BACS: blockchain and AutoML-based technology for efficient credit scoring classification

Fan Yang¹ · Yanan Qiao¹ · Yong Qi¹ · Junge Bo¹ · Xiao Wang¹

Accepted: 3 January 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Credit evaluation is of high scientific significance and practical use, especially in today's plight of the world suffering from the COVID-19 epidemic. However, due to the difficulties inherent in credit scoring model building which involves a large number of data mining steps and requires a lot of time to process the data and build the model, efficient and accurate credit scoring methods are urgently required. Aiming to solve this problem, we propose BACS, an blockchain and automated machine learning based classification model using credit dataset so that the credit modelling processes are performed in the pipeline in an automated manner to eventually obtain the classification results of credit scoring. BACS scheme consists of credit data storage to blockchain, feature extraction, feature selection, modelling algorithm and hyperparameter optimization, and model evaluation. Firstly, we propose a mechanism for credit data management and storage using blockchain to ensure that the entire credit scoring system is traceable and that the information of each scoring candidate is securely, efficiently and tamper-proofly stored on the blockchain nodes. Next, we design a pipeline using a random forest model to effectively integrate the key steps of credit data feature extraction, feature selection, credit model construction, and model evaluation. The experimental results demonstrate that our proposed automated machine learning-based credit scoring classification scheme BACS can assess the credit condition efficiently and accurately.

Keywords Credit scoring · Credit crisis · Hyperparameter optimisation · Blockchain technology · Classification model · Automated machine learning

1 Introduction

Credit scoring is an integral part of the modern economy. The economic activities of every individual and enterprise in society are based on credit relationships, and the intricate social credit relationships form the economic foundation of society. Therefore, a modern market economy can only continue to exist and develop if stable and reliable credit relationships are

✉ Yanan Qiao
qiaoyanan@mail.xjtu.edu.cn

¹ School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi, People's Republic of China

established on the basis of a set of strict credit scoring systems, and credit systems are an important support for credit transactions.

Credit risk and its associated investment risk are potentially huge risks in modern economies and societies (Gaganis et al. 2021; Doumpos and Zopounidis 2007; Mahbobi et al. 2021). 2008 was the year when the credit risk of subprime housing credit in the US triggered the investment risk of credit derivatives in the financial markets, causing a financial crisis that swept the world and had a major impact on global economic development. To this day, many countries are still mired in the debt crisis triggered by the financial crisis, with socio-economic development stagnating or even declining and societies in constant turmoil. At the same time, the COVID-19 outbreak, which began in 2020, has hit the global economy severely, while causing a significant corporate credit crisis. Therefore, it is of great practical significance and research value to study an efficient and stable credit scoring system, and thus effectively control credit risk, in order to alleviate the global credit crisis caused by the epidemic and stabilise the sustainable development of the global economy. Efficient credit scoring systems not only bring great value to creditors, but also have an important role for credit recipients and society. Specifically, for creditors, credit scoring systems can be used to identify markets and develop new customers, enabling creditors to better understand customers' consumption behavior and purchasing power, facilitating the establishment of good credit relationships with customers, and promoting the improvement of internal financial management quality.

The era of big data has arrived with the rapid expansion of data accumulated in all areas of society, including scientific research, production and consumption, at an unprecedented rate. Making full use of data information and tapping into the value contained therein has become a general consensus among academia, industry and governments. In recent years, machine learning has achieved remarkable results in terms of theory, methods and applications. As stated in a review article published by Science (Jordan and Mitchell 2015), machine learning has been one of the most rapidly developing areas of information science and technology. Therefore, with the advancement of deep learning and machine learning research, the efficiency and performance of the credit scoring system will also be improved considerably.

Credit scoring research, while impacting a wide range of industries and especially playing an significant role in the current global economic downturn due to the impact of COVID-19, still faces considerable challenges. Figure 1 shows the distribution of different credit scores in the credit data. We can see that the challenge of credit distribution lies mainly in how to efficiently classify a smaller group of people with poor credit. Currently, there are some notable challenges to the study of credit scoring schemes. First, the current credit scoring research focuses on the construction of models and the selection of various classifier combinations to improve performance, without paying attention to the storage and use of credit data itself, which makes it extremely difficult to trace the whole process of using credit data, and at the same time reduces the efficiency and fails to achieve the safe use of data. Second, the current credit scoring model is too high a threshold for the practical use of institutions and individuals. Although deep learning and artificial intelligence-based classification models can provide efficient performance, the processes involved, such as model training and tuning, require a substantial amount of expertise. Third, current credit scoring models are not systematic, and each processing stage is fragmented. Therefore, building credit scoring systems to provide high-efficiency credit scoring services is an urgent challenge to be addressed. Three key problems solved in this study include:

- How to trace credit data usage records so as to prevent data leakage during the storage and transmission of credit data?

Fig. 1 Distribution of Fico credit score results

- How to build a credible, reliable, and robust cross-validation scheme, which guarantees superior performance of the credit scoring model? (Fig. 1).

In conclusion, there is a high demand for automated and efficient credit scoring systems. Finding the optimal set of hyperparameter by hand is nearly impossible, therefore using automated pipelines can result in more optimal hyperparameters and possibly even better performances. Traditional methods require a large number of human settings of model parameters, finding hyperparameters, etc., and the model efficiency is low. Therefore, each aspect of machine learning needs to be integrated so as to improve the automation of the credit scoring model.

