table 1 provides detailed descriptions of each feature. Appendix Table A2 provides the descriptive statistics of the selected features.

Table 1

Selected Features with Description

Feature Name	Description	Type
Loan amount (€)	The listed amount of the loan applied by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.	Numeric
Installment (€)	The monthly payment owed by the borrower if the loan originates.	Numeric
Grade	LendingClub-assigned loan grade.	Numeric
Application type	Indicates whether the loan is an individual application or a joint application with two co-borrowers.	Nominal
Funded amount (€)	The total amount committed to that loan at that point in time.	Numeric
Open account	Number of open credit lines in the borrower's credit file.	Numeric
Total payment (€)	Payments received to date for total amount funded.	Numeric
Total received interest (€)	Interest received to date.	Numeric
Verification status	Indicates if income was verified by LendingClub, not verified, or if the income source was verified.	Nominal
Term	The number of payments on the loan. Values are in months and can be either 36 or 60.	Nominal
Fico range	The boundary range the borrower's FICO at loan origination belongs to. There are two ranges (High and Low) for upper and lower ranges.	Nominal
Annual income (€)	The self-reported annual income provided by the borrower during registration.	Numeric
Employment length of borrower	Employment length of the borrower in years. Possible values are between 0 and 10.	Nominal
Delinquency in 2 years	The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years.	Nominal

(continue next page)

(Table 1 continued)

Feature Name	Description	Type
Home ownership	The homeownership status provided by the borrower during registration or obtained from the credit report. The homeownership status can be rent, own, mortgage, and other.	Nominal
Purpose car or small business or renewable energy	A category provided by the borrower for the loan request.	Nominal
Dti	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.	Numeric
Total received principal (€)	Principal received to date.	Numeric
Outstanding principal by the investor (€)	Remaining outstanding principal for a portion of the total amount funded by an investor.	Numeric
Interest rate	Interest rate on the loan.	Numeric

Table 2

Loan Class Definition

Class Name	Loan Types	Labels
Non-default	Fully Paid/ Non-Default/ Issued	0
Distressed	In Grace Period/ Late (16–30 days)/ Late (31–120 days)	1
Default	Default/ Charged Off	2

3. Time Series Analysis

Time series is a sequence of observations based on discrete occurrences equally spaced in time. The issue date of the loan is used to determine the start date of the time series. During the time series data pre-processing, it is important to check the stationarity property of the time series. If the time series is non-stationary, it is required to first convert it into a stationary time series. For a stationary time series, the mean and variance must be constant with respect to time. For this, an ADF test is performed.

ADF Test for Stationarity

Example Features	ADF-statistics	p-value	Critical Values			Conclusion
			1%	5%	10%	
Loan amount	−6.842	<0.001	−3.679	−2.968	−2.623	Stationary
Installment	−3.241	0.002	−3.711	−2.981	−2.630	Stationary
Grade	−5.248	<0.001	−3.677	−2.962	−2.523	Stationary
Open account	−4.702	0.008	−3.711	−2.976	−2.633	Stationary

Note: A p-value > 0.5 indicates that the test does not reject Ho, i.e. the series is non-stationary, while a p-value < 0.5 indicates that the test accepts H1, i.e. the series is stationary.

According to *Yang/Shahabi* (2005), if all the univariate time series in a multi-variate time series item are stationary, then the MTS item is also stationary. The data set used in this article is also an MTS. The ADF test is applied for each individual univariate variable/feature of the time series. In our data set all the variables are stationary with respect to time. The mean and variance of these univariate variables are constant with respect to time. Some of the variables are presented in Table 3 as examples. Trends and seasonality for each variable are also verified, and it is observed that no specific trend or seasonality associated with any of the selected feature variables exists.

4. Model Generation

The following section illustrates the BiLSTM structure and discusses the LSTM and BiLSTM implementation, the hyper-parameters selection, tuning, and model performance evaluation. Appendix Figure A3 comprises a pseudo-code to explain the adopted algorithm with all the steps and details of implementations of the proposed procedure. In the first block train, test data splitting is applied. In the second block, LSTM/BiLSTM implementation with all hyper-parameter tuning is conducted. With class weight adjustment, cost-sensitive analysis is performed in order to mitigate the data imbalance problem. Feature importance is graphically visualized in the last block with the help of the SHAP library implementation.

5. Model Training and Summary

When the data set is imbalanced, the classifier predicts the majority class primarily, and the predictive accuracy of the minority class declines (*Ye* et al. 2018). There are many methods for handling the problem of data imbalance, such as

under-sampling, over-sampling, and cost-sensitive analysis. Cost is a penalty associated with misclassification. Some research has noted that over-sampling and under-sampling possess significant drawbacks compared to cost-sensitive analysis (*Xia* et al. 2017). In this study, the number of non-defaulters (majority class) is significantly higher than the number of defaulters (minority class). A weighted cost matrix is introduced in the proposed model training phase to reduce the effect of the imbalanced data set on the loan default prediction. This matrix increases the cost of misclassifications associated with defaulters. Therefore, due to the high penalty for misclassification, the classifier strives to make correct predictions of defaulters (minority class) to minimize the cost, hence improving the predictive accuracy for the defaulters class (*Ye* et al. 2018). During the model's training, two parameters are used for mitigating the effect of the imbalanced data set by introducing the penalty of misclassification: class weight and sample weight. The class weights are calculated based on the ratio of the majority class to the minority class. A ratio of 1 to 10 assigns a weight of 10 to the minority (ratio 1 class) and a weight of 1 to the majority class. During the training phase, these weights are multiplied with the loss function to impose a high penalty on the classifier for wrongly classifying a minority class. The aim of the classifier is to reduce the overall cost of misclassification. Similarly, the sample weights are used when weights are assigned for each sample of the whole data set based on the imbalanced data set. Each sample is element-wise multiplied with this corresponding sample weight matrix. The model summary for the training of the proposed model is described in Table 4.

Table 4
Model Summary of LSTM/BiLSTM

Layer type	Output shape	parameter#
Input Layer	(None, None, No. of features)	0
LSTM/BiLSTM	(None,128)	75776
Dropout	(None,128)	0
Dense	(None,2)	258
Output Layer (Dense)	(None,2)	6
Total trainable parameters	76040	

Note: The table reports the shape and number of parameters of input, output, dropout, and LSTM/BiLSTM layers. The table also indicates the total number of trainable parameters used for model training.

