



# Machine learning techniques for credit risk evaluation: a systematic literature review

Siddharth Bhatore<sup>1</sup> · Lalit Mohan<sup>1</sup> · Y. Raghu Reddy<sup>1</sup>

Received: 13 September 2019 / Accepted: 15 May 2020 / Published online: 10 June 2020  
© Institute for Development and Research in Banking Technology 2020

## Abstract

Credit risk is the risk of financial loss when a borrower fails to meet the financial commitment. While there are many factors that constitute credit risk, due diligence while giving loan (credit scoring), continuous monitoring of customer payments and other behaviour patterns could reduce the probability of accumulating non-performing assets (NPA) and frauds. In the past few years, the quantum of NPAs and frauds have gone up significantly, and therefore it has become imperative that banks and financial institutions use robust mechanisms to predict the performance of loans. The past two decades has seen an immense growth in the area of artificial intelligence, most notably machine learning (ML) with improved access to internet, data, and compute. Whilst there are credit rating agencies and credit scoring companies that provide their analysis of a customer to banks on a fee, the researchers continue to explore various ML techniques to improve the accuracy level of credit risk evaluation. In this survey paper, we performed a systematic literature review on existing research methods and ML techniques for credit risk evaluation. We reviewed a total of 136 papers on credit risk evaluation published between 1993 and March 2019. We studied the implications of hyper parameters on ML techniques being used to evaluate credit risk and, analyzed the limitations of the current studies and research trends. We observed that Ensemble and Hybrid models with neural networks and SVM are being more adopted for credit scoring, NPA prediction and fraud detection. We also realized that lack of comprehensive public datasets continue to be an area of concern for researchers.

**Keywords** Credit risk · Machine learning · Credit scoring · NPA prediction · Fraud detection

## 1 Introduction

Financial stability is sine qua non for sustained and rapid economic progress of banks. Among various indicators of financial stability, banks' assume critical importance on the asset or loan quality, credit risk and efficiency in the allocation of resources to productive sectors. Credit risk can be due to credit default, concentration of exposure to an industry or individual and, sovereign's unwilling or inability to meet obligations. Credit risk evaluation is a continuous activity that starts with underwriting a potential loan till

payments collection after on-boarding the loan. With growing uncertainty in the economy, the political and legal will to deal with defaulters, default prediction is a necessity to ensure that right practices of credit risk evaluation is practiced. Banks try to mitigate associated risks by insurance, covenants, diversification, risk based pricing, etc. Due to the global financial crisis of 2007–2008 and the regulatory concerns of Basel norms, credit risk evaluation has become a major focus of banking and financial industry. According to IMF, world economy is at the risk of another financial crisis [1] due to ineffective credit risk evaluation. Typically, credit risk is measured based on capacity to repay, capital, loans' conditions, credit history, and associated collateral. Jarrow Turnbull [2] proposed one of the first reduced form models for ascertaining credit risk. Banks and financial services rely on credit risk departments and engage agencies such as Standard and Poor's, Moody's, Fitch and others to perform credit risk evaluation for a fee. There are also credit scoring companies such as Experian, Equifax, CIBIL and others that focus on past credit history on financial contracts of a

✉ Siddharth Bhatore  
siddharth.bhatore@research.iiit.ac.in

Lalit Mohan  
lalit.mohan@research.iiit.ac.in

Y. Raghu Reddy  
raghu.reddy@iiit.ac.in

<sup>1</sup> IIIT Hyderabad, Hyderabad, India

customer. We focus on the credit risk evaluation that is associated with credit scoring of new loan, non-performing asset (NPA) prediction and fraud detection of existing loans, as these are risks associated with principal and interest or both that banks gave to good or bad customers. We do not focus on concentration and sovereignty risks for the current study as they are not necessarily related to customer behaviour.

NPA is a loan (including a leased loan) that becomes non performing when it ceases to generate income for the bank. With changes in global economies and inter dependent economies, an economic sneeze in one country leads to economic meltdown in another country. While Cyprus [3] leads the countries with approximately 46% NPA, in large economies, India is ranked first with NPA concerns and ranked fifth highest in the world [4]. The NPA accumulated by Indian lenders (approximately 114 Billion USD in 2018) is higher than those of banks in the major economies which include China, Japan, the US and UK. Countries with higher NPA ratios (to the total loans) than India's are part of the distressed PIIGS group— Portugal, Italy, Ireland, Greece, and Spain. India's gross NPAs stood at 18.77% by end of March 2019 for public sector banks and 14.33% for all scheduled commercial banks, lesser than the previous year due to recent initiatives including set up of Central Repository of Information on Large Credits, Asset Quality Review and Insolvency and Bankruptcy Code. Reserve Bank of India (RBI—India's central bank) continues to caution on gross NPAs and asset quality.<sup>1</sup> The process of income recognition is objective and based on record of recovery rather on any subjective considerations. Likewise, asset classification of banks is on the basis of objective criteria to ensure that the norms are applied uniformly and consistently. Also, the provisioning is on the basis of classification of assets, the period for which the asset has remained non-performing, the availability of security and the realizable value thereof. A report by India's Central Bank—RBI has identified 17 reasons for NPA.<sup>2</sup> Some of the rules for a loan or an advance to be considered as NPA are:

- “Interest and/ or installment of principal remain overdue for a period of more than 90 days in respect of a term loan”,
- “The account remains ‘out of order’ (outstanding balance remains continuously in excess of the sanctioned limit/ drawing power) for a period of more than 90 days”,
- “An overdraft or cash credit bill remains overdue for a period of more than 90 days in the case of bills purchased and discounted”,

- “Interest and/or installment of principal remains overdue for two harvest seasons but for a period not exceeding two half years in the case of an advance granted for agricultural purposes”, and
- “With respect to accounts, any amount to be received remains overdue for a period of more than 90 days.”

While rule based mechanisms are fairly successful, the mechanisms do not identify spending patterns, in/out cash flow, repayment patterns, new advances/loans, deposits and other collateral changes, seasonal, economic and political changes, ownership changes, and other social behavior of customer to predict a potential NPA. There also have been cases of window dressing of potential NPAs with overdrafts and short term advances to circumvent NPA classification. Identifying these patterns, guides banks for proactive intervention and take rightful steps including loan restructuring. Importance of pattern based recognition of anomalies, consumer behavior is studied by industry [5, 6] and academicians [7–10].

To reduce potential NPAs, it is crucial that credit score is evaluated more diligently. One of the key decisions of banking and financial institutions is to decide whether or not to grant a loan to a customer. Banks granting loans and advances are expected to have realistic repayment schedules based on borrowers' cash flows. The subprime mortgage crisis<sup>3</sup> between 2007–2010 was due to ineffective credit scoring while granting loans. A smart bank considers numerous factors to give a loan to a customer, for example, a coal company may not be given loans by banks for large expansion as there are global concerns over reducing the carbon emission. Credit scoring is an important analytical technique in credit risk evaluation based on customer history and environment factors. It becomes a binary or multi class problem to distinguish low credit risk customers from high credit risk ones. Some likely scenarios that state the need of good or bad Loan identification are

- “Lakshmi is an urban low-income factory worker with dependency on monsoon. She is in her mid thirties and stays in suburbs of a metro city with husband and two kids. Her family stays in a small rented house of 300 sqft. from last three years with basic amenities like TV, gas stove, etc. She has total work experience of 9 years and is working in the current organization for the past 18 months. Her monthly salary is 18000/- and she and her family has health coverage from company insurance. She requires personal loan for her kids' higher studies.”
- “Prasad is newly married, working as a software professional. Requires consumer loan to buy household furni-

<sup>1</sup> <https://tinyurl.com/y5ft3hd7>.

<sup>2</sup> <https://tinyurl.com/RBINPA>.

<sup>3</sup> <https://tinyurl.com/peupj3s>.

ture and new style motor cycle. His company has dependency on US visas that is going through lot of rejection.”

- “Roadside grocery seller has seen onion dumping in large volume at prices less than the cost of production. The farmer has taken farm loan and likes to get rid of the stock before onions get rotten.”

It is also likely that customers and banks can fall prey to fraudulent transactions. Examples of fraud include credit card fraud, insurance fraud, accounting fraud, etc. Fraud attempts have seen a drastic increase in recent years with increased digitization, making fraud detection more important than ever to reduce credit risk. As identified by Steve Albrecht [11], perceived pressure, opportunity and rationalization are the primary reasons for committing fraud. Financial frauds involve complex transactions with involvement of ‘white collar criminals’. Cyber, social engineering, mortgage application, merchant, rogue trading, financial statement cook-up, currency and other are some of the common fraud topologies. The recent reports of a central bank state that frauds have increased by 72% in the banks.<sup>4</sup> Along with external frauds, unfortunately, frauds due to internal employees are also increasing at more than 20% [12]. Some of the common frauds committed by customers and internal employees are

- “Customers create fake liens to obtain loans or obtain more than the prescribed amount.”
- “Employees collude with customers to provide loans at lower interest rates and unverified collaterals in return of bribe or favor.”
- “Customers use the loan amount for meeting their desires or other trivial activities rather using it for the purpose it was lent.”

Our focus in the paper is not related to frauds that are based on social engineering, compromise of accounts, cyber attacks and others that are not conspired by customers. In the study, we also do not focus on the fraudulent money laundering for terrorism and other illegal activities though they continue to be an area of concern for regulators, financial investigation units of countries and UN.

With the unprecedented growth in both banking and payment transactions in digital form, small or large banks, micro finance, self help groups and other financial institutions are becoming repositories of large volumes of varied data piling up at a great velocity. The cost of compliance, regulatory reforms and risk management needs careful attention for simplifying business functions. The increase in availability of data can lead to more informed

decisions provided the data is analyzed quickly and meaningfully. Existing data warehouses, MIS and other reports are becoming less important with emergence of pattern based data (structured and unstructured) analytics. Currently, banks with the help of credit rating agencies have started to use intelligence built over a period of time in the form of rules, statistics or pattern recognition techniques to perform credit risk evaluation. Even regulators such as European Central Bank [13] and others have suggested features based on structured and unstructured (natural language) data for early warning on credit risk. Data analytics involves complex processing that goes beyond statistics, into the field of computer science (via machine learning subsuming new wave of artificial intelligence) and operations research. Dr. Jim Gray of Microsoft, refers Artificial Intelligence (AI) as the fourth paradigm [14] of science with theoretical, experimental and computational paradigms that preceded it in the evolution of science. With AI, the hidden patterns are recognized and appropriate alerts are raised in a useful and usable way. According to IBM’s 2010 Global Chief Executive Officer Study,<sup>5</sup> 89% of banking and financial markets CEOs state that their top priority is to better understand, predict and give customers what they want leveraging analytics. Some of the broad use cases that analytics [15, 16] are expected to help with are: Acquiring customers, serving the extant customers and making them profitable, Targetted marketing, Market basket analysis (Cross-sell/Up-sell), Churn prediction including feedback processing, Customer sentiment analysis, Market risk (subsuming foreign exchange rate, interest rate and liquidity risks) modeling, Automated Teller Machine cash replenishment modeling, Productivity/profitability based ranking of banks, Portfolio optimization, Application screening and Channel optimization. More specifically, they can be used for (i) Credit scoring (ii) NPA prediction (iii) Fraud detection (transactional and non-transactional) and other use cases. In the recent times, Fintechs are giving a new vigor to innovation in banking and financial services with automation and customer experience. Fintechs are evaluating creditworthiness of loan applicants, and improve the interface between customers and their service providers [17]. Fintechs are also providing lending platforms for unsecured loans [18] while evaluating creditworthiness in few seconds leveraging various machine learning techniques consuming structured and unstructured data. However, the ML techniques and the information on hyper parameters is not available to public or research community. While there has been extensive research in industry and academia on credit scoring, NPA

<sup>4</sup> <https://tinyurl.com/y2vzvdqx>.

<sup>5</sup> <https://www-03.ibm.com/press/us/en/pressrelease/31670.wss>.

