Machine Learning in Credit Risk Forecasting

—— A Survey on Credit Risk Exposure

Shaoshu Li[1]

[1] Quant Researcher, MOYI Inc., USA

Correspondence: Shaoshu Li, Quant Researcher, MOYI Inc., USA.

Received: April 25, 2024 Accepted: May 12, 2024 Online Published: May 14, 2024

doi:10.5430/afr.v13n2p107 URL: https://doi.org/10.5430/afr.v13n2p107

Abstract

Credit risk is one of the most important elements in risk management area. Traditional regression types of credit risk models are straightforward to implement and model outputs are easy to interpret. However, the model accuracy can always be suboptimal to fit the real credit risk data series. Especially, the model performance even deteriorates under extreme economic scenarios. In contrast, the modern machine learning models can handle different drawbacks of regression types of models. In this paper, we survey the recent literatures on applying the machine learning or deep learning methods in credit risk forecast with special focus on study the superiorities of these techniques. Besides of delivering better prediction accuracies, we uncover other four advantages for machine learning type of default forecast which have been shown in few literatures. We also survey the less studied machine learning or deep learning type of prepayment forecast. By reviewing past literatures from both default and prepayment risk aspects, we can gain comprehensive overview of utilizing machine learning techniques in credit risk forecasting and valuable insights for future risk management research.

Keywords: credit risk forecasting, machine learning, prepayment rate, default rate, learning curve, feature selection, splitting rule, misclassification cost, systematic credit risk

1. Introduction

Machine learning is a subfield of artificial intelligence which uses computer programs to learn from past experience (Mitchell, 1997). Compared to the traditional statistical learning, machine learning can well intimate the pattern of human being decision making process and the explicit algorithms to handle learning procedure can be saved (Makridakis et al., 2018). In this way, machine learning facilitates the computers to automatically correct learning errors and improve model accuracies by iteratively training the models using underlying algorithms. The four major categories of machine learning algorithms are supervised learning, semi-supervised learning, unsupervised learning and reinforced learning (Sarker, 2021). In the previous three learning categories, the machine learning algorithms analyze and generate any of labeled, unlabeled or a mixed of labeled and unlabeled data. In the reinforced learning, the machine makes decisions of the optimal behaviors by analyzing rewards and penalties for each action in a dynamic environment (Kaelbling et al., 1996). Due to its flexibilities in model structures and data types, machine learning has become a crucial element in the next generation of time series forecasting models (Lim and Zohren, 2021). The primary advantage of machine learning in financial time series forecasting is in its broad applicable to all types of financial time series, both high frequency and low frequency series. Financial time series are most volatile in nature and rich in noise. Given that the financial time series are majorly non-linear and non-stationary, traditional statistical models may fail to capture the complexity of financial time series. In contrast, machine learning has its merit in handling extreme volatile financial data series with few assumptions imposed. Another major advantage of machine learning in financial time series forecasting is machine learning can efficiently analyze huge amount of data within a relatively short period and with greater accuracy (Rundo et al., 2019). With rapid processing time of large dataset and high prediction accuracy, machine learning can provide real time insights of the financial time series and enable fast decision-making process. With its remarkable performance in financial time series forecasting and financial datamining, machine learning can be applied to almost every subfield in financial area. The financial applications of machine learning have always been an extremely popular research area in the past decades. In the past survey papers, scholars demonstrate machine learning or deep learning's powerfulness in market predictions, risk assessment, fraud detection, bankruptcy and financial crisis predictions, derivatives pricing and portfolio management, cryptocurrency and blockchain studies, financial sentiment analysis and behavioral finance (Ozbayoglu et al, 2020; Tang et al., 2022; Nazareth and Ramana Reddy, 2023).

The methodologies of machine learning in financial time series forecasting can be utilized in the modern risk management environment. Risk management is a paramount aspect in the banking system especially after 2008 financial crisis. The roles of risk management can be versatile, ranging from helping financial institution to manage risks, assisting banking system to function efficiently and supporting government to maintain social stability. As one of the greatest risks faced by financial institutions, credit risk measures the risk that a borrower or counterparty fail to perform on an obligation (Board of Governors of the Federal Reserve System). Effective credit risk management is extremely vital for financial institutions to make lending decision, optimize capital allocation and maintain profitability. In banking, regression types of credit risk models are widely used due to these models are easy to implement and interpret. However, the model accuracy can largely suffer by concrete functional forms which fail to take account of the complexity of real economic data series. Especially, the model performance even deteriorates under extreme economic scenarios. The growing in scale and complexity of financial institutions require them to employ complex risk management techniques and monitor the changing credit risk exposures (Angelini et al., 2008). With more flexible and adaptive model format, the machine learning models can handle different drawbacks of traditional regression models in credit risk management. In the past studies, machine learning has been successfully applied in diverse subareas in credit risk management, namely consumer credit risk, corporate credit risk, wholesale risk, credit card risk, concentration risk, counterparty credit risk, collateral risk (Leo et al., 2019). For large banks, loans are the largest and most obvious source of credit risk (Board of Governors of the Federal Reserve System). In the following chapters, we specifically focus our investigations on the applications of machine learning in loan credit risk exposure, from both default and prepayment forecast aspects. Past literatures provide fruitful achievements in discussions of the underlying machine learning techniques and their credit risk management applications. In our survey, we would save the introductions of the detail mathematical setups for those machine learning algorithms and put more effort into demonstrations of their real-world economic rationale in credit risk management.