6. *Challenges during Training*

Over-fitting and under-fitting problems prevail even after adopting model classification and prediction algorithms. Thus, the over-fitting problem is addressed by the dropout layer and L2 regularization methods. The dropout layer sets the weights of the nodes to zeros at a certain ratio, thereby reducing the effect of the over-fitting problem. The L2 regularization slows down the weight update process, heavily reducing the over-fitting problem. For tackling the under-fitting problem, the learning rate and the number of epochs constitute the most important hyper-parameters that have to be deliberately adjusted to improve the performance of the proposed algorithm. The Adam optimizer[8] is selected following a grid search principle[9] since it performs better than other optimizers for the indented algorithm. Consequently, the updating speed of the inner parameters of the model is determined by the learning rate of the Adam optimizer. The model is able to achieve good performance when an appropriate learning rate is applied. The number of epochs is set to a large number to avoid under-fitting.

7. *Hyper-parameters Selection*

Table 5 presents the complete list of selected hyper-parameters. Adam, RMSprop, ADAdelta, and SDG[10] are the optimizers tested in the training phase, and the Adam optimizer performed the best among them. The learning rates of 0.1, 0.01, 0.001, and 0.0001 are tested and with a 0.001 learning rate, the model is converging at a moderate pace. The dropout rates of 0.05, 0.1, 0.2, and 0.5 are also tested and 0.1 has been selected after several simulations.

8. *Evaluation Matrices for Model Performance*

As opposed to standard evaluation metrics, imbalanced classification problems typically rate errors with minority classes more highly than errors with majority classes. Suppose we have a data set of 100 patients, of which 8 are cancer patients, and the rest of 92 is healthy. In this example, we have an imbalanced data set, as the majority class is about nine times bigger than the minority class.

---

[8] An optimization algorithm that is used in neural networks to update network weights iteratively.

[9] We refer interested readers to the following website for details of grid search technique: https://scikit-learn.org/ stable/modules/grid_search.html.

[10] We refer interested readers to the following website for details of deep learning optimizers: https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-on-deep-learning-optimizers/.

Table 5

Hyper-parameters of LSTM and BiLSTM

Hyper-parameters	LSTM	BiLSTM
Hidden layers	2	3
Nodes	128,128	128,128,128
Optimizer	Adam	Adam
Activation function	SoftMax	SoftMax
Learning rate	0.001	0.001
Epochs	500	500
Dropout rate	0.1	0.1
L2 regularization	Recurrent layers l2(0.01)	Recurrent layers l2(0.01)

Note: The table reports the shape and number of parameters of input, output, dropout, and LSTM/BiLSTM layers. The table also indicates the total number of trainable parameters used for model training.

Here the majority class is "healthy," and the minority class is "cancer." So, it is crucial to detect the 8 cancer patients accurately. False negatives, in this case, would be a life-threatening condition. Sensitivity/Recall is the ability of a test to identify patients with a disease correctly. While specificity is the ability of a test to correctly identify people without the disease or the ability to avoid false positives. These two parameters are important in imbalanced data set classification along with the AUC-ROC curve score.[11] On the basis of the threshold matrix for imbalanced classification, we use the following four evaluation parameters to evaluate the model's characteristics and performance.

Accuracy: Accuracy is the proportion of the correctly predicted labels (true positives (TP) and true negatives (TN)) among all our predictions (TP, TN, false positives (FP) and false negatives (FN)) [12].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Sensitivity/Recall: Recall addresses the question: "Among all the positive labels in the data set, how many of them are correctly captured by the model?"

$$Recall / Sensitivity = \frac{TP}{TP + FN}$$

---

[11] Different types of threshold matrices for imbalanced classification and when to apply them can be found on the following website: https://machinelearningmastery.com/tour-of-evaluation-metrics-for-imbalanced-classification/.

[12] https://tinyurl.com/4ukftyky.

Precision: Precision addresses the question: "Among all the positive labels predicted by the model, how many are indeed positive?"

$$\text{Precision} = \frac{TP}{TP + FP}$$

Specificity: Specificity measures the proportion of correctly identified negative labels among all the negative labels (TN + FP). Specificity also is a key ingredient in the ROC curve.

$$\text{Specificity} = \frac{TN}{TN + FP}$$

9. Data Division and Cross Validation

For time series cross-validation, the validation data set is evaluated on the "future" observations immediately following the training data set in time rather than random observations. In particular, we apply an expanding window cross-validation technique. The expanding window is also referred to as "time series split," in which the training set is divided at each iteration in a way that the validation set is always ahead of the training split (*Assaad/Fayek* 2021). This method is called the nested walk forward or expanding window method, where

Nested Cross-Validation
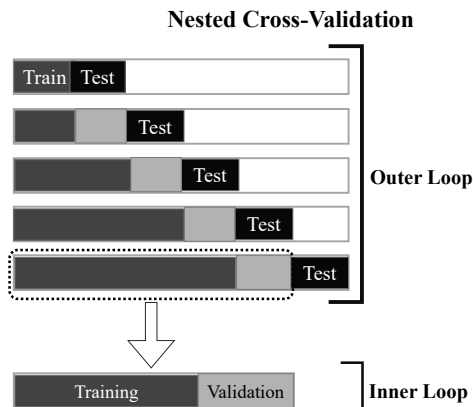
Note: Our study employs an expanding window cross-validation with the data set partitioned into 5 distributed sets, equating to a 5-split iteration procedure. The average accuracy values of the 5 splits were used to evaluate the models' overall performance.

Figure 5: Expanding Window Time Series Cross-Validation

the validation set is also defined inside the training set.[13] Nested cross-validation provides an almost unbiased estimate of the true error (*Varma*/*Simon* 2006). The data set is divided into five-time splits and accuracy is calculated for each split, as shown in Figure 5.