In this paper, we propose an automated machine learning (AutoML) pipeline with blockchain called BACS that integrates the key steps of credit data storage to blockchain, data preprocessing, feature extraction, feature selection, and modeling and hyperparameter optimization, and by means of a classification pipeline to automatically mine valid information from the credit data and obtain the classification results for credit scoring. The experimental results show that our BACS method improves the performance for credit scoring task. The contributions of our work are three-fold:

- We propose a storage mechanism for credit data in the blockchain, thus providing full traceability and tamper-evident characteristics for the use of credit data.
- We propose a credit data pre-processing method that can perform proper data balancing before the data is modeled, thus improving the distribution balance of credit data.
- Finally, we propose an AutoML based classification method that integrates the processes of data sampling, feature extraction, feature selection, and hyperparameter optimisation through the pipeline method with the input being credit data and the output the credit scoring results.

The overall structure of the paper is divided into five sections. Section 2 includes a literature review of current studies on credit scoring. Section 3 deals with the methods used for our BACS scheme. Section 4 analyzes the experimental results of the classification pipeline of credit scoring and the comparison with other state-of-the-art approaches. The fifth section summarizes the contribution of this paper and discusses possible future studies.

2 Related work

A large number of effective models and systems have emerged from the current research on data mining applied to credit scoring. Tripathi et al. (2019) suggested a hybrid model that combines feature selection and a multilayer ensemble classifier architecture to improve credit score predictive performance. The suggested hybrid credit scoring methodology is organized into three stages. The first phase is preprocessing, which assigns ranks and weights to classifiers. Based on soft probability, Feng et al. (2018) introduced a new dynamic ensemble classification method for credit scoring. The classifiers in this method are first chosen based on their classification performance and the relative costs of Type I and Type II error in the validation set. Zhang et al. (2019) introduced a novel multi-stage hybrid model that combines feature selection and classifier selection to achieve the best feature subset and best classifier subset, and then utilizes classifier ensemble to improve prediction performance based on the two best subsets. Kozodoi et al. (2019) apply the application of profit measurements to feature selection and create a multi-objective wrapper framework based on the NSGA-II genetic algorithm with two fitness functions: EMP and number of features. Experiments on several credit scoring data sets show that the suggested approach produces scorecards with a greater projected profit while utilizing fewer features than conventional feature selection strategies. To increase the accuracy, Munkhdalai et al. (2020) present a hybrid credit scoring model that employs deep neural networks and logistic regression. The suggested hybrid credit scoring model is divided into two stages. They train numerous neural network models in the first phase, and then integrate those models using logistic regression in the second phase.

To obtain strong performance for credit scoring, Zhang et al. (2021) suggested an unique multi-stage ensemble model with enhanced outlier adaptability. To mitigate the negative effects of outliers in noisy credit datasets, a local outlier factor algorithm is enhanced with the bagging strategy to detect potential outliers and then boost them back into the training set to create an outlier-adapted training set that improves the outlier adaptability of base classifiers. Xia et al. (2020) suggested a novel tree-based overfitting-cautious heterogeneous ensemble model (i.e., OCHE) for credit scoring that departs from previous literature on base models and ensemble selection approach. Tree-based approaches are used in base models to achieve a compromise between forecast accuracy and computational expense. Xiao et al. (2021) provided a new framework for comparing benchmark models for imbalanced credit scoring. They also present the balanced accuracy index and four other assessment measures in the framework, empirically examine the performance of ten benchmark resampling methods and nine benchmark classification models on six credit scoring data sets, and evaluate the optimal combinations of them. For credit scoring, Liu et al. (2021) suggested a multi-grained and multi-layered gradient boosting decision tree (GBDT). Multi-layered GBDT takes into account the benefits of the explicit learning process of the tree-based model as well as the representation learning ability to discriminate good/bad applicants; multi-grained scanning augments original credit features while improving the representation learning ability of multi-layered GBDT. Dumitrescu et al. (2022) presented penalised logistic tree regression (PLTR), a high-performance and interpretable credit scoring system that utilises information from decision trees to boost the effectiveness of logistic regression. Yang et al. (2021) proposed a credit scoring model construction method that combines memetic optimization algorithm and neural architecture search to achieve efficient search of credit scoring networks.

To optimize the random forest approach, Zhang et al. (2018) present a novel credit scoring model called NCSM, which is based on feature selection and grid search. To improve prediction accuracy, the model decreases the influence of irrelevant and duplicated features.

Deng et al. (2020) present a unique approach for doing feature cross-validation that is based on a convolutional neural network. This method is intended to automatically extract significant cross features and generate cross-feature embedding from structured data, eliminating the requirement for hand-crafted cross features. Wang et al. (2018) offer a consumer credit scoring approach based on the LSTM attention mechanism, which is a novel use of the deep learning algorithm. We treat each type of event as a word, build the Event2vec model to transform each type of event transformation into a vector, and then use an attention mechanism LSTM network to predict the likelihood of user default. Zhang et al. (2020) offer a new online integrated credit scoring model (OICSM) for peer-to-peer lending. OICSM combines a gradient boosting decision tree and a neural network to improve the credit scoring model's ability to handle two types of features and update it online. To validate the effectiveness and superiority of the proposed approach, offline and online experiments are done using real and representative credit datasets.

Furthermore, automated machine learning techniques are receiving increasing attention and are widely used to solve practical classification or regression problems. Khuzani et al. (2021) developed an effective machine learning classifier that can discriminate COVID-19 cases from non-COVID-19 cases with excellent accuracy and sensitivity using an AutoML-based dimensionality reduction method. Sun et al. (2021) performed gridwise GRACE-like data reconstruction using an automated machine AutoML approach. They showed the process over the conterminous United States (CONUS) by employing six different types of machine learning models and several groupings of meteorological and climatic variables as predictors. Ikemura et al. (2021) trained several machine learning algorithms using AutoML. They chose the model that predicted patients' chances of surviving a SARS-CoV-2 infection the best. The experimental results demonstrate the effect of AutoML in improving model performance. Yang and Zou (2020) created mAML, a machine learning model-building pipeline that can construct optimal and interpretable models for personalized microbiome-based categorization tasks in a reproducible manner. This pipeline performs well on 13 benchmark datasets, covering binary and multi-class classification tasks. Therefore, automated machine learning has a beneficial effect on the performance of classification models or regression models and can improve the efficiency of building machine learning models.