The motivations of this paper can be summarized into threefold. Firstly, this paper surveys the most typical machine learning type of credit risk exposure literatures. Instead of classifying past literatures by risk management areas or underlying machine learning algorithms, we proceed our study by identifying the potential advantages of machine learning or deep learning techniques in default or prepayment forecast, with supporting literatures analyzed. Secondly, past financial forecasting literatures have adequately shown machine learning models deliver more accurate predictions than traditional statistical models. Besides of model accuracies, we would also like to address other potential advantages of machine learning techniques. Most of these advantages are hardly mentioned in the past survey literatures and could bring some brand-new research topics. Thirdly, although default and prepayment forecast share many common properties, few literatures analyze prepayment risk using machine learning or deep learning techniques. We also analyze the existing machine learning or deep learning type of prepayment forecast literatures.

Our paper proceeds as follows. Section 2 reviews the past literatures related to machine learning type of financial forecasting and the application of machine learning techniques in general credit risk management area. Section 3 elaborates the scope of our survey and its unique merit in credit risk management area. Section 4 surveys the existing machine learning default literatures and demonstrates five salient advantages of machine learning default forecast, which can shed light on future research. Section 5 surveys the existing machine learning prepayment literatures and illustrates the importance of feature selection techniques in picking the most important variables enter the prepayment model. Section 6 concludes the paper and provides implications.

2. Literature Review

Substantial literatures in the past decades provide the methodologies and applications of machine learning or deep learning in financial forecasting. The merits of applying machine learning or deep learning in financial forecasting are discussed, either in numerical methodologies or by empirical evidences. In Rundo et al. (2019), authors review different machine learning financial forecasting literatures. Three machine learning algorithms, support vector machine, deep learning and recurrent neural network, are elaborated and corresponding financial forecasting literatures are enumerated. Authors also summarize past empirical findings to show the benefits of machine learning type of financial forecasting over its traditional statistical forecasting counterpart. In Masini et al. (2023), authors review the most recent development of supervised machine learning and high dimensional models in time series forecasting. Majorities of linear and non-linear methods as well as the hybrid models have been considered. Authors also survey recent literatures of machine learning in economics and finance area, and provide empirical illustration with stock index series. In Sezer et al. (2019), authors summarize the basic mathematical setups for several deep learning techniques in detail, such as Convolutional Neural Networks (CNN), Deep Belief Networks (DBNs), Long-Short Term Memory (LSTM). They separately group the past deep learning financial forecasting literatures according to the subjects being forecasted, the feature set used and deep learning methods involved. In Lim and Zohren (2020), authors survey the methodologies of the common encoder and decoder designs in deep learning time series forecasting, for both on-step ahead forecasting models and multi-horizon forecasting models. They also illustrate the recent development in hybrid deep learning models.

Credit risk forecasting is a special type of financial forecasting. The modern machine learning or deep learning techniques can be efficiently adopted in credit risk forecasting and escalate the potentials in credit risk management area. In Khandani et al. (2010), using machine learning techniques, authors significantly improve the classifications rates of credit card holder delinquencies and defaults. Authors show machine learning models can help the bank make credit lines related decisions for customers' accounts. They also show machine learning models can be applied in forecasting aggregate consumer credit delinquencies and can have substantial importance in systematic risk management. In Moscatelli et al. (2020), authors analyze the performance of random forest model and gradient based trees model in default risk forecasting, compared to statistical models. They show when only public information is available, machine learning models can depict better forecasting accuracy compared to traditional statistical forecasting models. Moreover, authors show machine learning types of credit allocation rules can help the lenders select more reliable borrowers and result in lower credit losses. In Fuster et al. (2022), authors show the machine learning models can generate better predictions than traditional statistical models. Due to the sophisticated functional forms of these machine learning models, structural relationships behind independent variables of default rate forecasting can be captured with little effort. Besides, authors prove machine learning can increase the disparity in default rate forecasting between and within borrowers' groups due to the machine learning can triangulate borrowers' hidden identities. In Barboza et al. (2017), authors show machine learning models have advantages in the corporate bankruptcy predictions compared to traditional bankruptcy models. They conclude machine learning models can lead to more accuracy forecasting under restrictive conditions with high correlated variables, outliers, missing values and fewer variables' transformations. By including growth or change of firm behaviors which are hardly captured in the traditional bankruptcy models, machine learning models can again improve their accuracies. In Dhankhad et al. (2018), authors apply supervised machine learning methods to detect credit card fraudulent transactions using credit card from European datasets. They also implement the super classifier using ensemble learning methods and show the ensemble classifier performs better than single classifier when using imbalance dataset.

Past literatures provide various illuminating applications of machine learning or deep learning techniques in every aspect of credit risk management. In the following chapters, we would like to narrow down our discussions to the machine learning or deep learning techniques used in loan credit risk exposure forecast, especially on default and prepayment forecast.