V. Results

The empirical analysis was performed on a SLURM-based HPC cluster with a machine equipped with three Nvidia V100S 32GB, 256GB memory, and 2 AMD EPYC ROME 7272 system. The algorithm is written in Python (v3.7.2) using TensorFlow and Keras (v2.7.0). The LSTM and BiLSTM models are implemented and evaluated against traditional classifiers. The most important step before the implementation of LSTM/BiLSTM is hyper-parameter tuning since this step highly affects the accuracy scores. By applying the grid search approach via maximizing the accuracy of the data set, the values of the hyper-parameters are adjusted, as described in Table 5. The number of nodes is set to 128. The parameter "mini-batch size" is set to 20, with a validation step of 20, following the grid search ranging from 10 to 100, as described in the subsection IV.7. *Hyper-parameters Selection*. The number of steps per epoch is set to 10, and the number of epochs is adjusted to 500, to ensure the algorithm's convergence. Figure 6 and Figure 7 show the training and validation curves for model accuracy
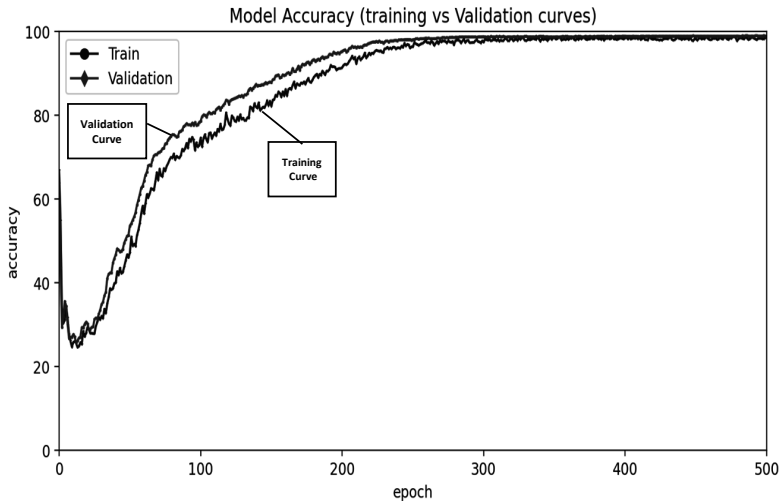

Figure 6: Training and Validation Accuracy Curves

---

[13] We refer interested readers to the following website for details of time-series nested cross-validation: https://www.angioi.com/time-nested-cv-with-sklearn/.
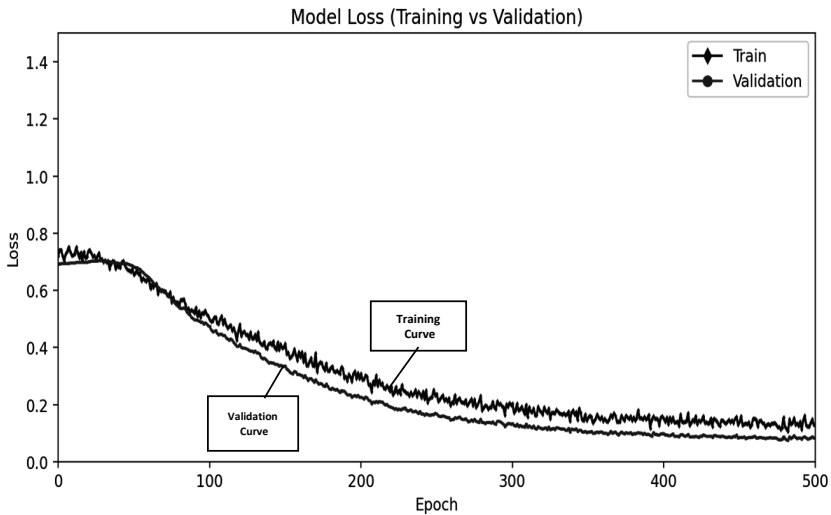
Note: Model accuracy increases while error decreases. Accuracy and error approach a constant value after around 300 iterations.

Figure 7: Training and Validation Loss Curves

and loss function. The loss function is set to be "binary cross entropy" and "categorical cross entropy" for the binary and the multi-class classification, respectively.

The data set is divided into five-time splits and in each split, the algorithm runs for 500 epochs as shown in Table 6. It is observed that with the increase in the training samples, the corresponding accuracy increases in each time split, consequently yielding an accuracy of 93 % for multi-class and 95 % for binary class prediction.

This article focuses on the four evaluation criteria (Recall, Specificity, Precision, and Accuracy) to evaluate and compare the classifiers' performance. Recall (or Sensitivity) is the important evaluation parameter because the concern is how many default/distressed loans are classified correctly by the classifier amongst all the default/distressed loans. LSTM, BiLSTM, RF, ExtraTree classifier, logistic regression, decision tree, and XGBoost classifier are tested for the loan default and distressed classification. Note that the class weight assignment for data balancing is performed in the training parameters of the classifiers. LSTM and BiLSTM perform better than the other classifiers in terms of Precision and Recall for binary as well as for multi-class classification, as illustrated in the confusion matrix for loan default classification presented in Figure 8. Table 7 proves that both LSTM and BiLSTM perform better than all traditional
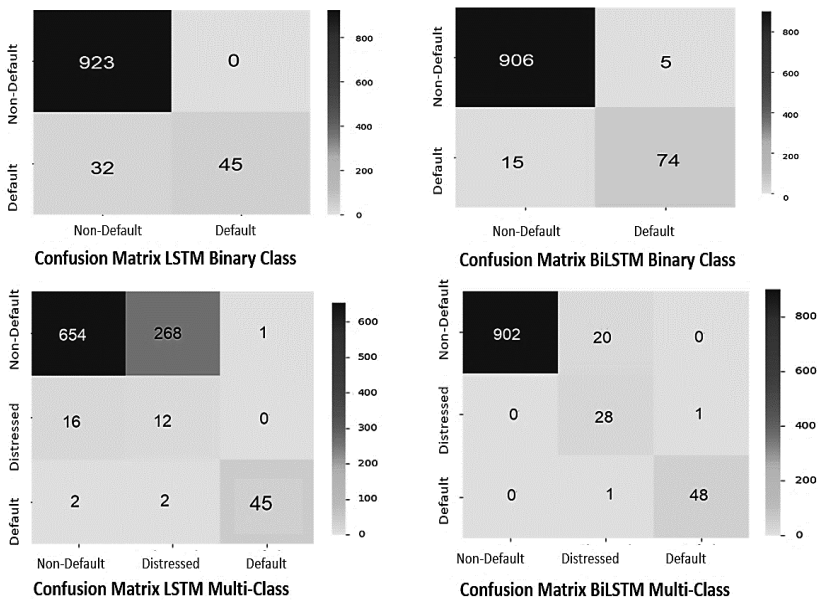
Note: Confusion matrices show the performance of the classifier algorithm on the test data set. The figure shows that the BiLSTM performs best in binary and multi-class classification.