It can be seen from the existing research that machine learning algorithms have been gradually applied to the study of credit scoring, and have achieved promising prediction results. However, the existing research is still focused on the classification model itself, and each part of machine learning does not integrate effectively. In addition, for a large number of non-specialists, using the methods currently studied requires a lot of knowledge in data science in feature selection, hyperparameter setting, etc., and cannot use the models for credit scoring quickly and accurately. Therefore, it can be observed that the main drawbacks of the current research on credit scoring include: (1) The models need to be constructed and tuned by humans, thus requiring a lot of human intervention; (2) The current research on credit scoring rarely takes feature engineering into account, thus missing the essential step of selecting the right features; (3) The current methods are separate for each module, thus the data needs to be manipulated and processed by humans between the different modules, which reduces the integrity of the method. manipulation and processing between them, reducing the holistic nature of the method. To address these issues, our proposed BACS method allows data to flow through the pipeline through an automated machine learning approach. The data are sequentially processed through feature extraction, feature selection, modelling, hyperparameter selection, and model evaluation, which significantly minimises human intervention in credit scoring methods.

3 Methodology of BACS

In our proposed BACS scheme, blockchain is introduced for storing credit data, thus guaranteeing full data traceability of credit scoring objects. In addition, the process of credit scoring models is efficiently integrated through a well-established machine learning pipeline approach, thus improving the efficiency and performance of credit scoring model construction. Figure 2 illustrates the workflow of the BACS approach, and we can see that the in-chain storage of credit data ensures that the data cannot be tampered with during use. In addition, the data is processed sequentially in the pipeline, thus ensuring a high degree of automation for the entire credit scoring model.

3.1 Blockchain technology

Blockchain, which is simply a public ledger running on a peer-to-peer network, is decentralized and tamper-proof, and may construct a safe and trustworthy data storage system. This method is well-known for its decentralization, transparency, and dependability (Khan et al. 2021; Vafadarnikjoo et al. 2021; Yadav et al. 2021). According to the degree of decentralization, the size of blockchain nodes, and a variety of other characteristics, blockchain can be categorized into three groups.

- (1) The public blockchain is entirely transparent. Users can join the blockchain network at any time to gain access to the public ledger's data. Bitcoin and Ethereum are the most common uses of public chains.
- (2) In comparison to the public blockchain, the consortium blockchain is less open. Only authenticated participants are allowed to join the network and view the data on the ledger. The most widely used consortium blockchain is Fabric.
- (3) A private blockchain is most commonly used within a single company or organization. It has the least amount of openness, as well as a high level of access control and authority management.

A blockchain-based data sharing platform is built in our proposed scheme to handle a uniform data interaction mechanism, which improves efficiency. It's a service platform that's powered by a blockchain consortium. On the one hand, it provides a decentralized and flexible data sharing platform, allowing the depression classification system to access all of the data. Meanwhile, no one can change the credit data on blockchain, which has excellent security protection.

3.1.1 Credit data storage on the blockchain

In order to fully portray the complex state, association relationship, ownership and other characteristics of credit information, this section treats subjects, credit information and contracts as identifiable objects in the bottom layer of blockchain, and separates credit information from contracts so that it has independent state transition control.

We establishes the expression and execution mechanism of association relationship between objects to ensure that the blockchain layer can recognize complex transaction in credit information. We also designs the complex state transition model to ensure the transition of credit information state to avoid double-spending problem. The complex state credit information is designed to avoid the occurrence of double spend problem. The blockchain consists of credit objects such as subject, credit information and contract, each of which has a

Fig. 2 Workflow of BACS. The whole pipeline consists of seven main parts: credit data storage to blockchain, data pre-processing, feature extraction, feature selection, cross-validation, model training and model evaluation

globally unique identity; the state of the information changes continuously with the transaction; the contract only acts as the controller of the credit information transition, encapsulating the transaction rules and state transition rules into the code, and the contract code is relatively constant.

Based on the key attributes such as type and ID of the credit information providers, the uniqueness identifier is designed by the provider of credit information with the following formula: $C_i \cdot UID = \text{Hash}(C_i \cdot \text{Attr}_x, \dots, C_i \cdot \text{Attr}_z)$; in the consensus process, from the block obtains the unique identifiers of all new credit information and verifies that there are no duplicate elements among them; also in ensuring that the block is free of duplicate credit information, the new credit information is verified in parallel with the existing credit scoring results. When a transaction indicates the creation of a new credit information, the smart contract generates a unique identifier based on the uniqueness formula. The smart contract generates the credit information's unique identifier based on the uniqueness identifier formula. The verification rules in the credit information contract, such as numerical type verification, status name and credit type verification, etc. etc., to prevent the emergence of multiple identical credit information caused by malicious and wrong data. After the block is executed, the unique identifier set UserIDSet of all newly created credit information in the block is obtained, and the unique identifier is verified in parallel. If there are duplicate identifiers, the verification process is terminated and the block is returned with an If there are duplicate identifiers, the verification process will be terminated and a block illegitimate message will be returned. If there are duplicate identifiers, the verification process will be terminated.