3. Methodology and Special Merit

Credit risk exposure estimations involve the projections of expected loss using probability of default (PD), loss given default (LGD) and exposure at default (EAD), while the three later elements are extensively discussed in Basel II accord (Basel Committee on Banking Supervision, 2005). In banking, the two most common risks faced by loan lenders are default risk and prepayment risk. Default risk captures the probability of the borrowers fail to make scheduled payment of principal and interest. Prepayment risk captures the probability of the borrowers pay off the loan earlier than anticipated and lower the income flows to lenders. By calibrating default and prepayment risks, financial institutions can access the appropriateness of the borrowers, determine interest rates of the loan products and settle corresponding loan terms. Since the Basel II accord, the traditional regression models such as logistic regression have been widely used in the banking to forecast default and prepayment, largely due to the traditional regression models are relatively easy to implement and interpret. In these regression models, default and prepayment rates are projected using borrowers' creditworthy, the loan characteristics and the macroeconomic environment variables. Although the regression models have many merits especially in their reporting convenience, the accuracies of forecast can be largely hampered by the concrete functional forms for these models especially under financial stressful scenarios. In contrast, the machine learning models impose minimum assumptions on the functional forms of underlying models and provide relatively higher forecasting accuracies. In the following two subchapters, we will separately discuss past machine learning types of loan default and prepayment risk forecast literatures with special emphasizing the following aspects: (1) Majority of past survey type literatures cover many risk management topics at one time. These risk management topics may not be directly comparable and underlying risk management process vary significantly. We narrow down our survey to those literatures which are discussing default and prepayment risk with machine learning or deep learning techniques heavily involved. We analyze the most cited literatures for both default and prepayment forecast to identify the superiorities of machine learning or deep learning techniques. Follow the configurations of past survey literatures, we also compare typical literatures from the specific machine learning or deep learning methodologies being used, the research goal being attained, the dataset being analyzed and the cost functions being selected. (2) Past literatures have uniformly shown machine learning models deliver more accurate predictions than traditional statistical models. Besides of the better prediction accuracies for these machine learning models, we would also like to discuss other potential advantages of machine learning techniques in credit risk forecasting documented in the past literatures. Some of these advantages are less mentioned in the past survey literatures and could bring some brand-new research topics. By reviewing past literatures about their merits respectively, we can gain many insights for future credit risk management research. (3) Default risk has been heavily studied in the past machine learning type of credit risk literatures. In contrast, prepayment risk has been relatively less analyzed using machine learning or deep learning techniques. We summarize the existing machine learning or deep learning type of prepayment literatures and illustrate the importance of feature selection techniques in picking the most important variables to enter the prepayment model.

4. Default Forecast

Most of the past survey literatures in machine learning type of default forecast proceed their surveys according to categorize the reference literatures by underlying machine learning methodologies involved (Baesens et al., 2003; Lessmann et al, 2015; Ampountolas et al., 2021), the credit risk problems analyzed (Leo et al., 2019; Ozbayoglu et al., 2020) or combinations of these two topics. The machine learning techniques commonly mentioned in these literatures are the unsupervised machine learning techniques (KNN) and supervised machine learning techniques (SVM and tree-based models). All of the above machine learning algorithms can handle dataset classification problems and can be used to classify default cases from non-default cases in default forecast. Among these machine learning techniques, the variants of tree-based models provide an affluent resource in default forecast. For tree-based model, ensemble learning methods like bagging and boosting methodologies are used to combine a set of weak learners to a single strong learner to improve model performance. More recent literatures apply gradient tree boosting, extreme gradient boosting (XGBoost) and AdaBoost in default forecast. Some past literatures also contain discussions about deep learning techniques, specifically neural network classifiers. The neural network contains an input layer, several hidden layers and an output layer which acts like human brains. These layers connect with each other via neurons which receive, process and send signals. In some literature, neural network classifiers are ensembled with boosting techniques. For instance, deep neural network and XGBoost can be ensembled to form the Hybrid DNN-GBT forecast (Albanesi and Vamossy, 2019). With respected to model performance comparisons, these machine learning models are being compared with each other and the logistic regression model according to the predicting powers indicators, such as ROC curve, confusion matrix, error rates for both true default cases and non-default cases. Some survey literatures also classify past machine learning types of risk management papers according to the risk management problem or specific risk involved, with more detail categories for the machine learning techniques used. Some examples of the credit risk management problems being solved are: credit risk exposure, corporate credit risk, financial crisis predictions and stress testing. In contrast with previous survey papers, we will enumerate machine learning's superiorities from machine learning algorithms improve default prediction accuracy, enhance model efficiency, contribute to feature selection, reflect real misclassification magnitude and manage systematic credit risk. While prediction accuracy is always the major concern, the later four advantages are rarely included in the past survey papers and could bring brand new research perspectives.

4.1 Improve Prediction Accuracy

The majority of the most recent literatures are aiming to show machining learning models deliver better forecasting accuracy in default forecasting than regression types of models. These literatures involve the comparisons of several machine learning models or deep learning models with traditional logistic regression types of models. Although the detail performance rankings for the machine learning models differ among these literatures, the machine learning models universally deliver more accurate performance than regression types of models. In these works, the most commonly used machine learning methods are decision tree type of algorithms. Specifically, classification and regression tree (CART), gradient boosting (GB) and extreme gradient boosting (XGBoost) are frequently adopted in the past literatures. Methodologically, the CART recursively splits the training dataset into small subsets according to certain rules and make prediction of the target value within each subset. By aggregating these predictions, the CART can generate a weak learner. On the contrary, the boosting methods use decision trees as weak learners and combine these weak learners to form a strong learner. Compared to decision tree method like CART, boosting methods can put more weightings on more difficult cases to overcome such barriers in the learning procedures. In Barbaglia et al. (2023), authors compare the performance of gradient boosting and extreme gradient boosting with traditional logistic regression in forecasting loan default behavior in European countries. They show both boosting methods deliver more accurate forecast than traditional logistic regression, with the extreme gradient boosting achieves the most out-of-sample prediction accuracy. Authors argue the success of both boosting methods in default forecasting can be largely contributed by machine learning techniques can better capture the non-linear relationships behind different independent variables and loan default. In Galindo and Tamayo (2000), authors compare four statistical and machine learning modelling in mortgage loan default estimations. By comparisons, the CART delivers the best forecasting of default, following by the Neutral Networks and K-Nearest Neighbor. These three machine learning techniques outperform the probity algorithm in general. In Yeh and Lien (2009), authors compare the performance of classification and predictive accuracy of probability of default for credit card clients among six data mining methods. Using the proposed “Sorting Smoothing Method” in real probability default estimation, authors compare these six methods and find out the artificial neural network outperform other five methods in classification performance. In the meantime, authors conclude artificial neural network is the only one which can accurately forecast the probability of default.