Figure 8: Confusion Matrix LSTM/BiLSTM

classifiers, with an accuracy of 95 % and 93 % for the binary and multi-class loan status prediction.

Figure 9 shows the Receiver operating curve ROC for the BiLSTM algorithm for the classification of three classes. Figure 10 shows the ROC for all tested classifiers (RF, DT, ExtraTree, LR, and XGB classifier) on the LendingClub time series data set.

1. Feature Importance by SHAP

A novel time series one-step-ahead prediction approach is applied for the MTS-based marketplace lending data set. The BiLSTM model has been trained one step ahead, and the SHAP feature importance summary is plotted on this basis. Analogously, SHAP feature importance has been calculated for two-step, three-steps, and four-steps-ahead prediction, as shown in Figure 11. The following 14 features that remain consistently important with one-step-ahead prediction are extracted: "loan amount (€)," "installment (€)," "grade," "funded amount (€)," "total payment (€)," "term 36 months," "interest rate," "verification status ," "employment length of borrower," "outstanding principal by investor (€)," "total
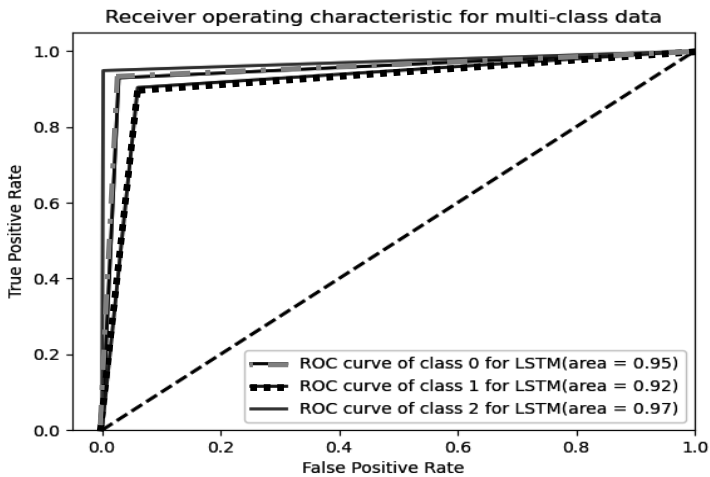
Note: ROC graphs show the performance of the classification model at all classification thresholds. The bigger the area under the curve, the better the classification model is.

Figure 9: ROC Curve for BiLSTM based Classification
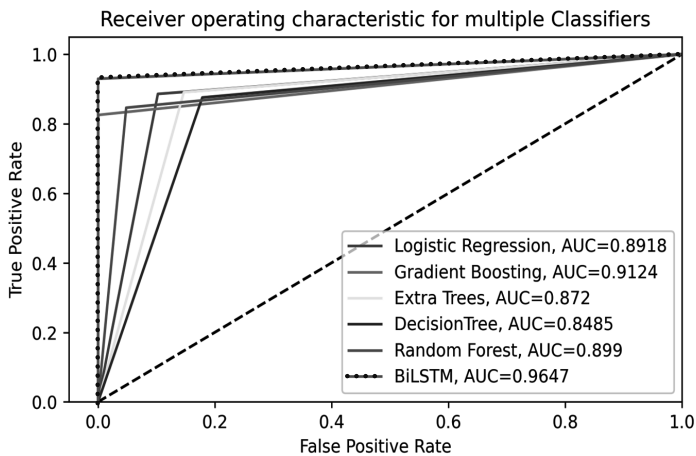
Figure 10: ROC Curves Binary Class Comparison with other Classifiers

received principal (€)," "application type," "fico range high" and "home owner-ship mortgage." The accuracy of all classifiers is re-calculated using the above most important and extensible features. Table 8 shows that the accuracy of LSTM and BiLSTM is improved with SHAP-extracted features as compared to Table 7, where 22 features were initially selected for loan default prediction.

Table 6
Cross-validation Accuracy

Time Split	Epochs/Split	Accuracy %	
		Binary Class	*Multi-Class*
1	500	92.81 (±1.45)	79.72 (± 4.53)
2	500	93.20 (± 1.87)	89.47 (± 1.67)
3	500	92.50 (± 0.03)	86.23 (± 3.83)
4	500	94.67 (± 0.2)	92.31 (± 1.34)
5	500	95.34 (± 0.53)	93.52 (± 0.63)

Note: The table reports the accuracy of the expanding window algorithm for time series cross-validation. Cross-validation is used for training and generalizing the model for predicting unseen future data. Increasing accuracy with an increase in data samples means the model is robust and better trained for future predictions.

Table 7
Classifiers' Performance Evaluation for Binary and Multi-Class Prediction

Classifier	Binary Classification				Multi-Class Classification			
	Accuracy	*Recall*	*Precision*	*Specificity*	*Accuracy*	*Recall*	*Precision*	*Specificity*
LSTM	92	91	88	93	90	80	92	93
BiLSTM	95	92	90	94	93	84	93	90
XGB	91	91	84	91	89	80	90	91
RF	90	60	90	87	84	40	35	86
ExtraTree	78	62	82	78	83	27	26	86
LR	80	72	78	80	68	75	58	71
DT	75	70	75	78	72	76	30	72

Note: The table reports different classifiers' performance evaluations for the period February 2020 to June 2020. The statistics show that the proposed LSTM and BiLSTM models outperform the other classifiers in binary and multi-class classification.