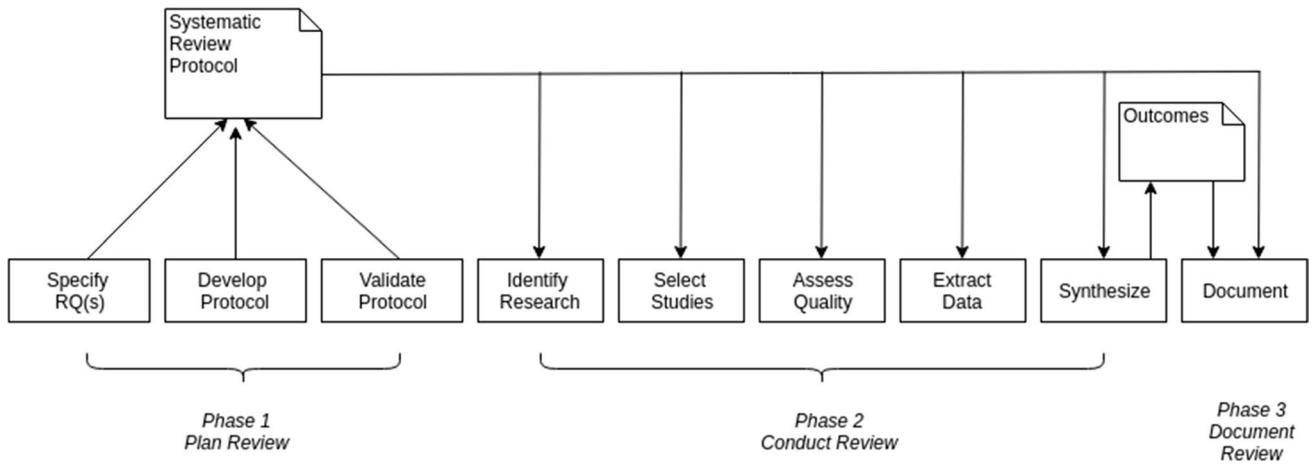


Fig. 1 Systematic literature review process [19]

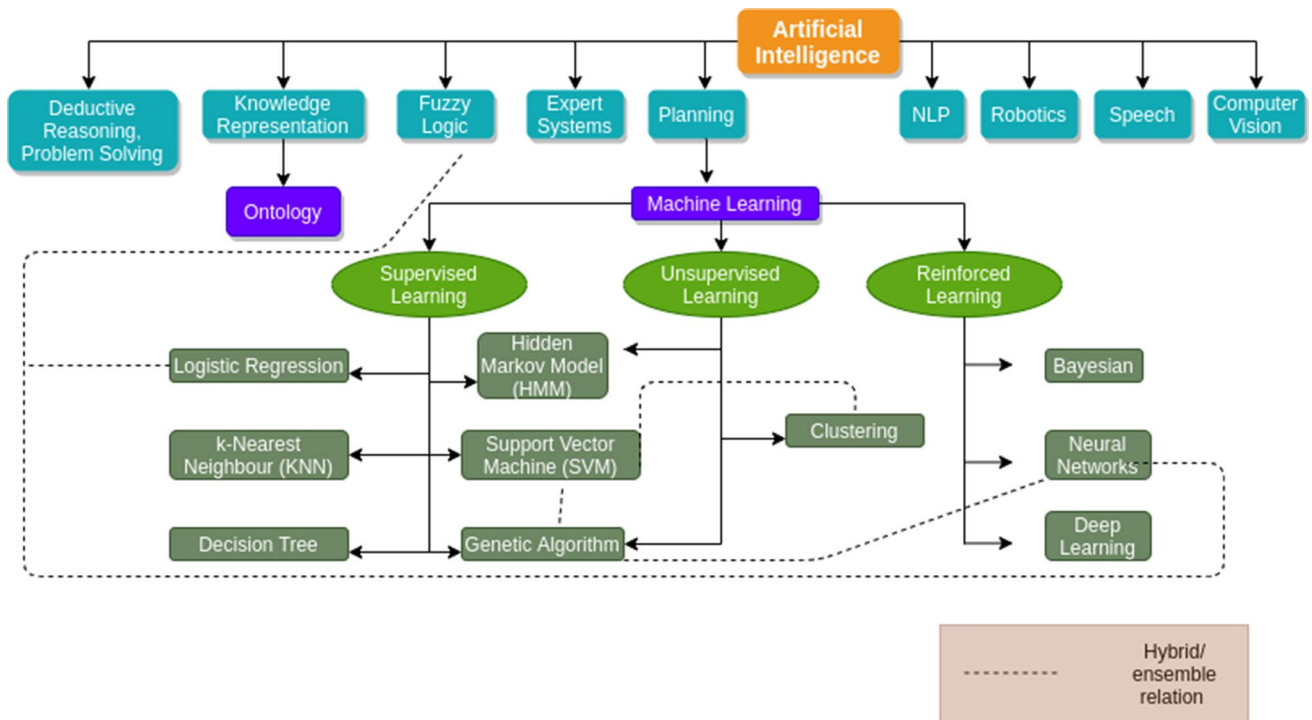


Fig. 2 Artificial intelligence landscape [20]

identification and fraud detection with rule based, statistical and pattern based approaches; we did not find any consolidated literature that discusses datasets, challenges and research gaps. In the remainder of this paper, we focus on literature gathered from academic and industry publications on credit scoring, NPA and Fraud. We perform a comprehensive systematic literature review so that various approaches and techniques can be studied. We further

analyze the shortlisted papers after applying inclusion and exclusion criteria and synthesize our findings.

## 2 Approach for literature review

There is extensive literature on various approaches on credit scoring, NPA prediction and fraud detection. The growing concerns and increasing benefits of automation on credit risk

evaluation provided us the motivation for conducting this study. We use the review process suggested by Kitchenham [19] to conduct the systematic literature review (SLR). The overview of the process is shown in Fig. 1. Following are the research questions that would be addressed by this study along with the rationale (Ra.) behind including them in the study:

- RQ1. Why and what are the AI techniques being used for credit risk evaluation?
- Ra. AI includes natural language processing (NLP), Machine learning (ML), information retrieval and extraction, expert systems, fuzzy logic and other approaches are shown in Fig. 2. While our primary interest lies in ML techniques, having broader research question would help us to analyze the advantages and disadvantages of ML and gaps in other techniques.
- RQ2. How are the ML techniques being used for credit risk evaluation?
- Ra. This research question can provide us an insight into various ML approaches including probabilistic, neural networks, optimization and ensemble based, for credit risk evaluation. This is also expected us to understand if there are any commonalities (models, feature extraction, datasets, etc.) within credit scoring, NPA prediction and fraud detection. The response to this question can provide us the scoring techniques (binary or a scale) for credit risk evaluation. The identified public datasets can be used by interested researchers in the area to improve their algorithms.
- RQ3. What are the challenges/limitations in this focus area?
- Ra. The identified challenges can help us to identify threats to validity associated with application in credit risk evaluation of various ML models, scoring techniques, regional (legal and compliance) and product specific issues. The usage of loss factor and hyper parameters will help us understand the implementation of ML techniques.
- RQ4. What are the research trends in credit risk evaluation?
- Ra. This can provide insights and potential guidance to researchers interested in credit risk evaluation.
- RQ5. Which are the universities that are working in this area?
- Ra. It is possible that researchers working in this area might want to get in touch with the universities and researchers working in the same area. This can enable researchers and industry to possibly collaborate on datasets and algorithms to improve credit risk evaluation.

A systematic review protocol is a documented plan describing all the details about how a review will be

conducted. We used a living document that was continuously updated during the review process. This protocol was used as a reference document by the reviewers and was evaluated by other fellow researchers with in our institute to provide feedback about the design of the study.

## 2.1 Search strategy and study selection

We used databases such as Springer, ScienceDirect, IEEE Xplore and ACM Digital Library to gather the relevant literature based on search query or search string. These databases were chosen as they cover most of the important journals and conferences. We did not perform any search on repositories such as Wiley, Taylor and Francis, IGI and others due to limited access and also the cost involved. The first step towards finding relevant studies is to find relevant keywords. As we are interested in only bank related credit risks, our first keyword is “banking”. The keywords that follow are “credit risk”, “credit score”, “default”, “NPA”, “Non Performing Asset”, “Non Performing Loan” and “Fraud Detection”. We identified the related keywords (synonyms) leveraging investopedia.<sup>6</sup> Then, we have a set of keywords that are related to the techniques for doing credit risk evaluation or models, such as : “AI”, “artificial intelligence”, “ML”, “Machine Learning”, “classification”, “Supervised”, “unsupervised”, “Deep Learning”, “Neural Network”, “Radial Basis Function Networks”, “SVM”, “Decision Tree”, “Discriminant Analysis”, “Naive Bayes”, “Nearest Neighbor”, “Random Forest”, “Hidden Markov”, “Markov Chain”, “Regression”, “Fuzzy Logic” and “Expert System”. The following query was formed to identify studies of relevance from the databases with the keywords :

(“banking”) AND (“credit score” OR “credit risk” OR “default” OR “NPA” OR “non performing asset” OR “non performing loan” OR “fraud detection”) AND (“AI” OR “artificial intelligence” OR “ML” OR “machine learning” OR “classification” OR “supervised” OR “unsupervised” OR “deep learning” OR “neural network” OR “radial basis function networks” OR “SVM” OR “support vector machine” OR “decision tree” OR “discriminant analysis” OR “naive bayes” OR “nearest neighbor” OR “random forest” OR “hidden markov” OR “markov chain” OR “regression” OR “fuzzy logic” OR “expert system”)

Let  $\Delta$  = (“banking”) AND (“credit score” OR “credit risk” OR “default” OR “NPA” OR “non performing asset” OR “non performing loan” OR “fraud detection”)

As each of the databases had their own query formats, we had to modify or breakdown the queries. For ACM

<sup>6</sup> <https://www.investopedia.com>.



Digital Library, the query was broken down into two parts due to limitations in query string size. The following queries were input and the results were compiled:

Δ AND (“AI” OR “artificial intelligence” OR “ML” OR “machine learning” OR “classification” OR “supervised” OR “unsupervised” OR “deep learning” OR “neural network” OR “radial basis function networks” OR “SVM” OR “support vector machine” OR “decision tree” OR “discriminant analysis”)

Δ AND (“naive bayes” OR “nearest neighbor” OR “random forest” OR “hidden markov” OR “markov chain” OR “regression” OR “fuzzy logic” OR “expert system”)

We used the original query for IEEE Xplore as the entire query string could be accommodated. As the limit for the size of query was low for ScienceDirect, following queries were used separately and the consequent results were compiled:

Δ AND (“AI” OR “artificial intelligence” OR “ML” OR “machine learning” OR “classification” OR “supervised”)

Δ AND (“unsupervised” OR “deep learning” OR “neural network” OR “radial basis function networks” OR “SVM”)

Δ AND (“support vector machine” OR “decision tree” OR “discriminant analysis” OR “naive bayes”)

Δ AND (“nearest neighbor” OR “random forest” OR “hidden markov” OR “markov chain” OR “regression”)

Δ AND (“fuzzy logic” OR “expert system”)

For Springer Library, we divided the search query into two parts because of limit set by the database. Following are the queries and their results were compiled:

Δ AND (“AI” OR “artificial intelligence” OR “ML” OR “machine learning” OR “classification” OR “supervised” OR “unsupervised” OR “deep learning” OR “neural network” OR “radial basis function networks” OR “SVM” OR “support vector machine” OR “decision tree” OR “discriminant analysis”)

Δ AND (“naive bayes” OR “nearest neighbor” OR “random forest” OR “hidden markov” OR “markov chain” OR “regression” OR “fuzzy logic” OR “expert system”)

We also did a manual search to find the seminal or important studies in this field. This was to ensure that all important studies are included in our search.

## 2.2 Inclusion, exclusion criteria and quality assessment

Here we describe the inclusion, exclusion and quality assessment criteria on the search results to identify the relevant papers. The literature is included if:

- It discussed new or improvements to existing techniques of credit risk evaluation including credit scoring, NPA prediction and fraud detection. In addition to studying many different techniques, this helped filter out repetition, especially the ones that discuss the same techniques.
- It is published between 1993 and March 2019. We have chosen studies after the AI boom in 1980’s and the subsequent AI winter that followed till 1993 [21]. This way we made sure that the studies chosen fall in the era of modern artificial intelligence when there are several studies being published on how artificial intelligence can be used to solve world problems.
- We also followed the forward-snowballing approach to identify the relevant studies. As a consequence, some additional papers were chosen as a part of our study.