Some other past literatures target on analyzing one machine learning algorithm and show its advantages over the other methodologies. The main machine learning techniques used in these literatures are: CART model and support vector machine (SVM). Literatures in this category always contain thorough review of the concrete mathematical setups of the algorithms used. In Feldman and Gross (2005), authors apply the CART algorithm to analyze Israeli real estate mortgage data. Authors explain the pros and cons of the CART model compared to traditional logistic regression, nonparametric additive logistic regression, discriminant analysis, partial least squares classification, and neural networks in their performance in forecasting mortgage default. Authors mathematically explain the misclassification cost functions of three splitting rules, Entropy, Gini and Twoing, as well as the tree pruning procedures. Using the CART model in forecast mortgage default, authors conclude the borrowers’ features have better default forecasting power rather than the mortgage contract features, given the cost of accepting the bad risks exceeds rejecting good ones. If both costs are equal, CART uses mortgage features as well. In Khandani et al. (2010), authors apply the CART algorithm to forecast credit card holders’ delinquencies and defaults. Authors depict the machine learning forecast differ substantially to CScore scores especially for the groups with high CScore scores but have already encountered delinquencies. Authors show machine learning models can help the bank make credit lines related decisions for customers’ accounts and forecast aggregate consumer credit delinquencies. In Bellotti and Crook (2009), authors apply the SVM to classify credit card customers into default cases and non-default cases. Author compare the SVM against logistic regression, linear discriminant analysis, KNN. They find the privileges of SVM in classifying credit card customers who default. The numerical problem for SVM’s optimization and different kernel functions to capture nonlinearities are detailly explained. Authors find out the superiorities of SVM in classification the good cases from the bad cases in default forecasting. They also show SVM can serve as a useful feature selection tool to determine the most important features.

Besides of applications of machine learning models into default forecasting, past literatures also involve the implementations of ensemble learning. The ensemble learning combines multiple machine learning algorithms to obtain better predictive performance than could be obtained by any of the constituent learning algorithms alone. Past literatures depict ensemble learning can bring additional improvement of prediction accuracy in default forecasting. In Albanesi and Vamossy (2019), authors apply the hybrid DNN-GBT model to predict consumer loan default. Authors show their hybrid DNN-GBT model performs better than the traditional logistic regression model and other machine learning types of models in default forecasting according to their specific sample. Such DNN-GBT model is interpretably and can also apply to a large class of borrowers relative to the standard credit scoring models. Authors conclude the complexity of the model can contribute the accuracy, while the deep neural networks can substantially improve the shallow models.

4.2 Enhance Model Efficiency

The machine learning models’ enhanced forecast efficiency can be explained by machine learning models’ steep learning curves. Compared to regression types of default forecast models, the machine learning models depict steeper learning curves for large sample which reveals machine learnings proceed large sample more efficiently. With a growing sample size, we expect models can achieve better accuracies since more information is available in learning. However, the speeds to achieve higher accuracies vary across models. To capture the speeds, the learning curve shows how the training or cross-validation errors change according to the sample size. With error rate on the vertical axis and sample size on the horizontal axis, the learning curve is upward sloping for training and downward sloping for cross validation. A steep learning curve reflects the cross-validation errors drop or training errors rise quickly as the sample size increases. In this way, steep learning curve indicates the model can achieve higher performance more rapidly. In Perlich et al. (2003), authors perform comparisons between logistic regression and tree types machine learning models using 36 binary classification data sets. By learning curve comparisons, authors conclude the logistic regression is better only for smaller training sets. Tree types models surpass logistic regression for large training sets both in terms of accuracy and AUR. Such relationship holds for dataset drawn from the same domain. Under higher signal-separable situation, the logistic regression learning curve is slightly steeper than the tree types of models when sample size is small. However, the logistic regression learning curve soon levels off and bypassed by tree types of models. The tree types of models keep learning and eventually achieve higher accuracy. In Galindo and Tamayo (2000), authors compare the learning curves for probit model, decision tree CART model, Neural Networks and KNN in mortgage loan default forecast. Authors show learning curves for testing by drawing test errors against sample sizes for each model and generate learning curves comparisons by adjusting model parameters or iterations. Given 2000 records, authors compare the error rates for above models and show the three machine learning models can deliver lower test errors than probit model. Authors also show the three machine learning models have much steeper learning curves than probit model for sample size greater than 128. The results depict machine learning techniques can reduce the model errors at a faster speed than traditional regression model when sample size is large.