2. The Impact of the Early Covid-19 Period

As a robustness check, we investigate whether the Covid-19 period or the recently rising interest rates in the USA may have had an effect on the model results. For this we divide the data set into a period from 2016 to mid-2019 and mid-2019 to mid-2020. The shares of the three loan level classes have somewhat changed from the former period to the latter. The share of distressed loans increased one percentage point from 1.0% to 2.0%, which is in line with related studies (*Anh* et al. 2021; *Kaveri/Narang* 2020; *Nigmonov/Daradkeh* 2022), and the

Figure 11: Feature Importance Summary by SHAP Deep Explainer for 1–4 Steps Ahead in Time
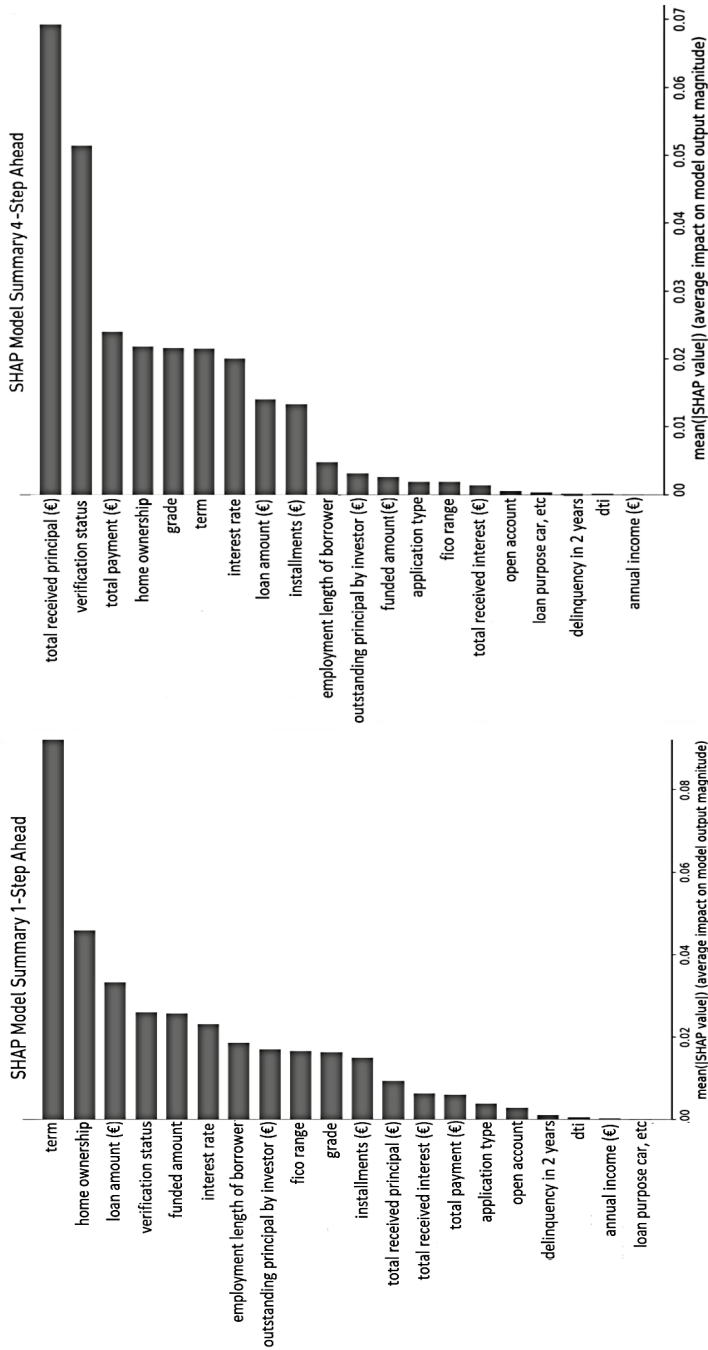
Note: This figure reports the most important features extracted by the SHAP Explainable AI model for the one-step- and four-steps-ahead BiLSTM model.

share of loans in default decreased almost twelve percentage points from 14.3 % to 2.7 %. While one might expect the share of loans in default to increase during the latter period, one should note the following: First, Covid-19 only started to affect the real economy starting in early 2020. Thus, Covid-19 did not yet fully affect the loans in our sample. Moreover, interest rates were rapidly declining in late 2019 and early 2020 in the US again because of the US-China trade war, potentially positively affecting loans. Second, the fact that loans in default have been decreasing in the period from 2019 to mid-2020 is also a statistical artefact. Because the number of new loans has been rapidly increasing on LendingClub, also because conventional banks increasingly rejected borrowers[14], and newly granted loans often do not immediately default, the share of new loans that are not in default relative to older loans that are potentially in default becomes larger and, consequently, the share of loans in default decreases in our sample.

Previous research has shown that in particular the loan amounts, interest rates, and term rates were all affected by coronavirus-related constraints (*Najaf* et al. 2022). Not least because of their non-linear nature, ML techniques are particularly suitable for reacting to such changes, showing a greater efficiency and accuracy than traditional statistical methods

(*Xu* et al. 2021). Given that the share of loan classes has changed somewhat during the validation period, if anything, one would expect the predictions to perform more poorly. However, in early 2020 the test accuracy for our model is generally still high.

Table 8
Classifiers' Performance Evaluation with SHAP Feature

Classifier	Binary Classification				Multi-Class Classification			
	Accuracy	Recall	Precision	Specificity	Accuracy	Recall	Precision	Specificity
LSTM	95	95	90	97	94	92	98	92
BiLSTM	97	97	93	94	95	92	99	93
XGB	95	94	88	92	91	84	99	92
RF	94	65	96	96	88	45	42	95
ExtraTree	87	63	88	88	83	30	30	94
LR	83	78	83	83	72	75	60	60
DT	78	77	80	80	89	81	31	89

Note: Model re-evaluation with feature selection by SHAP for the period from February 2020 to June 2020. The proposed LSTM and BiLSTM models outperform the other classifiers in binary and multi-class classification.

---

[14] *Najaf* et al. (2022) show that marketplace lending has become a viable alternative credit option for many borrowers.

VI. Conclusion

This article develops a general framework for classifying loan statuses, including the default prediction, for marketplace lending multivariate time series data sets by heavily orchestrating techniques such as LSTM/BiLSTM, SHAP, and various other statistical metrics for the required data pre-processing. In addition to loan status classification, the proposed framework identifies the most significant and extensible features using a one- to four-steps-ahead SHAP model. The empirical findings have shown that the proposed framework outperforms other existing techniques, such as logistic regression, random forest, and XGBoost, by achieving an accuracy of 95 %. For the first time, the proposed technique provides an early risk assessment that describes the potentially defaulting borrowers by classifying the "distressed" class. This work enables us to identify the most extensible features, which remain significant over time, by introducing an entirely novel approach by adapting the SHAP model for DNNs. The re-calculation of accuracy based on the SHAP features selection proved to be more promising with 97 % accuracy, forming an ideal foundation for prospective analysis.