The literature is excluded from the selection process if:

- Poster, short paper, doctoral symposium paper, thesis or dissertation, or grey literature are removed from evaluation. To maintain the scope of this study, we have confined ourselves to only full research papers.
- It is not written in English. Almost all major contributions in this area have their texts available in English.
- The full-text is not available or accessible.
- Papers that had reference to AI techniques but dealing with rules and statistics such as time series, regression, correlation and other methods were also excluded from the study.
- Papers that are related to fraud detection but did not involve conspiracy by fraudulent customers.
- Papers that were not published in Computer Science conferences and journals were removed. Conference proceedings that are not related to “Artificial Intelligence” were also removed during quality assessment.

To ensure quality papers are only included as part of the review, we included the following steps in our process

- It is published between 1993 and 2014 but lacks even a single citation (without considering self citation). The minimum citation count was set to one to ensure a bare minimal standard for the literature being studied. A relaxation is given to studies from last five years and they are included even if there is not a single citation to them. This was done in the light that research papers in

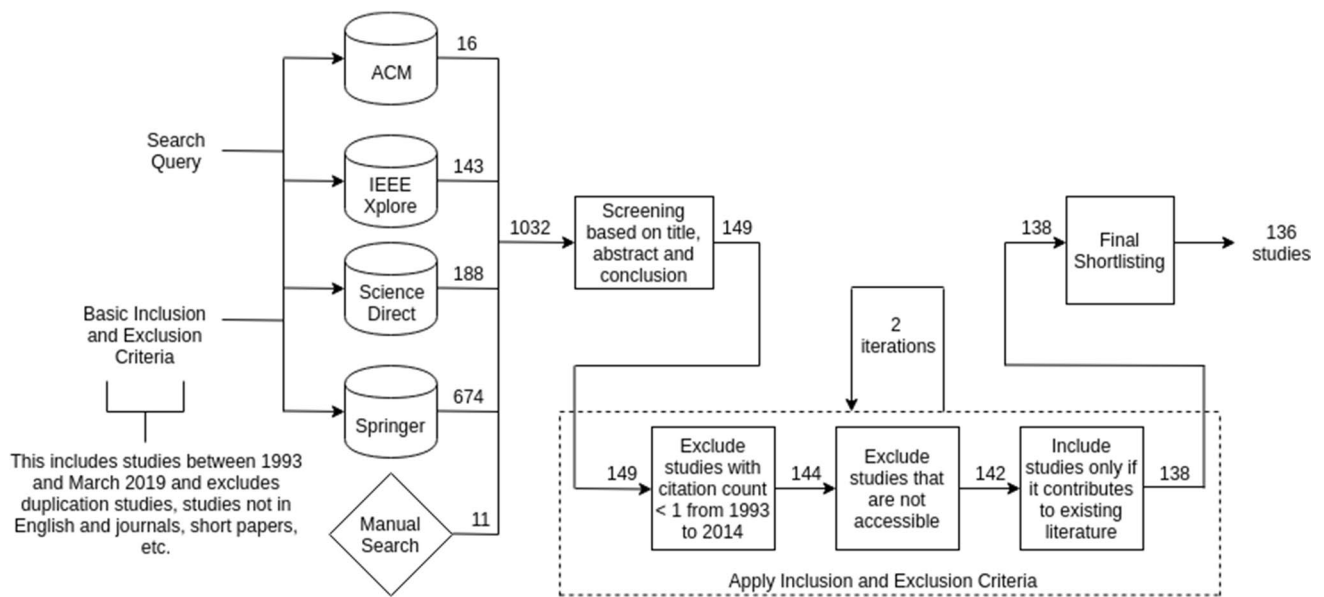


Fig. 3 Results of the search process

this area usually take some time to get noticed. We set the mark to five years based on our experience with other SLRs.

- We also did not include papers that are not peer reviewed research.
- It is a duplication study. i.e. it is found in other parts of the searching process or published in other sources.
- Papers that do not contain research objectives, experimental rigor and lack validation were excluded from study due to lack of quality in the research.
- Authors also did a manual search to find the seminal or impressive studies in this field. This was to ensure that the automated search did not exclude any important relevant studies. Studies were chosen according to their number of citations.
- Authors formed an internal review team to perform quality assessment on search criteria and search results. The authors met after every step in the SLR process to analyze the issues on hand. Emails and spreadsheets were used for recording the findings and observations.

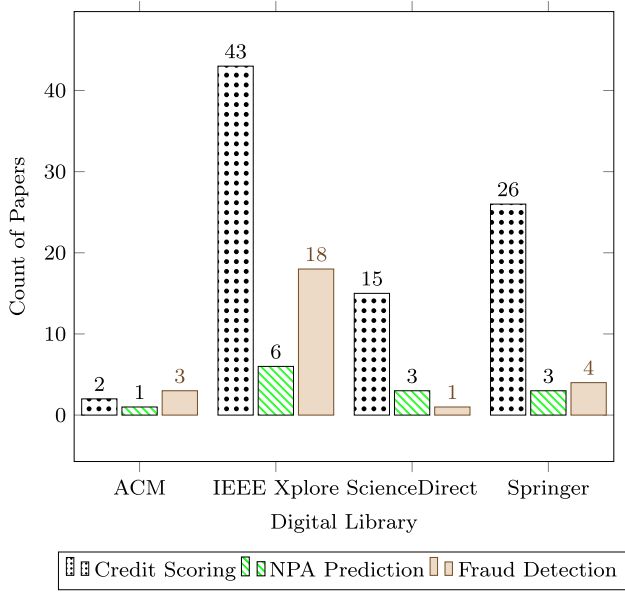
### 3 Results

The first step of the search process is to apply the search query along with basic inclusion and exclusion criteria to the search databases. Step two is to screen the resulting studies on the basis of their title, keywords, abstract and conclusion. Step three is applying remaining inclusion and exclusion criteria along with quality assessment. This involves critically going through the studies to see if they contribute to

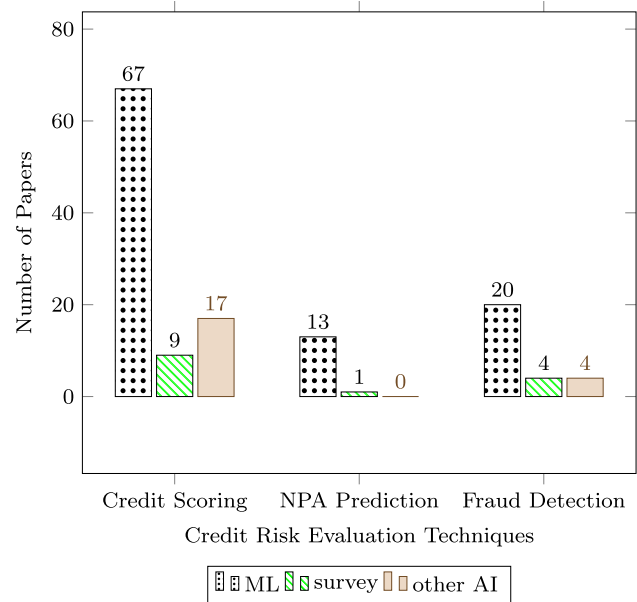
the existing methods on credit risk evaluation. The studies are also excluded if they don't have a single citation and is published between 1993 and 2014. Step three is repeated a fixed number of times to make sure that all the studies are relevant to our systematic review.

A total of 136 papers were shortlisted for review in our study. The results of the search process in terms of how the papers were shortlisted is shown in Fig. 3. The numbers in the figure refer to the number of studies shortlisted after each step. The initial studies and the results of the shortlisted studies can be accessed from [22]. After applying basic inclusion and exclusion criteria along with the search query on digital libraries and including the manual search results, we got 1032 studies. As can be seen in Fig. 3, we had a large number of studies from Springer library whereas only 16 studies from the ACM Digital Library. This is because Springer has a large collection of scientific journals from which we get our search results, there were duplicates among conference papers and book chapters. These studies were screened based on their title, abstract, keywords and conclusion, leaving us with 149 studies. Remaining inclusion and exclusion criteria were applied on these 149 studies such as papers that did not involve fraud detection papers conspired by customers. After repeating this step for 2 iterations with co-authors as reviewers, we were left with 138 studies. Then a final shortlisting is done on the 138 studies leaving us with 136 studies. The final numbers after the iterations are written in the figure after applying each step in inclusion and exclusion criteria.

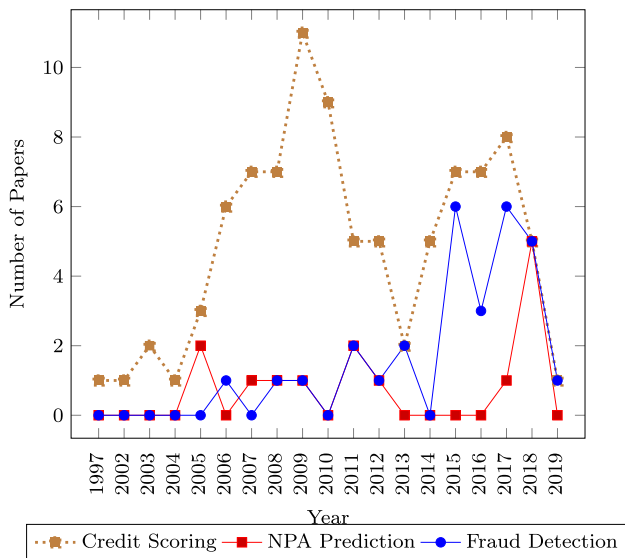
Figure 4 gives the number of papers for different risk evaluation techniques sorted according to the digital



**Fig. 4** Count of papers under different digital libraries categorized by risk evaluation technique



**Fig. 6** Count of papers on AI techniques for credit risk evaluation classified as ML, survey and others



**Fig. 5** Year wise trend of number of papers for different risk evaluation techniques

libraries. The number of papers on default prediction was on the lower side compared to other risk evaluation techniques. Credit scoring is the most widely used credit risk evaluation technique. Year wise trend of the number of papers falling under different credit risk evaluation techniques can be seen in Fig. 5. We observed that there was a peak in number of studies for the year 2009 for credit scoring, primarily, due to economic downfall in the year 2008. Number of studies for fraud detection has increased in recent years, primarily, due

to increased digitization in banks. Studies on NPA prediction have risen since 2016 indicating a need for more thorough credit scoring and echoing the observations of IMF [1] and banking regulators.

The different survey papers published before and how they differ from our paper is shown in Table 1. One of the things that was noticed was that only one study followed the systematic literature review approach for their study. The comprehensive and systematic nature of this study makes it unique.

### 3.1 Answer to RQ1: AI techniques

Artificial Intelligence can broadly be divided into categories shown in Fig. 2. Relevant categories to us are: Knowledge Representation, Planning and Deductive Reasoning and, Problem Solving. Planning involves Machine Learning and it is categorized into supervised, unsupervised and reinforcement learning. However, not all models under these categories are popular for credit risk evaluation. Under non-ML AI techniques, Ontology based and fuzzy logic based systems were proposed by the researchers for credit risk evaluation. The studies shortlisted after search process were divided into three categories namely credit scoring, NPA prediction and fraud detection. We have further classified them into studies that use ML Techniques, that are survey/analysis studies and studies that use non-ML AI techniques. The results can be seen in Fig. 6. The AI techniques (other than ML techniques) used for credit scoring and fraud detection are shown in Table 2.