4.3 Contribute to Feature Selection

The broadly used feature selection techniques in the machine learning domain can also contribute to develop machine learning types of default forecast. The major types of features involved in default forecast are loan characteristics variables, borrowers' creditworthiness and macroeconomic environment variables. The traditional regression types of models require calculations of performance measures such as MSE, R square and loglikelihood to evaluate the contributions of features and perform feature selections. Most of these performance measures only take account of model fitness, but fail to take account of model complexity. In contract, feature selections in machine learning techniques reduce model complexity on the one hand and reduce overfitting on the other hand. For instance, the well-known regularization methods Lasso (L1 regularization) and Ridge (L2 regularization) can combine reduce overfitting and the selection of useful features at the same time. Both regularization methods are commonly used in machine learning algorithms. Another advantage of machine learning lies in its automation the feature selection in learning procedure. For some machine learning models, namely SVM, feature selections are based on the output of the models. However, for some other machine learning models, namely the decision trees, the feature selections are automatically involved in the learning procedures. In decision trees, information gain is calculated at each node to select the feature for data splitting. Features with higher information gain are considered to be more important and appear closer to the root of the tree. By recursively select features for splitting, decision trees automatically select the relevant features and rank the feature importance. In Bellotti and Crook (2009), authors compare the feature selection procedures between support vector machine and logistic regression in consumer default forecast. The maximum likelihood method is used in logistic regression in determination of the most important features in the model. By maximizing the likelihood function, the Wald statistics for each coefficient and corresponding p value is generated. Compare these statistics with the critical values, one can determine the features to be included into the model. For support vector machine, authors refer to Guyon et al. (2002)'s feature selection criteria and use the magnitudes of weights from the hyperplane generated by SVM as feature selection criterion. They set a threshold of 0.1 and all features with weights greater than this threshold are selected as significant features. They further discuss the overlapping of the most important features in both models. In Barbaglia et al. (2023), authors compare the feature importance of gradient boosting and extreme gradient boosting default forecast models for seven European countries. The feature importance is calculated as reduction in the squared prediction error as a result of the split in the tree and attributed to the splitting variable. Authors repeat this procedure across all the nodes of a tree and take summation across all trees. The resulting quantity is then standardized and can be interpreted as the importance of a feature in default forecast. By comparisons, authors conclude the loan related variables, such as LTV and interest rate, are most important features in default forecast. Authors also use Accumulated Local Effect (ALE) plots to explore the marginal effect of the explanatory variables on default forecast. The ALE averages the prediction changes calculated on small partitions of the variable of interest and then accumulates them over all partitions. Authors show ALE plots for LTV and interest rate for different countries. Contrary to the linear relationships shown in logistic regression, the ALE plots depict nonlinear relationships between the probability of default and explanatory variables, which again reflect machine learning models' superiority in capturing the nonlinear interactions of variables.

4.4 Reflect Real Misclassification Cost Magnitude

The real misclassification cost magnitude can also be captured and built in the machine learning default forecast. Under many circumstances, the cost of errors is not always equal. The cost of making an error depends on both the predicted class of example and actual class of an example (Pazzani et al., 1994). The two misclassification cases in default forecast are the false positive cases (actual nondefault but predict as default) and false negative cases (actual default but predict as nondefault). Since missing actual default can be costly, the cost to misclassify bad borrower as good can be much higher than the reverse misclassification. Besides of simple calculations of the accuracy of the model, the machine learning default forecast can be designed appropriately to reflect the unequal misclassification costs by punishing more to the false negative cases. Although the misclassification cost is well studied in the machine learning field, it has been relatively less adopted in the past machine learning types of default forecast literature potentially due to its mathematical complexity. In Lessman et al. (2015), authors compare error cost of artificial neural network, random forest and HECS-Bag in default forecast for loans. 25 cost ratios between false positive and false negative cases are used to represent the more costly to grant credit to a bad borrower than rejecting a good application. Authors compare cost reductions for these three machine learning techniques against traditional logistic regression and show three machine learning classifiers can substantially reduce error costs of logistic regression type default forecast. However, the magnitude of error cost difference between false negative and false positive cases is considered in result comparison but isn't appeared in training step for three machine learning models. In contrast, the following paper provide a good example in addressing the unequal misclassification cost in the intermediate learning procedure. In Feldman and Gross (2005), authors include detail discussions of feature selections in CART mortgage default forecast. authors document three tree splitting rules for CART model and provide detail tree splitting examples. The tree splitting rules can handle both uniform cost and non-uniform cost cases. The non-uniform cost cases are the situations where the misclassification cost depends on both actual and predicted situations. Given the cost of false positive and false negative cases is not always equal in default forecast, the non-uniform cost cases can better reflect the real-world situation where the cost to misclassify bad borrower as good can be much higher than the reverse misclassification. Authors conclude after they impose such non-uniform cost function in model training, the trees process high sensitivity (probability of classifying cases as such) relative to specialty (probability of classifying non-cases as such) are obtained.

4.5 Manage Systematic Credit Risk

The applications of machine learning in systematic credit risk management can be explained from the broad society level and from the base individual financial institution level. The machine learning default forecast for individual consumers account can be aggregated to generate credit risk forecast for individual financial institution. As shown in the extension part in a few literatures, credit risk forecast for individual financial institutions using machine learning models can be aggregated into society level macroeconomic credit risk forecast under certain conditions. Using machine learning techniques to straightforwardly calibrate the systematic credit risk in the financial system is also a promising research area. The following papers demonstrate the possibilities of aggregate individual level credit risk to obtain systematic credit risk. In Galindo and Tamayo (2000), authors describe different ways aggregate risk of one institution as well as the entire financial system. Authors illustrate aggregate portfolio credit risk using the machine learning methods applied in this paper and introduce several types of aggregation methods. Authors also discuss the prerequisite conditions to aggregate risk faced by individual financial institutions to the entire financial system. In Khandani et al. (2010), authors show machine learning models can be applied in forecasting aggregate consumer credit delinquencies and can have substantial importance in systematic risk management. Authors construct aggregate delinquency probabilities in the year prior to the crisis of 2007-2009 and depict the machine learning type of systematic credit risk forecasting can generate higher forecasting delinquencies probabilities evening during the boom years. They conclude the machine learning techniques applied to combined transactions and credit scores are capable of generating leading indicators of deterioration in consumer creditworthiness. In some other literatures, systematic credit risk is estimated directly at a society level or social group level using machine learning techniques. In Fuster et al. (2022), authors compare logit models and tree-based models in mortgage default estimations using U.S. data. The show machine learning models especially the random forest model can deliver better performance than the logit and non-logit types of models. They also analyze the default predictions using machine learning across different ethnical groups. They find machine learning techniques can screen more extensively among minority groups than logit model, which brings more disperse predictions between and within different ethical groups.