The speed at which economic and financial crises are accelerating makes forecasting for credit defaults of personal loans increasingly important. Introducing new features into the current statistical framework, particularly socioeconomic and macroeconomic factors, and assessing their extensibility can be a valuable contribution to future computer science and finance research. However, our study also has significant limitations. While the empirical evidence suggests that the proposed framework outperforms other existing ML techniques, the empirical results may not be externally valid for other marketplace lending platforms. While we take some comfort in the fact that other studies have used very similar features to explain loan defaults across different marketplace lending platforms (*Lin* et al., 2013*; Iyer* et al., 2016), future research could test our framework to predict loan defaults on Prosper, Funding Circle, and other platforms.

References

Abiodun, O. I. et al. (2019): Comprehensive review of artificial neural network applications to pattern recognition. In: IEEE Access 7, pp. 158820 – 158846.

Akerlof, G. A. (1970): The market for 'lemons': Quality uncertainty and the market mechanism. Quarterly Journal of Economics, Vol. 84(3), 488 – 500.

Aleksandrova, Y. (2021): Comparing performance of machine learning algorithms for default risk prediction in peer to peer lending. Journal of Technology, Education, Management, Informatics, Vol. 10(1), 133 – 143.

Anh, N. T. T./*Hanh*, P. T. M./*Le Thu*, V. T. (2021): Default in the US peer-to-peer market with covid-19 pandemic update: An empirical analysis from lending club platform. International Journal of Entrepreneurship, Vol. 25(7), 1 – 19.

Ari, A./*Ratnovski*, L./*Chen*, S. (2020): The dynamics of non-performing loans during banking crises: a new database. Working Paper Series 2395. European Central Bank.

Assaad, R. H./*Fayek*, S. (2021): Predicting the price of crude oil and its fluctuations using computational econometrics: deep learning, LSTM, and convolutional neural networks. Econometric Research in Finance, Vol. 6(2), 119–137.

Bakker, E. (2015): PEER-TO-PEER LENDING: How digital lending marketplaces are disrupting the predominant banking model. BUSINESS INSIDER 23.

Barasinska, N./*Schäfer*, D. (2014): Is crowdfunding different? Evidence on the relation between gender and funding success from a German peer-to-peer lending platform. German Economic Review, Vol. 15(4), 436–452.

Bartholomew, D. (2010): Principal components analysis. International Encyclopedia of Education. Ed. by Peterson, P./Baker, E./McGaw, B. Third Edition. Oxford: Elsevier, pp. 374–377.

Chromik, M. (2020): reshape: A framework for interactive explanations in XAI based on SHAP. In: Proceedings of 18th European Conference on Computer-Supported Cooperative Work. European Society for Socially Embedded Technologies (EUSSET).

Cumming, D./*Hornuf*, L. (2022): Marketplace lending of small-and medium-sized enterprises. Strategic Entrepreneurship Journal, Vol. 16(1), 32–66.

Diamond, D. W. (1984): Financial intermediation and delegated monitoring. The Review of Economic Studies, Vol. 51(3), 393–414.

Dorfleitner, G./*Hornuf*, L./*Weber*, M. (2022): Paralyzed by shock: the portfolio formation behavior of peer-to-business lending investors. Review of Managerial Science, 1–37.

Dorfleitner, G./*Priberny*, C. et al. (2016): Description-text related soft information in peer-to-peer lending – Evidence from two leading European platforms. Journal of Banking & Finance, Vol. 64, 169–187.

Duan, J. (2019): Financial system modeling using deep neural networks (DNNs) for effective risk assessment and prediction. Journal of the Franklin Institute, Vol. 356(8), 4716–4731.

Duarte, J./*Siegel*, S./*Young*, L. (2012): Trust and credit: The role of appearance in peer-to-peer lending. The Review of Financial Studies, Vol. 25(8), 2455–2484.

Dzik-Walczak, A./*Heba*, M. (2021): An implementation of ensemble methods, logistic regression, and neural network for default prediction in peer-to-peer lending. Zbornik Radova Ekonomskog Fakultet au Rijeci, Vol. 39(1), 163–197.

Herzenstein, M./*Sonenshein*, S./*Dholakia*, U. M. (2011): Tell me a good story and I may lend you money: The role of narratives in peer-to-peer lending decisions. Journal of Marketing Research Vol. 48.SPL, 138–149.

Iyer, R. et al. (2016): Screening peers softly: Inferring the quality of small borrowers. Management Science, Vol. 62(6), 1554–1577.

Jin, Y./*Zhu*, Y. (2015): A data-driven approach to predict default risk of loan for online peer-to-peer (P2P) lending. International Conference on Communication Systems and Network Technologies. IEEE, 609–613.

Käfer, B. (2018): Peer-to-Peer lending – a (financial stability) risk perspective. Review of Economics, Vol. 69(1), 1–25.

Kaveri, V./*Narang*, D. (2020): Peer to Peer Lending in India under COVID-19 Scenario. Vinimaya Vol. 41(1), 5 – 19.

Kim, J. Y./*Cho*, S. B. (2019): Predicting repayment of borrows in peer-to-peer social lending with deep dense convolutional network. Expert Systems, Vol. 36(4), e12403.

Kumar, V. et al. (2016): Credit risk analysis in peer-to-peer lending system. In: IEEE International Conference on Knowledge Engineering and Applications. IEEE, 193 – 196.

Lee, J. W./*Lee*, W. K./*Sohn*, S. Y. (2021): Graph convolutional network-based credit default prediction utilizing three types of virtual distances among borrowers. Expert Systems with Applications, Vol. 168, 114411.

Li, Y./*Chen*, W. (2020): A comparative performance assessment of ensemble learning for credit scoring. Mathematics, Vol. 8(10), 1756.

Liang, L./*Cai*, X. (2020): Forecasting peer-to-peer platform default rate with LSTM neural network. Electronic Commerce Research and Applications, Vol. 43, 100997.

Lin, M./*Prabhala*, N. R./*Viswanathan*, S. (2013): Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending. Management Science, Vol. 59(1), 17 – 35.

Mohammadi, A./*Shafi*, K. (2017): How wise are crowd? A comparative study of crowd and institutions. Academy of Management Proceedings, 13707.

Mokhtari, K. E./*Higdon*, B. P./*Başar*, A. (2019): Interpreting financial time series with SHAP values. Annual International Conference on Computer Science and Software Engineering, 166 – 172.