**Table 1** Summary of non-ML AI approaches for credit risk evaluation

References	Studies searched	Comments	Differences
Popat and Chaudhary [170]	7	Done for credit card fraud detection Techniques are explained in detail	Done for credit risk evaluation in general Technique is not explained for any method of evaluation
Dastile et al. [171]	74	Study based on papers between 2010 and 2018 Different techniques are also explained Deep learning approaches studied in detail	Studies before 2010 not included Studies for fraud detection not included
Bakshi [172]	Not specified	Credit card fraud detection is studied Different techniques are explained	Studies used for analysis not listed Results are not detailed
Huang and Day [144]	Not specified	Data mining techniques explored	Only credit scoring is surveyed Study focused mainly on comparing classifiers

**Table 2** Non-ML AI techniques for credit risk evaluation

Credit scoring	NPA prediction	Fraud detection
Ontology based ([27]), fuzzy logic based ([23–26, 38–41]), decision table ([28]), mobiscore ([29]), bstacking ([30]), expert system ([31]), grey relational analysis ([32]), adaptive reference system ([37]), domain adaptation approach ([33])	No studies were found	rule based ([34, 42]), scatter search ([35]) APATE ([36])

Fuzzy logic based models for credit scoring are popular. Marikkannu and Shanmugapriya [23] proposed a fuzzy set based domain driven approach for customer credit data classification. Linear combinational sets of attributes for classification are built using domain expertise knowledge. Romanyuk [24] proposed a decision support system concept for granting of loan. It is based on the use of loan price function (which is continuous) of the credit score of borrower. Wei [25] also proposed a credit risk assessment model based on fuzzy theory. Hoffman et al. [26] proposed two evolutionary fuzzy rule learners. Evolution strategy is used in the first approach for generating approximate fuzzy rules, where every rule consists of membership functions which have their own definitions. Another learner is a genetic algorithm that extracts fuzzy rules which are descriptive. In this method, a common linguistically explainable definition of membership functions which are in disjunctive normal form is shared by all fuzzy rules. Other AI methods such as Ontology based, echo state network based, decision table, mobiscore, bstacking, expert system, grey relational analysis, adaptive reference system and domain adaptation approach have also been explored by researchers for credit scoring. Kotsiantis et al. [27] proposed an ontology-based system that predicts credit risk by using intelligent reasoning and searching mechanisms. The proposed ontology was designed and implemented such that it represents statements which are financial in nature. The domain could be modeled in a way that was shareable, efficient and reusable because of

the use of ontology. Pedro et al. [29] proposed MobiScore, an approach in which mobile phone usage data is used to build a model of the financial risk of user. This model could prove to be a good alternative when the applicant's financial history is not available. Xia et al. [30] proposed a novel ensemble credit model, which is heterogeneous in nature, that combines the bagging algorithm and stacking technique. Mahmoud et al. [31] proposed an expert system for assessing and supporting credit decisions on the banking sector. The main goal of the expert systems is to make skill available to technicians and decision making people. Lin et al. [32] proposed a grey relational analysis (GRA) approach for credit risk assessment of the banking sector. Huang and Chen [33] proposed a domain adaptation approach based data mining strategy for tasks which require credit risk assessment. In this method, the training of the algorithm is done on a source domain with numerous samples. Then the algorithm is applied on the target domain with relatively less number of samples. It does not require the equal distribution of the two domains.

There are quite a few non-ML artificial intelligence techniques used to detect frauds in transactions and loans. Gadi et al. [34] applied Artificial Immune System for credit card fraud detection. They also did a comparison of the results with that of other techniques such as Naive Bayes, Bayesian Networks, Neural Networks and Decision Trees. For parameter optimization, they used Genetic Algorithm (GA). Duman and Ozcelik [35] used genetic algorithm (GA) and

scatter search for detecting credit card fraud. Van Vlasselaer et al. [36] proposed APATE which is an approach to detect credit card transactions in online stores which are fraudulent in nature. The approach takes intrinsic and extrinsic features of the transactions and combines them. Combination of both these features leads to best performing models. The key observations from the research on non-ML AI techniques for credit risk evaluation are:

1. The results from various approaches show that their accuracy is comparable with Decision Trees, Support Vector Machine (SVM), Neural Networks, etc. However, there is no study which compares these proposed approaches to see which among them is better.
2. Fuzzy theory based systems for credit risk evaluation have potential to be used.
3. There are three studies which propose a tool for credit granting institutions [27, 31, 37] to help them in loan granting decisions. These tools are a good option for banks to adapt provided the administrator of the tool has domain and models knowledge to change the classification with changing scenarios of the outside world.

The key summary of Non-ML AI approaches for credit risk evaluation can be seen in Table 3. The past two decades have seen a growing interest in machine learning among the researchers with good computing capability to process large volume of data. We observed from the results in Fig. 6 that ML techniques are being more explored by researchers for credit risk evaluation—credit scoring, default prediction and fraud detection. In answer to RQ2, we discuss about ML techniques for credit risk evaluation.

### 3.2 Answer to RQ2: ML techniques

The ML techniques used for credit scoring, credit risk evaluation, NPA prediction and fraud detection are tabulated in Table 4. The distribution of different ML techniques for credit risk evaluation can be seen in Fig. 7. One major finding after the process of going through these studies is that ML techniques outperform the traditional statistical and optimization models [143]. The study by Malhotra and Malhotra [143] suggested that Neural Networks prove to be better than traditional statistical and optimization techniques. However, Huang and Day [144] showed that the support vector machine models have better accuracy rates among the 17 classification models investigated and therefore the past classification models are outperformed in the credit scoring context. This is supported by Khemakhem and Boujelbene [145] who did a study on credit risk evaluation for Tunisian banks and compared traditional models and modern ANNs and SVMs. The study also concluded that RBF kernel SVM was the best method in terms of sensitivity, specificity and

accuracy with the error rates which are least among others. Nwulu and Nnamdi [146] did a comparative analysis of SVM and ANN for credit scoring and concluded that ANNs perform slightly better than SVMs. Thus, we can say that growing interest of researchers towards developing ML techniques is justified as these models are better in terms of accuracy. In the following subsections, we give detailed explanation on how credit risk is actually computed and what datasets researchers use for their machine learning models.

Credit risk evaluation is done through the development of classification models, in order to distinguish between creditworthy and non-creditworthy clients [46]. A common approach for credit risk assessment is to apply some kind of classification technique to previous data of customers so that we find some kind of relation between the characteristics of the customer and failure of the loan. An important component of the modern techniques for credit risk evaluation is an accurate classifier that discriminates between good customers and bad ones. Due to its importance and better accuracy figures, there is an increasing research interest about credit risk assessment through machine learning techniques. Firstly, many statistical models and optimization techniques, such as linear discriminant analysis [147], logit analysis [148], probit analysis [149], linear programming [150], integer programming [151] and k-nearest neighbor (KNN) [152] are widely applied to credit risk evaluation and modeling tasks. There can be further improvement to these techniques although they can be applied for credit risk assessment. Recent studies have revealed that artificial intelligence (AI) techniques, such as SVM and neural networks perform better than traditional statistical models and optimization techniques for credit risk evaluation due to flexibility of tuning the weightages and ability to classify even though the features are not easily separable. We describe the research related to each credit risk type separately here and at the end of the each type, key observations are drawn.

#### 3.2.1 Credit scoring

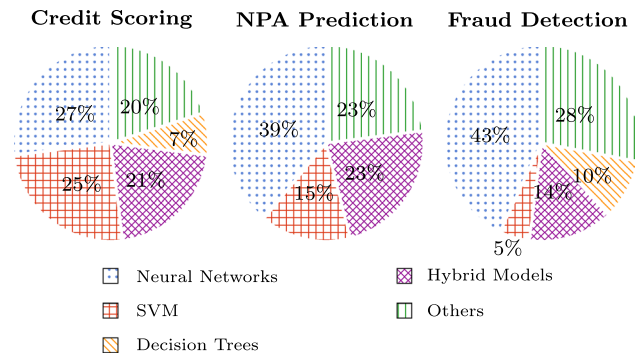
Credit scoring using machine learning is generally done using some kind of classifier which differentiates between creditworthy and non-creditworthy customers using the previous data of the customers. An important step in the classification process is to choose an accurate classifier for classification of good customers and bad customers. The ML techniques used by researchers for credit scoring can be seen in Table 4. The different techniques that can be seen are neural network techniques and its variants, SVM and its variants, Naive Bayes, Markov Chain, HMM, Bayesian Networks, Decision Tree, Bayesian Ensemble, HLVQ-C, Hybrid models and Ensemble models.

**Table 3** Summary of non-ML AI approaches for credit risk evaluation

Author	Approach	Term definitions	Comments
Kotsiantis et al. [27]	Ontology based	Ontology: an ontology consists of classes, instances, properties and rules. Classes and instances: Elements are modeled using instances which belong to classes. Properties: They establish relationships between the concepts of an ontology. For example, the property “BelongTo” associates an object with the parent it belongs to. Rules: They model logical sentences that are always true	This ontology has two concepts: ‘Customer’ and ‘Formal Financial Statement’. ‘Part Of’ and ‘Is A’ are two inter-conceptual relations
Marikannu and Shanmugapriya [23]	Fuzzy set based domain driven approach	Fuzzy set: The problems in which the source of imprecision is the lack of clearly defined criteria for class membership rather than random variables are solved naturally using fuzzy set algorithms	A hybrid approach of fuzzy set and multiple criteria and multiple constraints is used in linear combination to reduce the time complexity
Mahmoud et al. [31]	Expert system	CommonKADS methodology: CommonKADS follows a systematic approach in knowledge engineering that adopts the concept of modeling in the process of knowledge base system development	CommonKADS methodology is used in the developing process. The proposed development environment consists of document analysis and knowledge acquisition processes
Vac and Găban [37]	Adaptive Reference System (ARS)	ARS: ARS is used to evaluate organization’s performance	The implementation of the ARS includes evaluation of the tendency of bank to give the loan along with the eligibility of the customer. External factors are also taken into account
Huang and Chen [33]	Domain Adaptation Approach	Domain Adaptation Approach: Used when the training and test datasets have different domains	The challenge of domain adaptation approach lies in the gap between training domain and the test distribution
Romanyuk [24]	Fuzzy logic based	Loan price function: Loan price function is determined by loan value and bank’s possibility to convert private information into a loan price	Application of the proposed system provides a competitive advantage for a bank to attract borrowers with a low credit risk
Yazdani et al. [41]	Fuzzy logic based	FCMCR: Fuzzy Classification Method in Credit Risk (FCMCR) is inspired by the “Similarity”, and “Dominant gene”, approaches	The accuracy of the approach was found to be around 99% when tested on a dataset of a large bank

**Table 4** ML techniques for credit risk evaluation

	Credit scoring	NPA prediction	Fraud detection
	BP NN ([43–45]), ANN ([46–49]), echo state network ([50]), NN ([51–55]), Neural based ([56]), Restricted Boltzmann machine ([57]), Neural Tree model ([58]), Wavelet Network Model ([59]), Denoising-Autoencoder Approach ([60]), Markov Chain ([61]), SVM ([62–79]), Naive Bayes ([80, 81]), HMM ([82, 83]), HLVQ-C ([84]), decision tree ([85–88]), hybrid model ([89–91] [92–102]), ensemble model ([103–108])	Evolutionary NN ([109]), ANN [110], BP NN ([111, 112]), RBF NN ([113]), SVM ([114, 115]), HMM ([116, 117]) Bayesian ([118]) hybrid model ([119–121])	MLP NN ([122]), ANN ([123]), NN ([124–126]), Deep NN ([127]), Auto Encoder NN ([128]), Ant Colony Based ([129]), MLP and Naive Bayes ([130]), Deep Reinforcement Learning ([131]), Deep Learning ([132]), Active Learning ([133]), HMM ([134–136]), SVM ([137]), decision tree ([138, 139]), hybrid model ([140–142])



**Fig. 7** ML techniques for credit risk evaluation

Neural networks are becoming increasingly popular among researchers in recent years. Li et al. [44] proposed a model based on Back-Propagation (BP) algorithm to identify “good credit” groups from “bad credit” groups. Li and Wu [43] and Zhu et al. [45] proposed a credit risk assessment model based on BP Neural Network to identify potential defaulters.

Hu and Tang [47] proposed an artificial neural network (ANN) based credit risk assessment which measures the credit score of the applicant. This model has many characteristics such as self-adaptation, self-learning and parallel processing. The most suitable candidates for this model are the domestic commercial banks which have incomplete data and delayed data. Dima and Vasilache [46] proposed an ANN model for corporate credit risk evaluation to classify good creditors from bad ones. The paper uses probit regression and ANN model and the classification is based on the number of delay days.

Derelioğlu et al. [53] proposed a cascaded MultiLayer Perceptron (MLP) and Neural Rule Extraction (NRE) system for classification of customers as either creditworthy or uncreditworthy. In the rule extraction stage, the forwarded result is revealed to be of what condition the good customer was finalised in the decision. Zhang et al. [58] proposed a credit risk evaluation approach using flexible neural tree (FNT) model for classification of loan applicants. Zhaoji et al. [59] proposed a wavelet network model based on Particle Swarm Optimization (PSO) for classification of loan applicants. Fan and Yang [60] proposed a denoising autoencoder approach for training the neural networks. The paper proposes a denoising-autoencoder-based Neural Network model for credit risk analysis. This was proposed as the authors identified that the traditional ANNs learn not only from the training data but also from the noise in it. To overcome this drawback, this model was proposed. Lai et al. [51] built a Neural Network metalearning model for credit scoring. Marin-de-la-Barcelona et al. [56] proposed artificial metaplasticity (biological property of real neurons) applied on MLP. So neurons have this biological property of metaplasticity. Barcelona et al. applied this property on neural

networks and were able to propose a new machine learning method for credit scoring. Tomczak and Zieba [57] proposed a scoring model based on Classification Restricted Boltzmann Machines (ClassRBM). This model first trains the data on ClassRBM and then generates a scoring table. Geometric mean of sensitivity and specificity is used to take care of the imbalanced data. Baesens et al. [55] analysed three real life credit datasets and presented the results. The analysis was done using neural network rule extraction techniques. Decision tables were used to visualize the scores. The rules were extracted using three rule extraction techniques. It was concluded that neural rule extraction techniques have the potential to be used for credit risk analysis.