We summarize some most cited machine learning default forecast literatures in Table 1.

Table 1. ML default forecast literatures

Authors	Article	Methods	Main Goal	Dataset	Performance Criteria	Cost functions
Barbaglia et al., 2023	Forecasting loan default in Europe with Machine Learning	Logistic regression, Gradient Tree Boosting, Extreme Gradient Boosting	Loan default behavior in European countries	Residential loans in European Central Bank European Datawarehouse (ED)	AUC, H measure, Brier's score (BS), Logarithmic scores (LS)	Cross entropy loss function
Galindo and Tamayo, 2000	Credit Risk Assessment Using Statistical and Machine Learning: Basic Method and Risk Modeling Applications	Probit model, Decision tree cart model, Neural Networks, KNN	Compare predictive accuracy of different statistical and machine learning modeling methods on mortgage default	Mexico's security exchange and banking commission: Comision Nacional Bancaria y de Valores (CNBV)	Error rate, Confusion matrix, Complexity, Optimal sample size	Gini impurity
Yeh and Lien, 2009	The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients	KNN, Logistic regression, Discriminant analysis, Naïve Bayesian classifier, Artificial neural networks, Classification trees	Compare the predictive accuracy for probability of default among different data mining methods	Payment data in Oct, 2005 from an important bank in Taiwan	Area ratio, Error rate	Impurity Measure
Albanesi and Vamossy, 2019	Predicting Consumer Default: A Deep Learning Approach	Deep Neural Network, Classification and Regression Trees (CART), Extreme Gradient Boosting (XGBoost), Hybrid DNN-GBT Model	Predict default rate for consumer loans	Credit file data from Experian credit bureau	Confusion matrix, AUC	Cross entropy loss function
Bellotti and Crook, 2009	Support vector machines for credit scoring and discovery of significant	Support vector machines (SVM), Logistic regression,	Classify credit card consumers into default cases and	Four different credit card products data from a major financial	AUC, Error rate	Magnitude of weights on features

	features	Linear discriminant analysis, KNN	non-default cases	institution		
Feldman and Gross, 2005	Mortgage Default: Classification Trees Analysis	Classification and Regression Trees (CART)	Classify mortgage borrowers into potential defaulters and those unlikely to default	Residential mortgage contracts from a major Israeli mortgage bank	Confusion matrix	Entropy function, Gini index of diversity, Twoing function
Khandani et al., 2010	Consumer credit-risk models via machine-learning algorithms	Classification and Regression Trees (CART)	Forecast credit card holders' delinquencies and defaults	Bank's account-level transactions data and credit bureaus data	Confusion matrix, AUC	Gini impurity

5. Prepayment Forecast

Compare to abundant literature resources of machine learning type of default forecast, prepayment forecast with machine learning techniques involved has been less discussed in the past literatures. The prevailing prepayment forecast models shown in the past literatures are the regression types of models, either parametric regressions or nonparametric regressions, with different factors affect loan prepayment as independent variables (Maxam and LaCour-Little, 2001; Schwartz and Torous, 1993; Kang and Zenios, 1992; Green and Shoven, 1986). Due to the scarcity of the machine learning type of prepayment forecast literatures, we can only comprehensively study the most cited ones. Since default and prepayment forecast share many common natures, most machine learning and deep learning techniques shown in previous default forecast literatures can be transferrable to predict prepayment as well. However, due to the extremely small literature pool, both prepayment literatures we explore in this part adopt neural network model. The results in these papers demonstrate the neural network model can better capture the nonlinearities in prepayment risk predictions and deliver accurate predictions. In Sadhwani et al. (2020), authors use neural network to analyze the relationships of borrower specific behaviors on mortgage prepayment and delinquency with interactions between variables taken into account. Broad range of borrower specific variables, loan specific variables and macroeconomic variables have been taken account of. They uncover many variables have highly nonlinear influence on borrower's behavior and variable interactions represent a significant part of nonlinear effects. By addressing the nonlinearities in mortgage risk calibration via deep learning models, the mortgage risk forecast accuracies, investment performance of mortgage trading strategies and hedging of MBS can be improved. In Sitzia et al. (2022), authors calibrate prepayment rate for fixed rate mortgage and floating rate mortgage using behavior model based on neural network. The neural network is calibrated using past history of the observed prepayment event and the features are selected using random forest model for both contractual elements and market factors. By comparisons, the paper shows the neural network can deliver better prediction accuracies than Kaplan-Mayer estimator (KM) or logistic estimator. The neural network prepayment forecast is compared with actual repayment rate observed in the past, showing neural network is suitable for such behavior modelling.