Najaf, K./*Subramaniam*, R. K./*Atayah*, O. F. (2022): Understanding the implications of FinTech Peer-to-Peer (P2P) lending during the COVID-19 pandemic. Journal of Sustainable Finance & Investment, Vol. 12(1), 87 – 102.

Nigmonov, A./*Daradkeh*, H. (2022): From One Crisis to Another: Impact of COVID-19 Pandemic on Peer-to-Peer Lending Market. Financial Transformations Beyond the COVID-19 Health Crisis, 363 – 402.

Parker, G. G./*Van Alstyne*, M. W./*Choudary*, S. P. (2016): Platform revolution: How networked markets are transforming the economy and how to make them work for you. WW Norton & Company.

Pope, D. G./*Sydnor*, J. R. (2011): What's in a picture? Evidence of discrimination from Prosper. com. Journal of Human Resources, Vol. 46(1), 53 – 92.

Ravina, E. (2019): Love & loans: The effect of beauty and personal characteristics in credit markets. Available at SSRN 1107307.

Rochet, J. C./*Tirole*, J. (2003): Platform competition in two-sided markets. Journal of the European Economic Association, Vol. 1(4), 990 – 1029.

Serrano-Cinca, C./*Gutiérrez-Nieto*, B./*López-Palacios*, L. (2015): Determinants of default in P2P lending. PLOS ONE, Vol. 10(10), e0139427.

Sherstinsky, A. (2020): Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena, Vol. 404, 132306.

Siami-Namini, S./*Tavakoli*, N./*Namin*, A. S. (2019): The performance of LSTM and BiLSTM in forecasting time series. In: IEEE International Conference on Big Data. IEEE, 3285–3292.

Song, Y. et al. (2020): Multi-view ensemble learning based on distance-to-model and adaptive clustering for imbalanced credit risk assessment in P2P lending. In: Information Sciences, Vol. 525, 182–204.

Stiglitz, J. E./*Weiss*, A. (1981): Credit rationing in markets with imperfect information. In: The American Economic Review, Vol. 71(3), 393–410.

Tang, H. (2019): Peer-to-peer lenders versus banks: substitutes or complements? The Review of Financial Studies, Vol. 32(5), 1900–1938.

Teply, P./*Polena*, M. (2020): Best classification algorithms in peer-to-peer lending. In: North American Journal of Economics and Finance, Vol. 51, 100904.

Varma, S./*Simon*, R. (2006): Bias in error estimation when using cross-validation for model selection. In: BMC bioinformatics, Vol. 7(1), 1–8.

Wang, Y./*Ni*, X. S. (2020): Risk prediction of peer-to-peer lending market by a LSTM model with macroeconomic factor. ACM Southeast Conference 2020, 181–187.

Wei, Z./*Lin*, M. (2017): Market mechanisms in online peer-to-peer lending. Management Science Vol. 63(12), 4236–4257.

Xia, Y./*Liu*, C./*Liu*, N. (2017): Cost-sensitive boosted tree for loan evaluation in peer-to-peer lending. Electronic Commerce Research and Applications, Vol. 24, 30–49.

Xu, J./*Lu*, Z./*Xie*, Y. (2021): Loan default prediction of Chinese P2P market: a machine learning methodology. Scientific Reports, Vol. 11(1), 1–19.

Yang, K./*Shahabi*, C. (2005): On the stationarity of multivariate time series for correlation-based data analysis. IEEE International Conference on Data Mining, Vol. 4.

Ye, X./*Dong*, L./*Ma*, D. (2018): Loan evaluation in P2P lending based on Random Forest optimized by genetic algorithm with profit score. Electronic Commerce Research and Applications, Vol. 32, 23–36.

Zhu, L. et al. (2019): A study on predicting loan default based on the random forest algorithm. Procedia Computer Science, Vol. 162, 503–513.

Table A1

Eliminated Attributes

Category	Eliminated variables
Attributes that cannot be used for prediction	ID, member ID, employment title, url, loan description, loan title
Attributes with more than 50 % missing values	months since borrower's last delinquency, months since last major derogatory, months since recent bank card delinquency, months since recent revolving delinquency, open installment account since 6 months, open installment account since 12 months, open installment account since 24 months, months since recent installment account opened, total balance of installment accounts, installment loan utilization, months since last record, open revolving trades since 12months, open revolving trades since 24 months, max current balance, open account since 6 months, all utilization, finance inquiry, total finance trades, inquiry last 12 m
Redundant attributes	annual income joint, dti joint, verified status joint
Attributes that are filled after the borrower starts to repay	policy code, payment plan, outstanding principal, outstanding principal by investors

Table A2

Descriptive Statistics of Selected Features

Feature Name	n	mean	std	min	25 %	50 %	75 %	max
loan amount (€)	1587613	15912.63	9872.05	1000	8000	14000	21550	40000
installment (€)	1587613	466.71	283.92	7.61	257.75	391.5	627.55	1719.83
grade	1587613	1.47	1.19	0	1	1	2	6
application type	1587613	0.09	0.29	0	0	0	0	1
funded amount (€)	1587613	15912.61	9872.05	1000	8000	14000	21550	40000
open account	1587613	12.37	5.91	1	8	11	15	104
total payment (€)	1587613	11909.70	9878.53	0	4556	9014.86	16452.51	70826.17
total received interest (€)	1587613	2401.12	2573.70	0	738.04	1543	3137.72	33337.57
verification status	1587613	0.61	0.49	0	0	1	1	1
term 36 months	1587613	0.69	0.46	0	0	1	1	1
term 60 months	1587613	0.31	0.46	0	0	0	1	1
fico range high	1587613	706.70	34.20	664	679	699	724	850
fico range low	1587613	702.70	34.20	660	675	695	720	845
annual income (€)	1587613	85730.00	90555.89	14	50000	71000	100000	10999200
employment length of borrower	1587613	3.59	3.22	0	1	2	6	10
delinquency in 2 years	1587613	0.29	0.85	0	0	0	0	35
home ownership mortgage	1587613	0.50	0.50	0	0	0	1	1
home ownership own	1587613	0.11	0.31	0	0	0	0	1