As can be observed, researchers are moving towards hybrid systems with neural networks in it. Huang et al. [98] proposed classification of loan applicants of state-owned commercial banks using fuzzy neural networks. Huang and Tian [97] proposed a classification model of applicants for commercial banks based on Fuzzy Probabilistic Neural Network Model (FPNN). This model is a combination of the Probabilistic Neural Network (PNN) and relative membership degree in fuzzy mathematics. Oreski et al. [92] proposed a hybrid system with Genetic Algorithm (GA) and ANNs for credit scoring of applicants. In this model, the feature selection is done using GA and classification using ANNs. The proposed hybrid system was found to be competitive with other models for credit scoring tasks. Taremian and Naeni [93] proposed a Hybrid Intelligent Decision Support System (HIDSS) for credit risk evaluation and classified applicants as creditworthy or not, based on neural networks and GAs. MLP Neural Network was used for this purpose in which a co-evolutionary process was used to train the weights of the MLP network. Weidong et al. [95] proposed a hybrid model based on Back Propagation (BP) Neural Network and Logistic Regression. The primary advantage of using this model is that it gives better accuracy than simply applying logistic regression. Also, it is more robust than simply applying BP neural network. Djemaiel et al. [96] proposed a hybrid neural network model built using a combination of Radial basis function (RBF) neural network and Elman neural network. The context for the data was set using big data. The proposed model proved to be efficient when it was used to classify customers as “good” or “bad” based on their credit scores. Hence, the proposed hybrid model can be a good choice when opting for a classification technique for credit scoring. Fu and Liu [89] proposed a model in which Radial Basis Function (RBF) Neural Network is combined with Genetic Algorithm (GA). This model is called GA-RBF Neural Network Model. Genetic algorithm is used for optimization of weights in this model, position of center and spread of center of RBF neural network.

SVM is a widely researched classification technique for credit scoring due to many reasons. Not many data points or

support vectors are needed for determination of the optimal hyperplane. SVMs provide an excellent generalization ability. It is also relatively easy to train SVMs. SVMs also do not contain any local optimal like neural networks. SVMs scale relatively well to data with high dimensionality and trade off between classifier error and complexity. Many have used SVM and its variants to perform credit scoring. Farquad et al. [62] proposed a PCA-SVM model which performs PCA for dimensionality reduction on dataset and SVM for classification. The PCA-SVM model had good performances. When compared to SVM alone, it had better accuracy. Similarly, it outperformed PCA-Logistic Regression model. Harris [63] introduced the use of clustered Support Vector Machine (C-SVM) for credit scoring. This model was proposed in response to SVM being computationally expensive for high dimensions. C-SVM tries to address this challenge and provides us with credit score of a customer in relatively less time even if the dataset is non linear and large. Huang [64] integrated Kernel Graph Embedding (KGE) with SVM for credit scoring. In this model, KGE is a graph based technique used for dimensionality reduction. This SVM-KGE classifier was shown to be better than traditional SVMs and other multi-class SVMs. Li [65] proposed a model based on fuzzy integral support vector machine (SVM) in which the importance of the output of sub SVM is taken into account. This method proved to perform better than SVM applied alone. Feng et al. [66] and Yang et al. [68] proposed SVM classification model based on PCA for dimensionality reduction for commercial banks. It is similar to the PCA-SVM model proposed by Farquad et al. Lv and Peng [67] proposed a model which combines rough sets and SVM to evaluate credit risk in commercial banks. The indexing system was established in this model and the reduction of the number of indexes was done using rough sets. Comparison with back propagation (BP) Neural Network showed that the rough set-SVM method is more precise and efficient than it. Wei et al. [74] proposed classification of credit applicants using SVM with mixture of kernel. The model uses 1-norm and convex combination of basic kernels. Computational cost is greatly reduced as the quadratic problem is reduced to only one linear programming problem. Wei et al. [69] proposed a least squares support vector machine with mixture kernel (LS-SVM-MK). Just like previously Wei et al. used mixture of kernel with plain SVM, this time the researchers used it on LS-SVM. The problem of the traditional LS-SVM model such as the loss of robustness and sparseness for credit risk evaluation was solved using the mixture of kernel. It was found out that LS-SVM-MK can improve the generalization ability of LS-SVM and can obtain a smaller number of features. Sun and Yang [73] proposed a multi-layer support vector machines (SVM) classifier to evaluate the credit risk for commercial banks. The accuracy of this method is shown to be higher than BP neural network. Lai



et al. [75] proposed the use of least square support vector machine (LSSVM) technique to design a credit risk assessment system for classifying “good” customers and “bad” customers. A linear programming problem is all that needs to be solved unlike the traditional quadratic equation which saves us some computational complexity as a result. Gestel et al. [79] proposed a Least Squares SVM classifier for credit scoring that outperforms traditional SVM classifiers. Later, Gestel et al. [77] proposed a Least Squares Support Vector Machine (LS-SVM) classifier within the Bayesian evidence framework. It automatically inferred and analyzed the creditworthiness of potential corporate clients. This method of classification was shown to be better than traditional Linear Discriminant Analysis (LDA) and Logistic Regression models. Zhu et al. [70], Ma and Liu [71], Li et al. [154] and Li et al. [76] also proposed a SVM model for identifying good creditors from bad ones.

Ruiz et al. [90] and Gestel et al. [102] proposed a hybrid model which uses logistic regression and SVM to perform credit scoring. For loan classification processes, Ruiz et al. modeled credit score based on non-traditional data which is obtained from smartphones. Gestel et al. emphasize on good readability of the model and show that as the SVM model has a gradual increase in the complexity, starting with a basic model, the readability and performance of the model goes up. Huang et al. [91] proposed a data mining approach using SVM for credit scoring. The proposed hybrid GA and SVM integrated strategy simultaneously performs model parameters optimization and feature selection task. Zhou and Bai [99] proposed a SVM classifier using genetic optimization algorithm which is hybridized with rough set theory. A reduced information table is the result of the application of rough set theory. SVM is trained using this reduced information table and the classification rules are also crafted using the same. Hao et al. [101] proposed a Fuzzy SVM (FSVM) for credit scoring. FSVM assigns fuzzy membership to each data points which helps in improving the generalization ability of traditional SVMs. Jiang and Yuan [100] used Particle Swarm Optimization (PSO) for searching the SVM parameters. After the search is done, the SVM model is used for credit scoring. Martens et al. [78] proposed rule extraction techniques for SVMs and introduces two others Trepan and G-REX. The other two are taken from the AI domain. The proposed technique does not lose much accuracy and also provides comprehensibility or readability as compared to other models. The accuracy of this model is even comparable to C4.5 and logit.

Apart from neural network and SVM based approaches, several other classification techniques are proposed for credit scoring. Though not a popular classification model for credit scoring, Naive Bayes approach has also been proposed by some. Vedala and Kumar [80] proposed a Naive Bayes classification for credit scoring. This scoring is done primarily

on e-lending platforms. The paper uses social networks to extend its database. Okesola et al. [81] also proposed a Naive Bayes classification model for credit scoring. The input variables in this method are the demographic and material indicators. A modern approach for credit scoring is the decision tree method [155]. Szwabe and Misiorek [87] proposed a decision tree model for making credit decision. In this paper, several approaches for classification of loan applications are evaluated that provide a single decision tree as the final form of their results. Xia et al. [88] proposed a boosted Decision Tree approach for credit scoring. Bayesian technique was used for hyperparameter optimization. Wei et al. [85] proposed a model for credit risk evaluate using decision tree algorithm. Lang and Sun [86] studied the problem of class imbalance in credit risk early warning by applying decision tree algorithm. Empirical results have shown that there is strong sensitivity for decision tree algorithm to imbalanced data. This is when it is modeled for early warning of credit risk. Hidden Markov Model (HMM) has been explored for credit scoring. Benyacoub et al. [82] proposed a HMM combined with Baum-Welch procedure for credit scoring for iterative re-estimation of the parameters from a sequence of observations. Petropoulos et al. [83] used Student's-t hidden Markov models (SHMMs) for corporate credit scoring system. Capturing of correlations and high robustness to outliers is an extra advantage of using SHMMs. SHMMs are shown to have competitive performances as compared to other models. Timofeev and Timofeeva [61] proposed an estimation of Loan Portfolio Risk based on Markov Chain Model. Discrete time model is used and the system state is fixed through identical time intervals which is taken as once a month.

Another method used by the researchers for credit risk evaluation is ensemble learning method. It is similar to hybrid systems. The difference is that in ensemble learning, the decision is taken by pooling multiple classifiers while in hybrid method of classification, various techniques are used on the data and the final parameters and pre-processed data is passed on to a single classifier which does the classification. There are many examples of researchers using ensemble learning method for credit risk evaluation. Ensemble techniques outperform individual classifiers, hence, they are widely in use. Chen et al. [108] proposed an ensemble model which ensembles logistic regression analysis (LRA), MLP-NN and cluster. A Bayesian approach is followed for the ensemble. It was found that this method outperforms single classifiers. Hsieh et al. [107] proposed an ensemble classifier which incorporates various data mining techniques. Class-wise classification is introduced as a preprocessing step. Bayesian network, SVM and Neural Network are used for the augmentation of the ensemble classifier. Zięba and Świątek [105] proposed an ensemble classifier based on switching class labels

**Table 5** Studies of ML techniques in credit scoring

Author	Number of Features	Dimensionality Reduction		Model
		F.S.	F.E.	
Derelioğlu et al. [53]	27	DT, SVM-RFE	PCA, FA	MLP
Marin-de-la-Barcelona et al. [56]	16	–	–	Artificial metaplasticity with $\eta = 1$
Huang et al. [91]	24, 14	Grid, F-score, GA	–	SVM with 5-fold cross validation
Ma and Liu [71]	8	–	–	SVM with $C = 4$ and $\sigma = 25$
Zhou and Bai [99]	12	Rough sets	–	SVM with $C = 57.7$ and $\sigma^2 = 2.4$
Lv and Peng [67]	10	–	–	SVM with $C = 100$ and $\sigma^2 = 2.1$
Li et al. [154]	17	–	–	SVM with 5-fold cross validation and $C = 10$ and $\sigma = 0.1$

techniques. There are two data mining problems which are solved through using switching class label technique: first is that asymmetric cost matrix would be an issue, another is imbalanced dataset's predicament. Zhen and Wenjuan [104] proposed a SVM ensemble method based on fuzzy integral for credit risk evaluation. Different weights are given to separate components of SVM and their outputs are aggregated to give the result. The accuracy of the model was found out to be satisfactory.

Krishna and Ravi [169] proposed feature subset selection method by incorporating Adaptive Differential Evolution as a wrapper and tested it on three datasets for both credit scoring and fraud detection. The proposed method proved to be better than the previous ones.

We also found three studies on credit scoring using deep learning approaches. We found out that deep learning approaches can be useful in evaluating credit score. A summarized information about the techniques used in the studies for credit scoring can be found in Table 5. Here,  $\eta$  is the learning rate of the neural network and  $C$  and  $\sigma$  are the parameters for a nonlinear support vector machine (SVM) with a Gaussian radial basis function kernel. The key observations drawn from the research on credit scoring are:

1. Neural networks are the most widely studied models for credit scoring, most notably feed forward neural networks. The primary advantage in using feed forward neural networks is its excellent generalizability property. However, the interpretability of these models are an issue as they are black box models, which makes it difficult for the person in charge of giving loans to understand the process followed by the model.
2. SVM method has been used by many studies for classification, however, it becomes computationally expensive when large data sets are used. This problem has been addressed by some [63, 90].
3. Hybrid and ensemble models are becoming popular as the proposed models overcome the shortcomings of individual classifiers and provide better accuracy rates.