The five advantages for machine learning in default forecast shown in previous chapter can also be transferred to study machine learning or deep learning type of prepayment forecast as more literatures become available. Intuitively, the main difference between the default forecast and prepayment forecast should attribute to the features chosen in the model development, regardless the traditional regression type of forecast or the machine learning forecast. Past prepayment literatures demonstrate several common features to be included in prepayment forecast. In these common features, spread term, the age of the mortgage, burnout effect and seasonality are the most essential ones. Most of the past prepayment literatures, either regression type or machine learning type, use predetermined features. Within each literature, the prepayment risk is forecasted by a group of independent variables chosen subjectively. In Maxam and LaCour-Little (2001), authors summarize the common factors used by industry and academic research in prepayment models are the refinancing incentive, seasoning or mortgage age, premium burnout, seasonality, and various macroeconomic and interest rate variables. In Kang and Zenios (1992), authors summarize the characteristics of the mortgage that provide indicators for the prepayment behaviors of the borrowers are the age of the mortgage, the month of the year, the ratio between the mortgage contract rate and the prevailing rate at which mortgage can be refinanced. In Sitzia et al. (2022), authors summarize two reasons in determinations of prepayment rate for mortgages are non-financial reasons and financial reasons. The financial reasons are related to the movement of market reference rate. The non-financial reasons are idiosyncratic elements and contractual features. Examples of non-financial reasons are outstanding mortgage amount, remaining time before expiration and time effects. As shown above, the features included in the prepayment model are predetermined for majority of literatures. In contrast, the following literature depicts some feature selection procedures using gradient-based approach. In Sadhwani et al. (2020), the economic significance of a variable is measured by the magnitude of the derivative of a fitted transition probability with respect to this variable. Using such gradient-based approach, authors show the most influential pairs of variables for prepayment forecast are original interest rate, state unemployment, loan balance variables and FICO score while the most influential pairs of variables for delinquency forecast are original interest rate, interest rate differentials, original loan term, FICO score, loan balance variables, and past delinquency behavior. As can be seen from Sadhwani et al. (2020), not all commonly used features mentioned in the past literatures prove to have nonnegligible predicting power in prepayment forecast after the feature selections. Thus, it would be rewarding to perform some preliminary feature selections before actual model development. Meanwhile, establishing some appropriate feature selection techniques to identify the most important features in the intermediate learning procedure would again be a promising research area.

We summarize two papers mentioned above which contain machine learning prepayment forecast in Table 2.

Table 2. ML prepayment forecast literatures

Authors	Article	Methods	Main Goal	Dataset	Performance Criteria	Cost functions
Sadhwani et al. (2020)	Deep Learning for Mortgage Risk	Neural network	Prepayment and Delinquency Estimation	U.S. Mortgage data	AUC	Cross entropy loss function
Sitzia et al. (2022)	A Neural Network Approach for the Estimation of Mortgage Prepayment Rates	Logit model, Kaplan-Mayer estimator Neural Network,	Conditional Prepayment Rate Prediction	Italian Retail Mortgage	Compare realized series with the model predictions, RMSE	Binary cross entropy function

6. Conclusions and Future Research

Credit risk management is an important risk management area. Developing up to date credit risk model is crucial to meet regulatory purpose and maintain economic stability. In industry credit risk management, regression types of models have been widely used. However, the emerging machine learning methodologies deliver more accurate predictions as shown in the past academic studies. In this paper, we summarize and analyze previous literatures which utilize machine learning techniques in credit risk exposure forecast, especially default and prepayment risk forecast. We enumerate five unique advantages of machine learning type of default forecast with supporting literature evidences listed. Besides of outstanding prediction accuracies for machine learning models, other four advantages are shown in very few literatures. We also survey the less studied machine learning or deep learning type of prepayment forecast. Future research can be proceeded from diverse aspects, namely: (1) Theoretically, features enter the default and prepayment models have significant overlapping. Suitable feature selection method can be established to distinguish the features used in default forecast with the features used in the prepayment forecast. (2) The actual misclassification costs for false negative and false positive cases are not always equal. In credit risk forecast, more delicate cost function should be designed to reflect the non-uniform penalties for misclassification cases in the intermediate learning procedures. (3) Using machine learning techniques to meticulously capture systematic credit risk could also be a valuable research topic.

References

Albanesi, S., & Vamossy, D. F. (2019). Predicting Consumer Default: A Deep Learning Approach. *NBER Working Paper Series*, Working Paper 26165. https://doi.org/10.3386/w26165

Ampountolas, A., Nde, T., Date, P., & Constantinescu, C. (2021). A Machine Learning Approach for Micro-Credit Scoring. *Risks, 9*(3), 50. https://doi.org/10.3390/risks9030050

Angelini, E., Tollo, G., & Roli, A. (2008). A neural network approach for credit risk evaluation. *The Quarterly Journal of Economics and Finance, 48*(4), 733-755. https://doi.org/10.1016/j.qref.2007.04.001

Baesens, B., Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring. *The Journal of the Operational Research Society, 54*(6), 627-635. https://doi.org/10.1057/palgrave.jors.2601545

Barbaglia, L., Manzan, S., & Tosetti, E. (2023). Forecasting Loan Default in Europe with Machine Learning. *Journal of Financial Econometrics, 21*(2), 569-596. https://doi.org/10.1093/jjfinec/nbab010

Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications, 83*, 405-417. https://doi.org/10.1016/j.eswa.2017.04.006

Basel Committee on Banking Supervision. (2005). International Convergence of Capital Measurement and Capital Standards. *Bank for International Settlements*. https://www.bis.org/publ/bcbs118.pdf

Bellotti, T., & Crook, J. (2009). Support vector machines for credit scoring and discovery of significant features. *Expert Systems with Applications, 36*(2), 3302-3308. https://doi.org/10.1016/j.eswa.2008.01.005

Board of Governors of the Federal Reserve System. *Supervisory Policy and Guidance Topics: Credit Risk Management*. https://www.federalreserve.gov/supervisionreg/topics/credit_risk.htm

Dhankhad, S., Mohammed, E., & Far, B. (2018). Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study. *IEEE,* 122-125. https://doi.org/10.1109/IRI.2018.00025

Feldman, D., & Gross, S. (2005). Mortgage Default: Classification Trees Analysis. *The Journal of Real Estate Finance and Economics, 30*(4), 369-396. https://doi.org/10.1007/s11146-005-7013-7