3.1.2 Consensus mechanism

The blockchain technology we use requires a consensus mechanism to determine whether the credit data is stored and used within predefined rules. The whole consensus process is roughly divided into two modules: the sorting service and the synchronized ledger. The client blockchain network initiates a transaction proposal with the client's signature and sends it to the endorsing node. After receiving the transaction proposal, the endorsing node verifies the signature and simulates the execution of the transaction according to the endorsement policy and responds with the result to the client, while the transaction is forwarded to other non-endorsing nodes (including the master node Leader). The non-endorsing nodes that receive the transaction cache the transaction locally. After receiving the correct endorsement response, the client indicates that the endorsement is successful and notifies the master node to pack the transactions in the cache. Otherwise, the endorsement fails and the client notifies each node to delete the cache, and this client submits a failed transaction. Next is the replication, validation and submission of the block.

- Step 1: Similar to the previous description, the sorting node sends the offer for the next block to other nodes. this process of propagating the offer message can be abstracted as a layered model, where at each layer a proxy node can be selected based on the completeness of the current node's block. the master node sends the offer to the nodes that are closer to it's network, and the nodes then propagate the block-generated offer message in turn in layers. In case the current node finds that the block sent by the node in layer x does not match the local cache transaction or the signature verification of the sorting node fails, it is judged that there is a Byzantine node in layer x . That is, if the current layer fails to verify the previous layer, the node at that layer communicates with the node at the previous layer of the Byzantine node, i.e., layer $x-1$, to complete the

propagation of the block generation proposal message, and continues to the next layer in turn. If after a certain number of nodes determine that the master node Leader of the first layer is evil, then each layer proxy node completes the preparation work locally and runs for Leader respectively.

- Step 2: After each node receives the block generation proposal and verifies it passes, the nodes will forward the results to each other, and if the node receives more than a certain number $2f + 1$, f is the number of Byzantine nodes) of successful verification messages, it completes the second verification and submits it locally. The secondary validation process consists of reconfirming whether the block matches the local cache transaction, and discarding it if it does not. Compare the current term and block term, if the block term is less than the current term, verify that the block fails and reject the submission; also carry out the traditional Fabric block verification process such as block height, block body hash value. In addition, the conflict verification of read/write set is carried out, and the key value change of valid transactions is update. The block will also be submitted locally after it is fully verified. When the block height of the submitted block is larger than the block height of the node's local block, it means that there is a block stub submission in the process of the previous block copy submission, and the block of a node is incomplete due to the node's downtime or network problems.

3.2 Credit data preprocessing

Credit data pre-processing is the process of re-examining and verifying credit data, including checking data consistency, handling invalid values and missing values, etc. Data cleaning steps are mainly divided into missing value processing, outlier processing, variable renaming, etc. The purpose is to remove duplicate information, correct existing errors, and provide data consistency. When the percentage of missing values is small, samples containing missing values are deleted directly. (1) When a variable has a large number of missing values and the variable is not particularly important for the problem under study, then the variable can be considered for deletion. (2) Estimation. Replace missing values with the sample mean, median, or plural of the variable. (3) Interpolation by correlation analysis or logical inference between variables.

3.3 Feature extraction

We intend to create understandable and computationally efficient models in this paper by automatically generating and selecting time series features. Rather different features are important depending on the domain and the credit data analysis. Traditionally, such features are designed by hand from credit datasets, which requires technical expertise and expert knowledge. Feature extraction is designed to obtain a large number of credit related features, and then select the most appropriate features for the overall classification task in order to establish a systematic and automated approach to generating features for credit datasets.

In our study, the overall feature mapping is a real-valued function which takes all credit data as input (for each dataset and epoch) and outputs an m -dimensional real-valued feature vector $\mathcal{F} : \mathbb{R}^{l \times n} \rightarrow \mathbb{R}^m$, where m represents the number of features, l denotes the average length of each credit dataset, and n indicates the number of credit datasets.

$$\forall(S, epoch), \mathbf{C}^{(S, epoch)} \mapsto \mathcal{F}(\mathbf{C}^{(S, epoch)}) \quad (1)$$

We'll denote the extracted feature vector ($S, epoch$) for each target and measurement epoch pair ($S, epoch$), and it'll go through the feature selection procedure.

3.4 Feature selection using Boruta

Feature selection is the method of selecting the most efficient features from the original features to reduce the dimensionality of the dataset, thus reducing the computational overhead and improving the classification performance. Since we have to select features extracted from credit datasets, it is extremely challenging to select the most suitable combination of features for the credit scoring model. We adopt the Boruta algorithm in BACS to evaluate all extracted features and select the most suitable features for the credit scoring model (Kursa et al. 2010).

The core idea of the Boruta algorithm is to evaluate the importance of each feature variable through a loop method (Kursa et al. 2010). By copying the original feature set, randomly mix each feature value to construct a random shadow feature. The final sample data set of the model is original feature and shadow feature. The primary reason we chose Boruta as the feature selection method is that the Boruta algorithm is based on the idea of a random forest classifier and incorporates randomness into the system by collecting the results of different feature combinations from a random set of samples. The randomness of Boruta will provide us with a clearer picture of the results of feature selection, providing us with a solid understanding of which features are truly important and thus keeping those significant features. For each feature in the extracted credit data, calculating Z_{score} corresponding to that feature is an important step in Boruta's algorithm and is calculated in the following way.

$$Z_{score} = \frac{\text{average_feature_importance}}{\text{feature_importance}} \quad (2)$$

After calculating the Z_{score} of all features, the maximum value of Z_{score} is taken as Z_{max} . All features with Z_{score} value higher than Z_{max} will be selected as essential features by the Boruta algorithm and be retained. Therefore, Boruta algorithm adds randomness to the system and collects results from the group of randomized samples that assists to reduce the misleading impact of random fluctuations and correlations.