4. There are survey studies that compare the results by different classifiers [144, 156]. Comparison of individual and ensemble classifier is done by Singh [157].

### 3.2.2 NPA prediction

Another type of credit risk evaluation technique is NPA or default prediction. This is performed to predict which loan is likely to become a default so that appropriate measures can be taken to deal with the situation. The ML techniques that are used for default prediction are different types of neural networks, SVM and hybrid models. Zhang [111] proposed an early warning default risk model based on rough sets and BP Neural Network algorithm. First, a default index is created for the personal loans and then rough sets is applied to it. This helps in streamlining the indexes. Then a BP Neural Network is trained on the data samples to determine the default risk. Makrygianni and Markopoulos [110] proposed default prediction using feedforward ANN which considers economic and personal information of the loan applicant. The proposed model was found to give satisfactory accuracy. Ribiero et al. [115] proposed enhanced default risk models using SVM+. Generalization is improved even further when using SVM+ as it not only takes training data into account but also additional information. Baseline SVM was outperformed by SVM+ on a French company dataset. Feki et al. [114] proposed methods of discrimination of banks as per the rate of Non Performing Loans (NPLs). It was done using different approaches of multiclass SVM and Gaussian Bayes models. Strategies for variable selection are also proposed. Ni et al. [117] proposed an extension of Factorization machines called RobustFM. Class imbalance problem and noisiness problem in default prediction is supposed to be addressed by RobustFM. In terms of F-measure, RobustFM outperforms traditional state-of-the-art classifiers. Chen et al. [113] proposed a loan default prediction model in which a hybrid undersampling method is used. The name of this undersampling method is DSUS

**Table 6** Studies of ML techniques in default prediction

Author	Number of features	Dimensionality reduction		Model
		F.S.	F.E.	
Zhao et al. [119]	18	GA	–	BP neural network with $\eta = 1$
Yao et al [121]	16	–	–	Fuzzy neural network
Feki et al. [114]	30	Unique to this study	–	SVM and Bayes classification
Masmoudi et al. [118]	9	–	–	Bayesian network
Zhang [111]	–	Rough sets	–	BP Neural Network

and a stochastic sensitivity measure and the RBF Neural Network is combined with k-means clustering method for default prediction. Data was taken from a P2P company in China and used for the validation of the performance of the method. Su and Zhang [120] proposed an early-warning model by optimizing the weights and thresholds of BP neural network using GA. It is based on nonlinear combinatorial forecasting principle. The accuracy and the simulation error are known to be improved on use of GABP method as opposed to the traditional methods. Miglionico and Parillo [112] proposed an early warning indicator system using ANN. The implementation was done using a custom developed sfloat24 Math library. The ANN consisted of 3 layers, and a low cost FPGA device was used for its development. Fault tolerance and good accuracy are the characteristics of ANN when concerned with loan risk evaluation. Yao et al. [121] proposed a indicator system to evaluate credit risks of commercial banks based on fuzzy neural network. The results were good and it was found out that this model served as a better model than the black box neural network models. Oguz and Gurgun [116] explored the Hidden Markov Model(HMM) for the task of probability of default (PD) modeling and classification. The credit customers are assigned default bankruptcy probabilities using PD modeling instead of classifying them as credit-worthy and uncreditworthy customers. The HMM method is shown to be robust and powerful for default prediction tasks. Table 6 shows the summary of the studies under default prediction.

The key observations drawn from the research on NPA or default prediction are:

1. SVM and neural networks are mainly used for default prediction.
2. Recently, hybrid models have gained a lot of popularity as they outperform individual classifiers. Performance of SVM can be enhanced by incorporating methods such as rough set theory and fuzzy theory with it.
3. Lack of public datasets for default prediction and governmental regulations that are primarily rule based seemed to have curtailed the research on NPA prediction. However, recent guidelines from regulators on

early warning looks encouraging to use ML techniques for NPA prediction.

### 3.2.3 Fraud detection

Fraud in financial transactions can endanger their reputation among customers as well as cause heavy damages. As said by Abakarim [132], banks and financial institutions are investing in perfecting the machine learning algorithms and big data analytics to identify fraud and come up with fraud detection systems which are accurate and competitive. There can be many types of frauds in the banking sector. However, as stated earlier, we will focus only on credit card frauds, banking transaction frauds and loan application frauds conspired by customers. Fraud detection is a binary classification problem in which the loan is categorized as either 'fraudulent' or 'non-fraudulent'. The idea is to apply a well suited classifier on the problem, however, the classifier should also be trained on a suitable dataset. The major approaches for coping with credit card fraud in banking are either statistical or based on artificial intelligence. The ML techniques applied by the researchers in these studies are NN, HMM, SVM and its variants and decision tree.

Mubarek and Adalı [122] proposed a MLP neural network technique for fraud detection. The proposed MLP ANN was shown to yield average better performance when compared to Naive Bayes and Decision Tree models. Patil and Dharwadkar [123] also worked on customer retention and fraud detection and proposed a supervised ANN for classification purpose. This supervised ANN showed competitive results and better accuracy than similar models. Ghobadi and Rohani [124] proposed a credit card fraud detection model to prevent credit card frauds using the Artificial Neural Networks. The model also includes a Meta Cost Procedure. It is added to deal with the problem of class imbalance of data. Zhan and Yin [126] proposed a fraud detection method for loan applications based on Neural Network and Knowledge Graph. Borrower's phone network is used to extract features which is a time consuming process when done using other methods. Kazemi and Zarrabi [127] proposed deep neural networks for fraud detection in credit card transactions. Deep autoencoder

**Table 7** Studies of ML techniques in fraud detection

Author	Number of features	Dimensionality reduction		Model
		F.S.	F.E.	
Abakarim [132]	30	–	–	Deep learning using auto encoder with 3 encoders and 3 decoders
Jiang et al. [134]	7	Sliding window based	–	Random forest based with clustering and aggregation strategy
Ghobadi and Rohani [124]	18	–	–	ANN with topology [17 60 50 1]
Zamini and Montazer [128]	28	–	PCA	Autoencoder based clustering
Charleonnann [130]	23	–	–	RUSMRN ensemble

is used to extract features from the information provided by credit card transactions. Deep learning has proved to be beneficial in several fields and this model has shown to do well for credit card transaction fraud detection. Zamini and Montazer [128] proposed an unsupervised fraud detection method using autoencoder based clustering. The autoencoder consists of 3 layers and the k-means clustering is used for the clustering purposes. The model proved to be better in comparison to other models. Liu et al. [129] proposed an Ant Colony based approach for fraud detection in business. The model performs better as compared to the traditional ANNs as the local optima problem is solved in the ant colony optimization based approach. Charleonnann [130] proposed a credit card fraud detection technique using RUS and MRN algorithms, so the technique for fraud detection was named as RUSMRN. Classification of unbalanced data is done using boosting and data sampling. A Taiwanese bank is used for data collection. Bouchti et al. [131] used deep reinforcement learning (DRL) for fraud detection in banks. Various interesting facts about DRL are covered in the paper and competitive performance is shown by DRL method. The paper is rather technical, however, a new approach for fraud detection has appeared in front of the research community. Karlos et al. [133] predicted fraudulent financial statements (FFS) using active learning. Supervised learning methodology has been used for this purpose. Active learning strategy seemed to perform better than supervised models. Jiang et al. [134] proposed an approach for credit card fraud detection using feedback mechanism and aggregation strategy. Rahmawati et al. [135] proposed fraud detection in business processes in the bank credit application using Hidden Markov Model (HMM). The accuracy of the method was found to be competitive and was benchmarked at 94%. Khan et al. [136] proposed a credit card fraud detection system using Hidden Markov Model (HMM). The system is compatible with scaling to large databases or to say large volumes of credit card transaction. Kotsiantis et al. [138] predicted fraudulent financial statements (FFS) using decision trees. Published financial data was

used for detecting fraudulent financial statements and the performances of the machine learning techniques in using this data was evaluated in this paper. Decision tree was shown to achieve the best performance among all the classifiers taken into consideration. The input vector of the decision tree contained only financial ratios. Ravishankar et al. [158] did an analysis on detection of financial statement fraud using data mining techniques. The dataset was taken from 202 Chinese companies and the comparison was done with feature selection and without it. Probabilistic Neural Network (PNN) outperformed all others which was without feature selection techniques. PNN along with Genetic Programming (GP) outperformed the ones with feature selection.

Hybrid methods have been adopted in fraud detection techniques. Mareeswari and Gunasekaran [140] proposed prevention of credit card fraud using hybrid Support Vector Machines (HSVM). Communal and spike detection are used as hybrid techniques. Scalability is efficient in this method upon updating the evaluation of data. Montini et al. [141] proposed a hybrid sampling model for bank fraud diagnosis. The MLP model is used for training the bank transaction data. Kamaruddin and Vadlamani [142] employed a one-class classification approach in big data paradigm for detecting credit card fraud. It was an implementation of a hybrid architecture of PSO and Auto-Associative Neural Network for one-class classification. Big data analytics is used in this method and this method is also known as PSOANN. Table 7 summarizes major studies for fraud detection. The key observations drawn from the research on Fraud detection are:

1. Neural network based classifiers are most popular among researchers for fraud detection with 43% of the studies of the selected studies being based on neural networks.
2. ANN based models perform better than linear models [132] for classifying loans as fraudulent or not.
3. SVM has proved to be better than back propagation neural networks [137] for classification of loans as fraudulent or not.

**Table 8** Public datasets for credit risk evaluation

Dataset	Used for	Studies
German credit data	Credit scoring, default prediction, fraud detection	[23, 27, 28, 30, 52, 57, 62, 63, 82, 87–89, 91, 103, 105, 107, 110, 116, 117, 123, 127]
Japanese credit screening	Credit scoring	[51, 54, 103, 106]
Australian credit approval	Credit scoring, default prediction	[30, 52, 56, 57, 82, 88, 91, 103, 116, 117]
Lending club loan data	Credit scoring	[30, 33, 60, 88]
Default of credit card clients	Credit scoring, default prediction	[87, 117]
Give me some credit	Credit scoring, default prediction	[87, 117]
Credit card fraud detection	Fraud detection	[132]
ccFraud	Fraud detection	[142]

**Table 9** Attributes of German credit dataset

German dataset
The status of the client's existing checking account
The duration of the credit period in months
The client's credit history
The purpose for the credit
The credit amount requested
The client's savings account/bonds balance
The client's present employment status
Installment rate in percentage of disposable income
The client's personal (marital) status and sex
Whether the client is a debtor or guarantor of credit granted by another institution
The number of years spent at present residence
The type of property possessed by client
The client's age in years
Whether the client has other installment plans
The client's housing arrangements (i.e. own their home, rent, or live for free)
The number of existing credits the client has at the bank
The client's job
The number of people for whom the client is liable to provide maintenance for
Whether the client has a telephone
Whether the client is a foreign worker

- There is no significant survey or analysis study on fraud detection using machine learning best to our knowledge.
- There is extensive research on Social Engineering, Cyber attacks, Software vulnerabilities based frauds that is beyond the scope of our current study as they are not initiated by customer.

### 3.2.4 What are the public datasets available?

As there were observations on lack of public datasets, we analyzed available datasets for various credit risk

evaluation techniques, shown in Table 8. Some studies did not make their dataset public or simulated their own dataset making it difficult to compare credit risk algorithms. German, Australian and Japanese Credit datasets are the most used datasets for credit scoring. The datasets are explained below one by one.

The German Credit dataset has 1000 instances of which 700 are delinquent and the remaining are non-delinquent customers. The dataset contains 20 attributes (see Table 9). Interestingly, Microsoft Azure Studio also demonstrates the German Credit Dataset to do credit risk evaluation. From the research studies and examples on



various patterns that can lead to NPA and fraud, we opine that the dataset does not represent the world scenarios.

The Japanese Credit dataset has 125 instances which represent creditworthy and un-creditworthy clients.

This Australian Credit Approval dataset concerns credit card applications. It has 14 classification features with 690 instances. All attribute names and values have been changed to meaningless symbols to protect confidentiality of the data. This dataset consists of a good mix of nominal and continuous attributes. There are also a few missing values.

Lending Club dataset file is a matrix of about 890 thousand observations and 75 variables for loans issued between 2007-2015. The details of all the features can be viewed here.<sup>7</sup> This is probably the biggest dataset available for loan.