Feature Name	n	mean	std	min	25 %	50 %	75 %	max
home owner-ship rent	1587613	0.39	0.49	0	0	0	1	1
purpose car	1587613	0.01	0.10	0	0	0	0	1
purpose house	1587613	0.01	0.08	0	0	0	0	1
purpose medical	1587613	0.01	0.11	0	0	0	0	1
purpose renewable energy	1587613	0.00	0.02	0	0	0	0	1
purpose small business	1587613	0.01	0.10	0	0	0	0	1
dti	1587613	20.59	14.19	−1	13.56	19.29	25.92	999
total received principal (€)	1587613	9355.36	8537.10	0	3000	6529.99	12983.73	40000
outstanding principal by investor (€)	1587613	5258.53	8064.72	0	0	0	8460.71	40000
interest rate	1587613	13.03	5.10	5.31	8.81	12.4	16.02	30.99

```
#Algorithm for LSTM and BiLSTM
Inputs: Time series Outputs: Accuracy of the forecasted data
```

```
# Split data into: # 80% training and 20% testing data          Dataset Splitting
# Split training data into 80% training and 20% validation set
1. size1 ← length(series) * 0.64 , size2 ← length(series) * 0.16
2. train ← series[0...size1]
3 validation ← series [size1.....size2]
4. test ← series[size2...length(series)] # Set the random seed to a fixed value
5. set option = "L" (LSTM) or "B" (BiLSTM)
```

```
# Fit an LSTM or BiLSTM model to training data          Model Generation
  Procedure fit_lstm_Bilstm(train, epoch, neurons, dropout)
5. X ← train
6. y ← train
7. model = Sequential()
8. if option = 'L':
      model.add(LSTM(neurons), kernel_regularizer=l2,, dropout))
      model.add(LSTM(neurons), , dropout))
      layers.Dense(nb_classes, activation='softmax')
9. else option = 'B':
      model.add(lBidirectional(LSTM(neurons, return_sequences=True, dropout))
      x = layers.Bidirectional(layers.LSTM(neurons, return_sequences=False, dropout))
      x = layers.Dense(nb_classes, activation='softmax')
      end for return model
10. model.compile(loss=' categorical_crossentropy', optimizer='adam', learning rate=lr)
```

```
# Class Weights to Handle Imbalanced Data          Cost Sensitive Analysis
history=model.fit( train_data, epochs, steps_per_epoch,
validation_data=validation_data, validation_steps, class_weight=class_weight )
```

```
# Forecast the training dataset          Cross Validation
11. lstm_model.predict(train)

# Five fold Time Series Split expanding window validation on the test data
12. for each i in range(length(test)) do
# make one-step forecast
      X ← test[i]
      yhat ← forecast_lstm(lstm_model, X)
# record forecast
      predictions.append(yhat)
      expected ← test[i]
      end for
13. Accuracy← Model Evaluate(expected, predictions)
14. Calculate Confusion Matrix and Classification Report
```

```
15. SHAP Deep explainer to visualize the feature importance          Feature Importance
and their extensibility
```
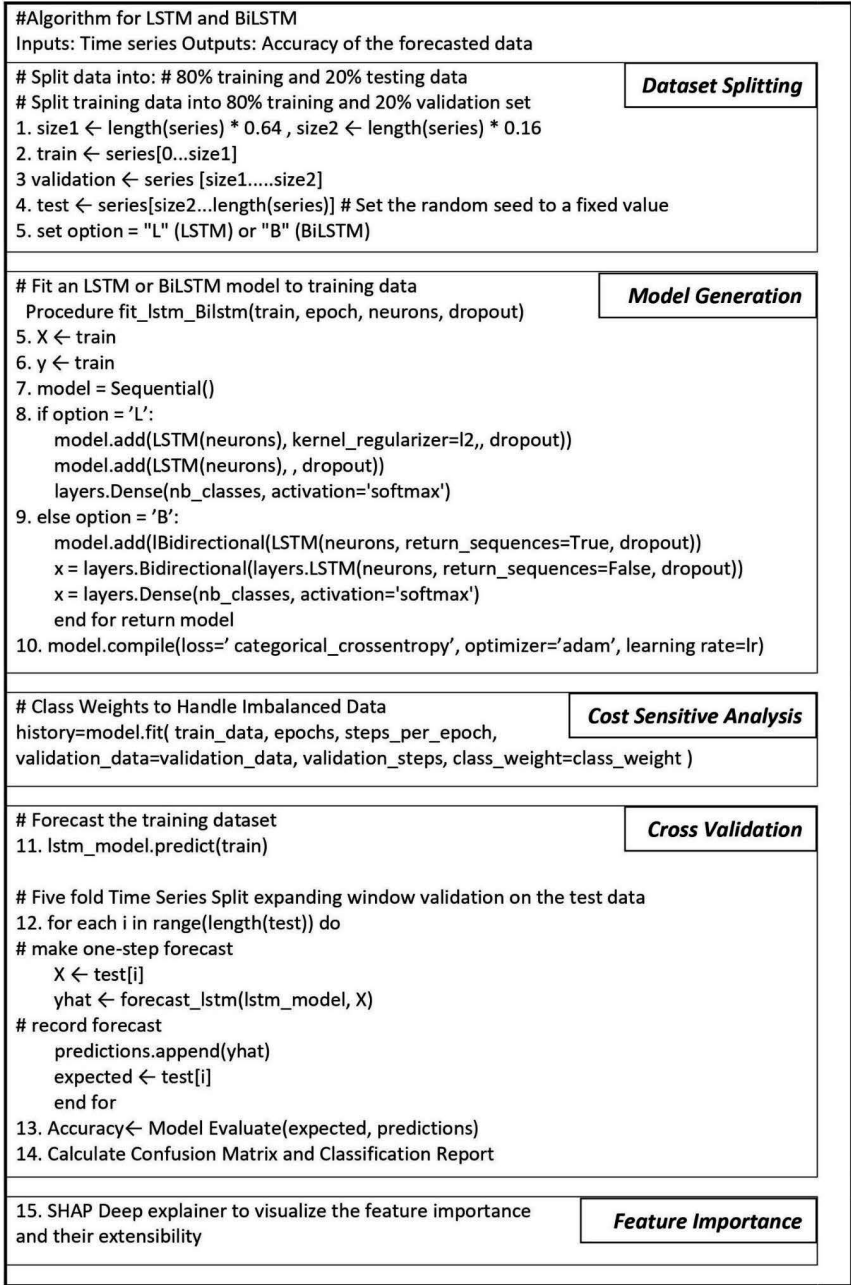
Figure A3: Algorithm of LSTM/BiLSTM