3.5 Credit scoring modeling

Breiman proposed Random Forest as an integrated machine learning method in 2001 (Schonlau and Zou 2020). This approach is a classifier with several decision trees that produces the majority of all decision trees for the categories. This method is frequently used in credit data analysis because of its high classification accuracy, quick calculation speed, and ability to discover main correlation factors. The random forest's generalization error is determined by the classification strength of the forest's base classifiers and their correlation. Currently, random forest study concentrated mostly on two areas: the application of the random forest algorithm in various fields, and the parameter setting, improvement, and optimization of the random forest method. The calculation procedure is as follows: first, the training sample set is created by randomly selecting N samples from the dataset; next, each training sample produces a decision tree using random splitting attributes; finally, the N decision trees form a forest, and the final classification result is calculated by voting on the classification results of the N decision trees.

The random attribute selection is included in the decision tree training process, and when the split attribute is chosen, the normal decision picks the optimal attribute from all the

attributes of the current node. When the random forest decision tree chooses an attribute, a candidate attribute set is chosen at random from all of the attributes, and the optimal attribute is chosen from the candidate attribute set. The technique generates a series of classification models using k -round training as follows.

$$\{\gamma_1(x), \gamma_2(x), \dots, \gamma_m(x)\} \quad (3)$$

where a credit scoring classification model is made out of a series of Formula(3), the model's final classification result is generated using a simple majority voting approach, as shown below.

$$\Gamma(x) = \arg \max_y \sum_{i=1}^m \xi(\gamma_i(x) = y) \quad (4)$$

where $\Gamma(x)$ represents a combined classification model, $\gamma_i(x)$ provides a single decision tree classification model, x represents an input variable, y represents an output variable, and $\xi(x)$ indicates the indicator function. Random forests generate different training sets at random, and specific training sets generate distinct decision trees. The model classification ability is increased by merging the decision results of decision trees. The most significant advantage of using random forest in our pipeline in BACS is that each decision tree utilizes a portion of all credit data and only a subset of the attributes are extracted for modeling. This approach enhances the model diversity greatly and minimizes the correlation of each decision tree. Therefore, due to the advantages of random forests, which are less prone to overfitting, faster training and simpler to implement, we adopted the random forest to introduce into our BACS.

3.6 Hyperparameter optimization

The proper configuration of hyperparameters is critical to the performance of a machine learning model (Feurer and Hutter 2019). As a result, throughout this research, hyperparameter optimization is required to improve the performance of the model on credit dataset. It is worth noting that the search space for hyperparameter optimization contains both integer and categorical variables. Grid Search, Evolutionary Algorithms, and Bayesian Optimization are some of the well-established methods for our credit scoring task (Pedregosa et al. 2011). The Bayesian Optimization algorithm was chosen for this paper due to its efficiency when optimizing expensive problems, such as training a machine learning algorithm, which can be time consuming. Grid search is also a simple, efficient, and widely used method in hyperparameter optimization problems (Kaur et al. 2020). Another reason we adopt Bayesian optimization is that Bayesian optimization uses a Gaussian process that takes into account previous parameter information, while grid search fails to consider previous parameter information. In addition, Bayesian conditioning is fast with fewer iterations, but grid search is slow and prone to dimensionality explosion when there are many parameters.

In detail, Bayesian optimization is run for 100 iterations, with each iteration generating a candidate hyperparameter setting whose goodness is measured by the performance of the corresponding model on a test data set. The hyperparameter optimization procedure is used to enhance model performance in a 10-fold cross validation configuration, where the dataset is divided into ten parts, with nine of them rotating as training data and one as test data. The primary reason for choosing 10-fold is that a large number of experiments using different learning techniques with a large dataset have demonstrated that 10-fold is an appropriate choice for obtaining the best error estimates while ensuring efficient training as well as not causing redundant time and space wastage (Fushiki 2011; Wong and Yeh 2019).

Table 1 Description of the dataset in our study

Dataset	Good/bad	Samples	Features
Credit dataset A (German dataset)	700/300	1000	24
Credit dataset B (Taiwan)	23364/6636	30000	24
Credit dataset C (Australia)	307/383	690	14
Credit dataset D (Credit card)	600/492	1092	25

4 Experimental results and analysis

4.1 Datasets and evaluation metrics

We compare our BACS method with existing classification algorithms through extensive experiments to further illustrate the effectiveness of BACS applied to credit scoring. Meanwhile, the datasets we used include three UCI datasets and one Kaggle dataset, where three credit datasets A,B,C are from UCI and dataset D is from Kaggle, and the Table 1 shows the properties of the datasets.

Three generally used measures are used to evaluate the model in this paper: Accuracy (ACC), Specificity, and AUC. The evaluation criteria of ACC and specificity are introduced in the following manner.

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

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

We use 10-Fold validation for hyperparameter tuning. The original credit datasets are divided into 10 groups (10-Fold), and each subset of data is made into a validation set separately, and the remaining subsets of data are used as training sets, and the final errors are summed and averaged to get the cross validation errors. Compared with the 5-fold cross validation, the 10-Fold validation results are more accurate, so we choose the 10-fold cross validation in the model. After that, the performance of different methods are calculated separately to compare the performance of credit scoring models. Table 2 presents the search space of the hyperparameters involved in the random forest regressor in our pipeline.