Credit Card Fraud dataset contains transactions made by credit cards in September 2013 by European cardholders. This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions. The features names have been changed to meaningless values and the number of columns in the dataset is 31.

### 3.3 Answer to RQ3: What are the challenges or limitations in this focus area?

As seen in previous subsections, there are numerous AI techniques for credit risk evaluation which include credit scoring, NPA/default prediction and fraud detection. We have statistical techniques, rule based techniques and ML based techniques to evaluate credit risk. The main advantage of computer aided credit risk evaluation is that human work is minimized since it learns from a pre-collected database to make accurate and reliable predictions. However, this research area like any other area comes with challenges.

One problem that is difficult to deal with in the area of credit risk evaluation is the changing domains of training dataset and testing dataset. The training dataset could be from a different geographic area or from a different bank when compared to the testing dataset. As there can be different rules and regulations in different areas and banks, the dataset will vary significantly and so will the relation between its features. Thus, this change in domain is responsible for inaccuracy in the sample classification and hence, there is a need to address this problem of changing domains. This problem has been tried to tackle by Huang and Chen [33] but this problem needs further exploration.

Another limitation of using machine learning models in credit risk evaluation is the influence of external factors or parameters. As an example, a farmer taking an agricultural

loan may not be able to pay his loan interests because of factors such as poor rainfall. Thus, these unknown factors (in this case, weather) hinder the ability of machine learning models to make accurate predictions. These factors require accessing information that is outside banking environment and are not part of customer profile. Also, some of the data can be in the form of images and unstructured text that needs to be extracted and gleaned for training the models. Another example could be the various macro economic issues such as country's GDP and inflation.

A common challenge that the researchers face during credit risk evaluation is pre-processing of data. Noisy data or data that contains outliers can have heavy effect on performance of model and so can redundant and irrelevant features [103]. Researchers use feature selection step or data-filtering to overcome this problem. Fan and Yang [60] tried to overcome the problem of noise using denoising-autoencoder as discussed before.

One of the prime challenges researchers face in evaluating credit risk is when datasets get large. This is when nonlinear approaches in classification become more and more computationally expensive. In credit risk evaluation, there are usually many irrelevant variables in the sample data which need to be removed. These variables make computation more expensive and we have to do redundant computation. For SVM classifier, size of the matrix of the quadratic programming is directly proportional to the number of training points [159]. This means that as number of training points increase, the size of the matrix increases. Thus, the quadratic programming problem becomes more and more difficult. To support the claim that SVMs that long training time, we can look at the SVM-GA model of Huang et al. [91]. It takes a long training time. This means that people are now in search of patterns in the datasets that would help in bringing down the time complexity [63]. SVMs are also black box models and hence improving the comprehensibility of these models is an area that needs further exploration. Parameter selection in SVM learning is a critical process if one wants to successfully model SVM for credit risk evaluation. Nowadays, grid search, rough sets, trial and error and genetic algorithm based techniques are becoming increasingly popular for parameter selection. GA is a parameter optimization technique while rough set is an indexing technique. Grid search is another technique for parameter selection, however, it is known to affect the computational complexity of SVMs in a negative manner. SVMs can become more robust if the parameter selection is explored properly and the techniques applied to it.

Studies have suggested that neural networks outperform many statistical techniques such as discriminant analysis, logistic regression and optimization techniques. However, they are not stable. This means only specific samples can be used for application of model. When there is a change in

<sup>7</sup> <https://tinyurl.com/LCDataDictionary>.

sample, the model's accuracy will change greatly. A large number of parameters, such as training methods, learning rate and network topology have to be refined before the neural networks can be successfully deployed. Another major drawback when neural networks are used for credit risk evaluation is that they lack the capability of explaining themselves. While high predictive accuracy rate can be achieved through them, the reasoning behind their decision making is not readily available [160]. There are many more drawbacks of neural networks such as trapping into local optimum and overfitting. Also since neural networks are non-linear in nature, sometimes there is huge time required for computation when there is a large dataset. It is still a challenging issue to find the optimal neural network model.

Another major concern in the field of credit risk evaluation is data shortage. Given a method, it is difficult to say that its performance is better than another method under all situations. Due to competitive press and privacy, in a realistic situation, a researcher can collect fewer data about credit risk. This makes it difficult for statistical methods and machine learning algorithms to obtain a continuously good result for credit scoring. To cope with the challenges of data shortage and poor performance, oversampling and other approaches are imperative to be introduced. Thus, we can safely say that further research is required in the area of data availability and data collection for credit risk evaluation.

While one may be still be convinced to use machine learning models for credit risk evaluation, it is good to keep in mind that machine learning models like any other models are not 100% accurate. Thus, relying on them for making decisions comes at a risk. It is up to the user to decide that to what extent he/she wants to involve them in the decision making process of credit risk evaluation.

### 3.4 Answer to RQ4: What are the research trends in credit risk evaluation?

Since, there were structural changes in the global financial market as well as an increase in the overall risk level was observed, it has become imperative to study credit risk evaluation. Over the last 20 years, much progress has been done in the area of credit risk evaluation. Credit scoring models are constructed by two fundamental and yet popular statistical tools: Linear Discriminant Analysis (LDA) and logistic regression (LR). As the times are changing, new methods have arrived such as Neural Networks, SVMs, k-NNs and Decision Trees. There are many other methods as described in the previous sections. However, hybrid models and ensemble models are becoming increasingly popular. Neural network and SVMs have their limitations which are being tackled by the current generation of researchers.

The prime research that is being carried out in the field of credit risk evaluation use classification algorithms that are

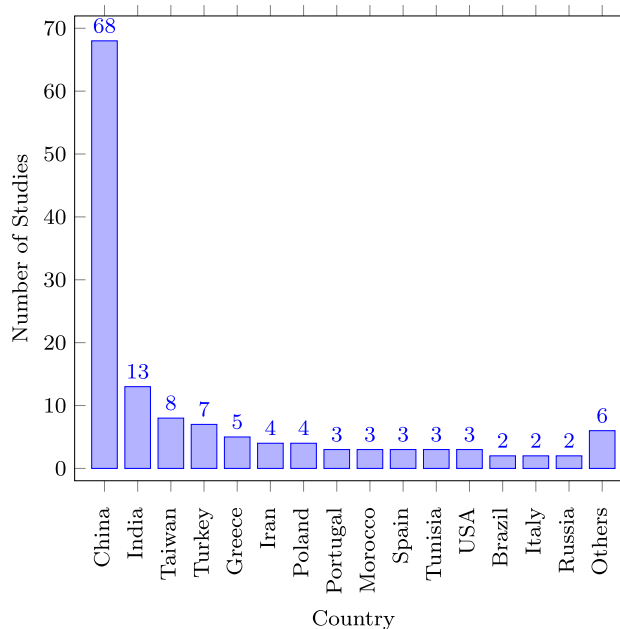


Fig. 8 Number of studies per country

non-linear in nature, such as neural networks and SVM. The research works related to neural networks and SVMs can be found in the previous subsections. SVM has received a lot of attention in the machine learning community because of its excellent generalization ability. Few have tried to perform credit scoring using Naive Bayes classification [80, 81]. For all three type of credit risk evaluation techniques, the researchers have also proposed many hybrid models that combine parts of two or more algorithms. Ensemble models for credit scoring are also becoming popular. The proposed ensemble models outperform single classifiers [162]. The HMM that has made remarkable achievements in speech recognition, engineering and many other fields is also applied in credit scoring and fraud detection, Benyacoub et al. [82] proposed an HMM based model for credit scoring. Decision trees are another widely used classification technique for credit scoring. But neural networks and SVM are still most popular machine learning models for credit scoring, default prediction and fraud detection.

### 3.5 Answer to RQ5: Universities working in the area of credit risk evaluation

The details on authors of the papers included in this study can be found in [163]. We noticed from our observations that considerable amount of the studies are from Chinese universities (see Fig. 8). The notable researchers in the field of credit risk evaluation according to number of studies published are shown in Table 10.

**Table 10** Notable authors for credit risk evaluation according to the count of published studies

Author	University	Count
Bart Baesens	Faculty of Economics and Business, KU Leuven, Belgium	8
Tony Van Gestel	Department of Electrical Engineering (ESAT), KU Leuven	6
Kin Keung Lai	Department of Management Sciences, City University of Hong Kong	6
Shouyang Wang	Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences	5
Ligang Zhou	Department of Management Sciences, City University of Hong Kong	4
Vadlamani Ravi	Centre of Excellence in Analytics Institute for Development and Research in Banking Technology Hyderabad, India	4
Lean Yu	Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences	3

### 3.6 Important Results

Important information about some of the studies included in the SLR can be viewed in Table 11. The comments give an insight into how some of the challenges posed in front of authors are tackled.

## 4 Conclusion

As per the protocols of our SLR, we extracted 1032 research papers and 136 studies were shortlisted for review. As we analyzed the papers, we found out there were multiple challenges in the field of credit risk evaluation. Each model comes with its own risks and challenges and cannot be relied completely upon for evaluation. A single complex classifier is not a solution to credit scoring and even for fraud detection according to the famous “no free lunch theorem” [164]. This is because of the problem of changing domains as discussed previously. Different banks from different geographic locations or even the same location will have different rules and regulations and thus the dataset will vary significantly. Hence, if we train the model on a dataset from one domain and test the model on a dataset from another domain, we will lose accuracy. Researchers are exploring this problem by applying ensemble techniques [88]. Ensemble techniques have proved to perform better than single classifiers [165–167]. Interpretability or readability of the results is a major drawback of ensemble learning. Therefore, improving the interpretability of ensemble models is another important research area which needs further exploration.

This study included only four digital databases for study selection, so it is possible that we may have missed some good studies on the topic. However, we are hopeful that we would have covered most of the major studies as we used snowballing approach in our search process as well with manual search to identify good studies. Another limitation of the study is that we did not validate or compare the findings or observations stated in some of the studies.

## 5 Discussion and future work

To solve the curse of dimensionality, applying feature selection methods is an important task. For feature selection approaches, there has been an increase in the use of GAs and Rough Sets [91, 99, 111, 119]. These algorithms are hybridized with other classifiers such as SVM to increase the accuracy of the model. Thus, hybridized models are becoming popular as more and more researchers are building such models. Their use has opened up a new area for exploration among researchers.

Another area which can be improved upon is data pre-processing of datasets. Datasets are made up of varying features or attributes. There can be redundant or recurring features in a dataset. This can lead to unnecessary computation and low accuracy. Thus, data pre-processing is an important step to improve the performance of a model. Piramuthu [168] discussed a few means to improve the performance of the classifiers through data pre-processing. However, there is room for improvement with more instances representing world scenarios. The data for NPA prediction that factors external, customer and bank features would help banks to implement early warning systems more effectively. Having datasets for various fraud topologies will enhance the usage of ML techniques with minimal false positives.

Deep learning is another area of machine learning which uses artificial neural networks (ANNs). We found out that deep learning can be useful in credit scoring and fraud detection. Further exploration is required in this area.