Fuster, A., Goldsmith-Pinkham, P., Ramadoral, T., & Walther, A. (2022). Predictably Unequal? The Effects of Machine Learning on Credit Markets. *The Journal of Finance, 77*(1). https://doi.org/10.1111/jofi.13090

Galindo, J., & Tamayo, P. (2000). Credit Risk Assessment Using Statistical and Machine Learning: Basic Methodology and Risk Modeling Applications. *Computational Economics, 15*(1-2), 107-143. https://doi.org/10.1023/A:1008699112516

Guyon, I., Weston, J., & Barnhill, S. (2002). Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning, 46,* 389-422. https://doi.org/10.1023/A:1012487302797

Green, J., & Shoven, J. (1986). The Effects of Interest Rates on Mortgage Prepayments. *Journal of Money, Credit and Banking, 18*(1), 41-59. https://doi.org/10.2307/1992319

Kaelbling, L., Littman, M., & Moore, A. (1996). Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research, 4*, 237-285. https://doi.org/10.1613/jair.301

Kang, P., & Zenios, S. (1992). Complete Prepayment Models for Mortgage-Backed Securities. *Management Science, 38*(11), 1665-1685. https://doi.org/10.1287/mnsc.38.11.1665

Khandani, A., Kim, A., & Lo, A. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance, 34*(11), 2767-2787. https://doi.org/10.1016/j.jbankfin.2010.06.001

Leo, M., Sharma, S., & Maddulety, K. (2019). Machine Learning in Banking Risk Management: A Literature Review. *Risks, 7*(1). https://doi.org/10.3390/risks7010029

Lessmann, S., Baesens, B., Seow, H., & Thomas, L. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: And update of research. *European Journal of Operational Research, 247*(1), 124-136. https://doi.org/10.1016/j.ejor.2015.05.030

Lim, B., & Zohren, S. (2021). Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 379*(2194). https://doi.org/10.1098/rsta.2020.0209

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PlOS ONE, 13*(3). https://doi.org/10.1371/journal.pone.0194889

Masini, R., Medeiros, M., & Mendes, E. (2023). Machine learning advances for time series forecasting. *Journal of Economic Surveys, 37*, 76-111. https://doi.org/10.1111/joes.12429

Maxam, C., & LaCour-Little, M. (2001). Applied Nonparametric Regression Techniques: Estimating Prepayments on Fix-Rate Mortgage-Backed Security. *Journal of Real Estate Finance and Economics, 23*, 139-160. https://doi.org/10.1023/A:1011102332025

Mitchell, T. (1997). Machine Learning. *McGraw-Hill Science*.

Moscatelli, M., Narizzano, S., Parlapiano, F., & Viggiano, G. (2020). Corporate Default Forecasting with Machine Learning. *Risk Management Magazine, 15*(3). https://doi.org/10.47473/2020rmm0071

Nazareth, N., & Ramana Reddy, Y.V. (2023). Financial applications of machine learning: A literature review. *Expert Systems with Applications, 219*. https://doi.org/10.1016/j.eswa.2023.119640

Ozbayoglu, A.M., Gudelek, M.U., & Sezer, O.B. (2020). Deep learning for financial applications: A survey. *Applied Soft Computing Journal, 93*. https://doi.org/10.1016/j.asoc.2020.106384

Pazzani, M., Merz, C., Murphy, P., Ali, K., Hume, T., & Brunk, C. (1994). Reducing Misclassification Costs. *Machine Learning Proceedings*, 217-225. https://doi.org/10.1016/B978-1-55860-335-6.50034-9

Perlich, C., Provost, F., & Simonoff, J. (2003). Tree Induction vs. Logistic Regression: A Learning-Curve Analysis. *Journal of Machine Learning Research, 4*, 211-255.

Rundo, F., Trenta, F., Stallo, A.L., & Battiato, S. (2019). Machine Learning for Quantitative Finance Applications: A Survey. *Applied Sciences, 9*(24), 1-20. https://doi.org/10.3390/app9245574

Sadhwani, A., Giesecke, K., & Sirignano, J. (2021). Deep Learning for Mortgage Risk. *Journal of Financial Econometrics, 19*(2), 313-368. https://doi.org/10.1093/jjfinec/nbaa025

Sarker, I. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science, 2*(160). https://doi.org/10.1007/s42979-021-00592-x

Schwartz, E., & Torous, W. (1993). Mortgage Prepayment and Default Decisions: A Poisson Regression Approach. *Journal of the American Real Estate and Urban Economics Association, 21*(4), 431-449. https://doi.org/10.1111/1540-6229.00619

Sezer, O.B., Gudelek M.U., & Ozbayoglu, A.M. (2020). Financial Time Series Forecasting with Deep Learning: A Systematic Literature Review: 2005-2019. *Applied Soft Computing, 90*. https://doi.org/10.1016/j.asoc.2020.106181

Sitzia, L., Baccaglini, R., Malacchia, V., & Cozzi, F. (2022). A Neural Network Approach for the Estimation of Mortgage Prepayment Rates. *Social Science Research Network (SSRN) Working Paper.* https://doi.org/10.2139/ssrn.4179429

Tang, Y., Song, Z., Zhu, Y., Yuan, H., Hou, M., Ji, J., Tang, C., & Li, J. (2022). A survey on machine learning models for financial time series forecasting. *Neurocomputing, 512*, 363-380. https://doi.org/10.1016/j.neucom.2022.09.003

Yeh, I., & Lien, C. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications, 36*(2), 2473-2480. https://doi.org/10.1016/j.eswa.2007.12.020

Copyrights