4.2 Analysis of experimental results for different methods

In order to compare our automated machine learning pipeline approach with the currently available classification methods (Zaidi et al. 2020), we list the models to be compared for testing and comparing the performance of our propose method. Table. 3 demonstrates the advantages and disadvantages of the four commonly used classification methods in credit scoring. To evaluate our proposed BACS based on automated machine learning pipeline, we use the credit dataset and observe the performanche for different methods and time consumption of the whole process of feature extraction, feature selection, model construction, and hyperparameter extraction under different methods:

Table 2 Hyperparameter optimization result

Hyperparameter	1	2	3	4	5	6	7	8	9	10
Is bootstrap used for training samples	True	True	True	True	True	True	True	True	True	True
Max number of features when splitting a node	Auto	Auto	Auto	Auto	Auto	Auto	Auto	Auto	Auto	Auto
Max depth of each tree	8	8	9	9	10	11	9	10	10	8
The number of n_estimators	1200	800	1300	600	1000	1000	1300	900	1100	1000
The number of min_samples_split	23	18	22	18	24	17	20	16	21	19
The number of min_samples_leaf	12	10	15	9	13	10	15	9	13	9

Table 3 Comparison of four commonly used classification methods

Methods	Description	Advantages	Disadvantages
CNN	In Shin et al. (2016), CNN was originally used for image problems, but good results can be achieved in natural language processing using CNN for text classification	CNN does not require human feature selection; we simply train the model's weights to obtain the features. CNN also outperforms the competition.	CNN requires tuning and a large sample size, and training is recommended to use GPU, which consumes more time and memory
SVM	In Tsang et al. (xxxx), SVM is a binary classification model, and its basic model is a linear classifier with maximum interval defined on the feature space	SVM is effective for solving classification and regression problems with high-dimensional features, and still works well when the feature dimension is larger than the number of samples	SVM is computationally overloaded when the sample size is very large and the kernel function mapping dimension is very high, thus reducing the classification performance of the model
RF	In Shi and Horvath (2006), Random Forest (RF) is a classification method made up of many tree decisions. When generating each individual tree, it employs bagging and randomization in order to create an uncorrelated tree forest whose forecast by the committee is more reliable than that of any individual tree	The accuracy of the RF algorithm is higher than that of other algorithms due to the use of an integrated algorithm	When the random forest has a significant number of decision-making trees, the space and time required for training will be rather considerable
LSTM	In Gers et al. (2000); Sundermeyer et al. (2012), LSTM is a type of neural network that is made up of LSTM blocks or other types of neural networks. LSTM blocks are intelligent network units because they can remember values for varying periods of time, and one gate in the block can decide if the input is important enough to remember	In sequence modeling issues with extended time memory, LSTM has some advantages. It is simple to put into action. In the extended sequence training procedure, this solves the problem of gradient disappearance and gradient explosion	LSTM has a disadvantage in parallel processing, but the effect is minor in comparison to some of the most recent networks

Fig. 3 Average time consumption using different credit scoring methods

- **SVM:** Support vector machines are supervised learning models that use associated learning methods to analyze data in classification and regression analysis. The SVM training process generates a model that assigns fresh examples to one of two classes given a set of training cases, making it a non-probabilistic binary linear classifier (Mavroforakis and Theodoridis 2006).
- **CNN:** Convolutional neural networks are made up of one or more convolutional layers with a fully connected layer on top (corresponding to a classical neural network) (Li et al. 2016).
- **LSTM:** The LSTM architecture is a type of artificial recurrent neural network (RNN) that is utilized in deep learning. Breuel (2015), long-term dependencies can be learned via LSTM.

The comparison results of the system in Fig. 3 reveals that the BACS method using pipeline has less time consumption. There is more blockchain data storage and consensus process compared to other methods, but the time consumption of this part is not much and it helps the traceability process of credit data. By comparing the running time we are able to see the following.

- The running time of our BACS system is reduced compared to other methods. The main reason is that the automatic machine learning parts are efficiently connected without spending a lot of human intervention, which reduces unnecessary time consumption.
- We use a feature selection step, so that irrelevant features have been effectively removed in feature selection, thus reducing the time consumption of unnecessary features for the model.
- The introduction of blockchain has a performance improvement for credit data traceability and automatic system consensus, but the additional time consumption added is minimal, and the whole BACS system is still able to quickly complete the whole process from credit data up-chain storage to output credit scoring results.

In summary, our BACS method combines blockchain and automatic machine learning process for the first time, and is highly efficient and automated for the whole process of traceability from the use of credit data to model construction and output results. Regarding

Fig. 4 Performance comparison for accuracy

Fig. 5 Performance comparison for specificity

the model building process in the system, the introduction of neural architecture search technology design will be considered in the subsequent research to further enhance the model performance.

Figures 4, 5, 6, 7, 8 demonstrate the credit scoring performance comparison under different metrics with choosing different classification models in BACS, and the accuracy performance improvement compared to only using RF and SVM, respectively. From the results in Figs. 4, 5, 6, 7, 8, we can see that the performance of our proposed model achieves a significant advantage in different evaluation metrics. In addition to the performance of the RF model itself, the Boruta feature algorithm introduced in our automatic machine learning is itself a way to calculate the importance of different features by RF, so as to filter out the more important features. In addition, in Figs. 4, 5, 6, 7, 8, we can also find that the efficiency and performance of different models join our proposed automatic machine learning pipeline are improved, which can further prove that the introduction of blockchain technology and automatic machine learning method can improve the efficiency of credit scoring, and also

Fig. 6 Performance comparison for AUC **Fig. 7** Performance improvement compared to RF **Fig. 8** Performance improvement compared to SVM

0 1 DAYS 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

First, our BACS scheme can be applied to the financing process of micro and small enterprises, which are facing a serious credit crisis due to the global economic downturn caused by the COVID-19 epidemic. Therefore, the impact of the epidemic on the business of micro and small enterprises can be greatly reduced.

Thirdly, our BACS system was able to respond to the economic downturn and credit crisis caused by the global COVID-19 outbreak. The ongoing epidemic has caused serious

credit risks in different countries and companies, so our BACS is able to provide stable, automatic and efficient credit scoring services to financial institutions and companies in different countries, and obtaining timely credit scoring results is important to alleviate the economic crisis caused by the epidemic.

6 Conclusion

As an essential component of the credit system, the evolution of credit scoring systems plays a key role in the progress of the overall credit scoring system. With the rapid progress and popularity of big data technology, credit scoring based on big data has received widespread attention and has generated a series of credit scoring applications based on big data and machine learning. In particular, the economic downturn and credit crisis caused by the global COVID-19 epidemic has also heightened the urgent need for an efficient and stable credit scoring system. The aim of this paper is to develop a automated credit scoring classification model that reliably and quickly differentiate the credit status of credit assessors. Therefore, we propose BACS method for efficient credit scoring, including credit feature extraction, feature selection, regression model construction and model hyperparameter optimization. Experimental results demonstrate the efficient and accurate credit scoring classification performance of our BACS scheme.

Although a multitude of credit scoring methods have made some progress in the field of big data credit scoring, the data utilized in current credit scoring research is still only a fraction of big data, and many data sources, especially social data scattered in cyberspace, have not been fully utilized. How to facilitate the sharing and exchange of credit data among agencies will be an essential part of our future research. In addition, for future research we will consider optimizing the efficiency of blockchain data storage in BACS system, such as the methods utilized in Yang et al. (2021); Qi et al. (2020), to guarantee the privacy security of disparate data sources while storing credit data securely. Moreover, we will adopt credit model building mechanisms such as neural architecture search technology (Mellor et al. 2021; Yan et al. 2020) to enhance the performance and efficiency of classification models in the BACS scheme.

Acknowledgements This research is supported by the National key R&D Program of China under Grant No. 2018YFB1402700. This work is also supported by the Fundamental Research Funds for the Central Universities, China under Grant No. xzy022020056. This work is also supported by a scholarship from the China Scholarship Council (CSC) under Grant No. 201906280499. This work is also supported by the the Blockchain Core Technology Strategic Research Program, China under Grant No. 2020KJ01080.

References

- Brueel, T. M. (2015). Benchmarking of LSTM networks. arXiv preprint [arXiv:1508.02774](https://arxiv.org/abs/1508.02774).
- Deng, S., Li, R., Jin, Y. & He, H. (2020). CNN-based feature cross and classifier for loan default prediction. In 2020 International conference on image, video processing and artificial intelligence, Vol. 11584, International Society for Optics and Photonics (p. 115841K).
- Doumplos, M., & Zopounidis, C. (2007). Model combination for credit risk assessment: A stacked generalization approach. *Annals of Operations Research*, 151(1), 289–306.
- Dumitrescu, M., Hue, S., Hurlin, C., & Tokpavi, S. (2022). Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. *European Journal of Operational Research*, 297, 1178–1192.

- Feng, X., Xiao, Z., Zhong, B., Qiu, J., & Dong, Y. (2018). Dynamic ensemble classification for credit scoring using soft probability. *Applied Soft Computing*, 65, 139–151.
- Feurer, M., & Hutter, F. (2019). Hyperparameter optimization. In: Automated machine learning (pp. 3–33). Springer, Cham.
- Fushiki, T. (2011). Estimation of prediction error by using k-fold cross-validation. *Statistics and Computing*, 21(2), 137–146.
- Gaganis, C., Papadimitri, P., & Tasiou, M. (2021). A multicriteria decision support tool for modelling bank credit ratings. *Annals of Operations Research*, 306(1), 27–56.
- Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. *Neural Computation*, 12, 2451–2471.
- Ikemura, K., Bellin, E., Yagi, Y., Billett, H., Saada, M., Simone, K., Stahl, L., Szymanski, J., Goldstein, D., & Gil, M. R. (2021). Using automated machine learning to predict the mortality of patients with covid-19: Prediction model development study. *Journal of medical Internet research*, 23(2), e23458.
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260.
- Kaur, S., Aggarwal, H., & Rani, R. (2020). Hyper-parameter optimization of deep learning model for prediction of Parkinson's disease. *Machine Vision and Applications*, 31(5), 1–15.
- Khan, S. A. R., Godil, D. I., Jabbar, C. J. C., Shujaat, S., Razzaq, A., & Yu, Z. (2021). Green data analytics, blockchain technology for sustainable development, and sustainable supply chain practices: Evidence from small and medium enterprises. *Annals of Operations Research* 1–25.
- Khuzani, A. Z., Heidari, M., & Shariati, S. A. (2021). Covid-classifier: An automated machine learning model to assist in the diagnosis of covid-19 infection in chest x-ray images. *Scientific Reports*, 11(1), 1–6.
- Kozodoi, N., Lessmann, S., Papakonstantinou, K., Gatsoulis, Y., & Baesens, B. (2019). A multi-objective approach for profit-driven feature selection in credit scoring. *Decision Support Systems*, 120, 106–117.
- Kursa, M. B., Jankowski, A., & Rudnicki, W. R. (2010). Boruta—A system for feature selection. *Fundamenta Informaticae*, 101(4), 271–285.
- Li, Y., Hao, Z., & Lei, H. (2016). Survey of convolutional neural network. *Journal of Computer Applications*, 36(9), 2508–2515.
- Liu, W., Fan, H. & Xia, M. (2021). Multi-grained and multi-layered gradient boosting decision tree for credit scoring. *Applied Intelligence* 1–17.
- Mahbobi, M., Kimiagari, S., & Vasudevan, M. (2021). Credit risk classification: An integrated predictive accuracy algorithm using artificial and deep neural networks. *Annals of Operations Research* 1–29.
- Mavroforakis, M. E., & Theodoridis, S. (2006). A geometric approach to support vector machine (SVM) classification. *IEEE Transactions on Neural Networks*, 17(3), 671–682.
- Mellor, J., Turner, J., Storkey, A., & Crowley, E. J. (2021). Neural architecture search without training. In *International conference on machine learning*, PMLR (pp. 7588–7598).
- Munkhdalai, L., Lee, J. Y., & Ryu, K. H. (2020). A hybrid credit scoring model using neural networks and logistic regression. In: Advances in intelligent information hiding and multimedia signal processing (pp. 251–258). Springer.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. et al. (2011). Scikit-learn: Machine learning in python. *The Journal of machine Learning research*, 12 2825–2830.
- Qi, S., Lu, Y., Zheng, Y., Li, Y., & Chen, X. (2020). CPDS: Enabling compressed and private data sharing for industrial internet of things over blockchain. *IEEE Transactions on Industrial Informatics*, 17(4), 2376–2387.
- Schönlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *The Stata Journal*, 20(1), 3–29.
- Shi, T., & Horvath, S. (2006). Unsupervised learning with random forest predictors. *Journal of Computational and Graphical Statistics*, 15(1), 118–138.
- Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., et al. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, 35(5), 1285–1298.
- Sun, A. Y., Scanlon, B. R., Save, H. & Rateb, A. (2021). Reconstruction of grace total water storage through automated machine learning. *Water Resources Research* 57(2), e2020WR028666.
- Sundermeyer, M., Schlüter, R., & Ney, H. (2012) LSTM neural networks for language modeling. In *Thirteenth annual conference of the international speech communication association*.
- Tripathi, D., Edla, D. R., Cheruku, R., & Kuppli, V. (2019). A novel hybrid credit scoring model based on ensemble feature selection and multilayer ensemble classification. *Computational Intelligence*, 35(2), 371–394.

- Tsang, I. W., Kwok, J. T., Cheung, P.-M., & Cristianini, N. Core vector machines: Fast SVM training on very large data sets. *Journal of Machine Learning Research*, 6(4).
- Vafadarnikjoo, A., Ahmadi, H. B., Liou, J. J., Botelho, T., & Chalvatzis, K. (2021). Analyzing blockchain adoption barriers in manufacturing supply chains by the neutrosophic analytic hierarchy process. *Annals of Operations Research*, 1–28.
- Wang, C., Han, D., Liu, Q., & Luo, S. (2018). A deep learning approach for credit scoring of peer-to-peer lending using attention mechanism LSTM. *IEEE Access*, 7, 2161–2168.
- Wong, T.-T., & Yeh, P.-Y. (2019). Reliable accuracy estimates from k-fold cross validation. *IEEE Transactions on Knowledge and Data Engineering*, 32(8), 1586–1594.
- Xia, Y., Zhao, J., He, L., Li, Y., & Niu, M. (2020). A novel tree-based dynamic heterogeneous ensemble method for credit scoring. *Expert Systems with Applications*, 159, 113615.
- Xiao, J., Wang, Y., Chen, J., Xie, L., & Huang, J. (2021). Impact of resampling methods and classification models on the imbalanced credit scoring problems. *Information Sciences*, 569, 508–526.
- Yadav, V. S., Singh, A., Raut, R. D., & Cheikhrouhou, N. (2021). Blockchain drivers to achieve sustainable food security in the Indian context. *Annals of Operations Research*, 1–39.
- Yan, S., Zheng, Y., Ao, W., Zeng, X., & Zhang, M. (2000). Does unsupervised architecture representation learning help neural architecture search? *Advances in Neural Information Processing Systems* 33.
- Yang, F., Qiao, Y., Huang, C., Wang, S., & Wang, X. (2021). An automatic credit scoring strategy (ACSS) using memetic evolutionary algorithm and neural architecture search. *Applied Soft Computing*, 113, 107871.
- Yang, F., Qiao, Y., Wang, S., Huang, C., & Wang, X. (2021). Blockchain and multi-agent system for meme discovery and prediction in social network. *Knowledge-Based Systems*, 229, 107368.
- Yang, F., & Zou, Q. (2020). maml: an automated machine learning pipeline with a microbiome repository for human disease classification. Database.
- Zaidi, T., Samy, A., Kocatürk, M., & Ateş, H. F. (2020). Learned vs. hand-crafted features for deep learning based aperiodic laboratory earthquake time-prediction. In *2020 28th signal processing and communications applications conference (SIU)* (pp. 1–4). IEEE.
- Zhang, W., He, H., & Zhang, S. (2019). A novel multi-stage hybrid model with enhanced multi-population niche genetic algorithm: An application in credit scoring. *Expert Systems with Applications*, 121, 221–232.
- Zhang, W., Yang, D., Zhang, S., Ablanedo-Rosas, J. H., Wu, X., & Lou, Y. (2021). A novel multi-stage ensemble model with enhanced outlier adaptation for credit scoring. *Expert Systems with Applications*, 165, 113872.
- Zhang, X., Yang, Y., & Zhou, Z. (2018). A novel credit scoring model based on optimized random forest. In *2018 IEEE 8th annual computing and communication workshop and conference (CCWC)* (pp. 60–65). IEEE.
- Zhang, Z., Niu, K., & Liu, Y. (2020). A deep learning based online credit scoring model for P2P lending. *IEEE Access*, 8, 177307–177317.