SVMs seem to be a better choice for solving the classification problem. SVM based approaches overcome the hurdles of overfitting and local optimum in ANN-based models. However, there are several challenges in applying SVM as discussed previously. More concrete research is needed if we want to increase the accuracy of the classifier using SVM. Deeper data processing and more suitable kernel function will help in increasing the accuracy. As the historical datasets are growing, there is a need to find out computationally inexpensive models that can deal with the dimensionality curse of the SVM. Harris [63] proposed a clustered SVM to address the problem. However, this work can further be

Table 11 Additional important results for different approaches in evaluating credit risk

Author	Approach	Dataset	Evaluation measure	Comments
Baesens et al. [55]	Neural Network Rule Extraction	German, two datasets from Benelux (Bene1, Bene2)	German - Train Acc: 81.38% (C4.5rules), Test Acc: 77.84% (Pruned NN), Bene1 - Train Acc: 77.76% (C4.5), Test Acc: 71.85% (Pruned NN), Bene2 - Train Acc: 82.80% (C4.5), Test Acc: 74.13% (Neurorule)	Comprehensibility problem addressed using neural network rule extraction and decision tables
Harris [63]	SVM, LR and variations	German Credit, Barbados Credit	German - Training Acc: 84.525% (CSVM-RBF), Test Acc: 78.700% (SVM-Linear), Balanced Accuracy (BAC): 65.428% (SVM-RBF), AUC: 65.983%, (SVM-linear)	To address the imbalanced nature of the German dataset, oversampling of the un-creditworthy class was done to get a 1:1 ratio
Martens et al. [78]	Comprehensible SVM	Australian Credit	Accuracy: 85.7% (SVM), Fidelity: 99.0% (Trepan)	Comprehensibility problem addressed using rule extraction techniques
Ribeiro et al. [115]	SVM+	Diane database	False Positive Rate: 3.51, False Negative Rate: 12.59, F1-score: 89.35%, Accuracy: 89.40%	Generalizability increased using additional information such as industry trends, economic environment, etc.
Feng et al. [66]	PCA-SVM	-	Training Accuracy: 97.55%, Testing Accuracy: 94.61%	Dimensionality reduction using PCA to reduce the computational time
Lai et al. [75]	Least Squares SVM	An English Financial Company's dataset	False Negative: 79.37%, True Positive: 93.27%, Overall: 89.16%, KS-stat: 58.88%	Quadratic problem converted to linear resulting in reduced complexity
Szwabe and Misiorek [87]	Tree based classifier	German Credit, Give Me Some Credit, Lending Club, Default of Credit Card Clients	AUC (%) for German: 80.79 (Random Forest), GMSC: 86.57 (Gradient Boost) LCL: 73.69 (Gradient Boost), DoCCC: 79.35 (Gradient Boost)	Pedagogical Tree Generation Algorithm and Random Samples-based Tree Generation Algorithm supported by black box classifiers (LR, Gradient Boosting and Random Forest) enable generation of decision trees that are interpretable by humans
Zięba et al. [105]	Ensemble classifier with switching labels	German Credit	Empirical risk value: 0.281, False Negative Rate: 23%	Switching class labels technique enabled solving dataset imbalance problem and asymmetric cost matrix problem
Pedro et al. [29]	LR, LSVM and Gradient Boosted Trees	Dataset of a credit card offering Latin American company	Average Precision: 29.2%, AUCROC: 72.5%	The problem of non existent credit history has been solved here using customer's mobile phone usage data

Table 11 (continued)

Author	Approach	Dataset	Evaluation measure	Comments
Krishna and Ravi [169]	Adaptive Differential Evolution Wrapper per	German, Financial Statement Fraud, Australian Credit	Sensitivity, Specificity, Accuracy: 85.3%, AUC: 0.821, Matthews Correlation Coefficient: (MCC) 0.64 for German Credit Dataset. Accuracy: 98.02%, AUC: 0.98, MCC: 0.96 for financial statement fraud dataset. Accuracy: 94.19%, AUC: 0.943, MCC: 0.88 for Australian Credit Dataset	Novel Feature Subset Selection Method ensures that there is no need to manually tune the hyperparameter values

improved in terms of area under the curve (AUC) and mean model training time.

The studies from our review collection state a need to develop more concrete tools which can address the problem of changing domains of datasets and also provide flexibility in adding any type of model to evaluate credit risk. A possible future work could be to combine rule-based, statistical and machine learning models into a single tool which would help in evaluating credit risk as per the requirements of the financial body. As most of the staff in banks are not technology savvy, building interfaces that do not require technical understanding but provide parallel processing, self-adaptation, self-learning, robustness and flexibility to assessors will enhance adoption of ML techniques.

## References

1. IMF says. <https://tinyurl.com/y5c4hyvj>. Accessed 13 Aug 2019
2. Jarrow RA, Turnbull SM (1995) Pricing derivatives on financial securities subject to credit risk. *J Financ* 50(1):53–85
3. Global Economy - NPA Concerns. <https://tinyurl.com/y5t17t5>, 2018. Accessed 13 Aug 2019
4. NPA countries. <https://tinyurl.com/y4a6n6bt>. Accessed 13 Aug 2019
5. IBM - Banking Analytics Services. <https://tinyurl.com/y2exerck>, 2018. Accessed 13 Aug 2019
6. McKinsey - Analytics in Banking. <https://tinyurl.com/y23ny7mv>, 2018. Accessed 13 Aug 2019
7. Khandani AE, Kim AJ, Lo AW (2010) Consumer credit risk models via machine learning algorithms. *J Bank Financ* 34(11):2767–2787
8. Akoglu L, Tong H, Koutra D (2015) Graph based anomaly detection and description: a survey. *Data Min Knowl Disc* 29(3):626–688
9. Bagherpour A (2017) Predicting mortgage loan default with machine learning methods
10. Galindo J, Tamayo P (2000) Credit risk assessment using statistical and machine learning: basic methodology and risk modeling applications. *Comput Econ* 15(1–2):107–143
11. Albrecht WS, Albrecht CO, Albrecht CC, Zimbelman MF (2011) *Fraud examination*. Cengage Learn
12. Hicks D, Caplain J, Faulkner N, Olcina E (2019) *Global banking fraud survey*
13. Hannes L, Peter SJ, Peltonen TA (2018) A framework for early-warning modeling with an application to banks. *European Central Bank*
14. Tony H, Stewart T, Kristin T (2009) The fourth paradigm: data-intensive scientific discovery. *Microsoft Research*
15. Bose I, Mahapatra RK (2001) Business data mining—a machine learning perspective. *Inf Manag* 39(3):211–225
16. Ravi V (2017) IDRBT Staff Papers - Analytics. <https://tinyurl.com/yy9s85pd>. Accessed 13 Aug 2019
17. Vives X (2017) The impact of FinTech on banking. *Eur Econ* 2:97–105
18. Jagtiani JA, Lemieux CM (2019) The roles of alternative data and machine learning in fintech lending: evidence from the lending-club consumer platform
19. Ann KB, David B, Pearl B (2015) *Evidence-based software engineering and systematic reviews*. Chapman & Hall/CRC, New York



20. Russell Stuart J, Peter N (2016) Artificial intelligence: a modern approach. Pearson Education Limited, Malaysia
21. AI Boom and subsequent winter. <https://tinyurl.com/y2yk9kqb>. Accessed 13 Aug 2019
22. SLR files. <https://tinyurl.com/SLRfiles>. Accessed 13 Aug 2019
23. Marikkannu P, Shanmugapriya K (2011) Classification of customer credit data for intelligent credit scoring system using fuzzy set and mc2—domain driven approach. In: 2011 3rd International conference on electronics computer technology, 3:410–414
24. Romanyuk K (2015) Concept of a decision support system for a loan granting based on continuous price function. In: 2015 SAI intelligent systems conference (IntelliSys), pp 105–111
25. Wei R (2008) Development of credit risk model based on fuzzy theory and its application for credit risk management of commercial banks in china. In: 2008 4th International conference on wireless communications, networking and mobile computing, pp 1–4
26. Hoffmann F, Bart B, Christophe M, Van GT, Jan V (2007) Inferring descriptive and approximate fuzzy rules for credit scoring using evolutionary algorithms. *Eur J Oper Res* 177:540–555
27. Kotsiantis SB, Kanellopoulos D, Karioti V, Tampakas V (2009) An ontology-based portal for credit risk analysis. In: 2009 2nd IEEE international conference on computer science and information technology, 165–169
28. Baesens B, Mues C, De Backer M, Vanthienen J, Setiono R (2005) Building intelligent credit scoring systems using decision tables. In: Camp O, Filipe JBL, Hammoudi S, Piattini M (eds) *Enterprise Information Systems V*, pp 131–137, Springer, Amsterdam
29. Pedro JS, Proserpio D, Oliver N (2015) Mobiscore: towards universal credit scoring from mobile phone data. In: Ricci F, Bontcheva K, Conlan O, Lawless S (eds), *User modeling, adaptation and personalization*, pp 195–207, Springer, Cham
30. Xia Y, Liu C, Da B, Xie F (2017) A novel heterogeneous ensemble credit scoring model based on bstacking approach. *Expert Syst Appl* 93:10
31. Mahmoud M, Algadi N, Ali A (2008) Expert system for banking credit decision. In: 2008 International conference on computer science and information technology, 813–819
32. Lin S, Wu S-J, Ma H-L, Wu D-B (2009) Development of credit risk model in banking industry based on gra. In: 2009 International conference on machine learning and cybernetics, 5, pp 2903–2909
33. Huang J, Chen M (2018) Domain adaptation approach for credit risk analysis. In: *Proceedings of the 2018 international conference on software engineering and information management, ICSIM2018*, pp 104–107, New York, NY
34. Gadi MFA, Wang X, do Lago AP (2008) Credit card fraud detection with artificial immune system. In: Bentley PJ, Lee D, Jung S (eds), *Artificial immune systems*, pp 119–131, 2008. Springer, Berlin
35. Duman E, Özçelik M (2011) Detecting credit card fraud by genetic algorithm and scatter search. *Expert Syst Appl* 38:13057–13063
36. Van Vlasselaer V, Bravo C, Caelen O, Eliassi-Rad T, Akoglu L, Snoeck M, Baesens B (2015) Apatate: a novel approach for automated credit card transaction fraud detection using network-based extensions. *Decis Support Syst* 75:05
37. Vac Gelu I, Găban Lucian V (2016) A new perspective over the risk assessment in credit scoring analysis using the adaptive reference system. In: Abramowicz W, Alt R, Franczyk B (eds), *Business information systems*, pp 130–143, Springer, Cham
38. Liu K, Lai KK, Guu S (2009) Dynamic credit scoring on consumer behavior using fuzzy markov model. In: 2009 Fourth International Multi-Conference on Computing in the Global Information Technology, 235–239
39. Anzilli L, Facchinetti G, Pirotti T (2017) Credit risk profiling using a new evaluation of interval-valued fuzzy sets based on alpha-cuts. In: 2017 IEEE international conference on fuzzy systems (FUZZ-IEEE), 1–6
40. Qiao H, Dong X-C (2009) Research on the risk evaluation in loan projects of commercial bank in financial crisis. *Int Conf Mach Learn Cybern* 2:776–781
41. Yazdani H, Kwasnicka H (2012) Fuzzy classification method in credit risk. In: Nguyen N-T, Hoang K, Jedrzejowicz P (eds), *Computational collective intelligence. Technologies and applications*, Springer, Berlin, pp 495–504
42. Kültür Y, Çağlayan M U (Oct 2015) A novel cardholder behavior model for detecting credit card fraud. In: 2015 9th International conference on application of information and communication technologies (AICT), pp 148–152
43. Li G, Wu Y (2010) Empirical research about credit risk on neural network based bp algorithm. In: 2010 3rd international conference on information management, innovation management and industrial engineering, 3:461–463
44. Li R-Z, Pang S-L, Xu J-M (2002) Neural network credit-risk evaluation model based on back-propagation algorithm. In: *Proceedings of international conference on machine learning and cybernetics*, 4:1702–1706 vol.4
45. Zhu C, Zhan Y, Jia S (2010) Research on bp neural network evaluation model of credit risk of bank clients. In: 2010 International conference on management and service science, 1–5
46. Dima AM, Vasilache S (2009) Ann model for corporate credit risk assessment. In: *International conference on information and financial engineering*, 94–98
47. Hu X-Y, Tang Y (2006) Ann-based credit risk identification and control for commercial banks. 2006:3110 – 3114, 09
48. Wang F, Song Z (2008) Research on the credit risk evaluation and forecast of housing mortgage loans. In: 2008 International symposium on intelligent information technology application workshops, pp 940–943
49. Peng Y, Tu X (2005) A study on the ann-based credit risk prediction model and its application. In: Li D, Wang B (eds), *Artificial intelligence applications and innovations*, pages 459–468, Boston, MA, Springer US
50. Bozsik J, Ilonczai Z (2012) Echo state network-based credit rating system. In: 4th IEEE international symposium on logistics and industrial informatics, pp 185–190
51. Lai KK, Yu L, Wang S, Zhou L (2006) Neural network meta-learning for credit scoring. In: Huang D-S, Li K, Irwin GW (eds), *Intelligent computing*, Springer, Berlin, pp 403–408
52. Hsieh N-C (2005) Hybrid mining approach in the design of credit scoring models. *Expert Syst Appl* 28:655–665
53. Derelioğlu G, Gürgeç F, Okay N (2009) A neural approach for sme's credit risk analysis in turkey. In: Perner P (ed), *Machine learning and data mining in pattern recognition*, Springer, Berlin, pp 749–759
54. Lean Y, Shouyang W, Keung LK (2008) Credit risk assessment with a multistage neural network ensemble learning approach. *Expert Syst Appl* 34:1434–1444, 02
55. Baesens B, Setiono R, Mues C, Vanthienen J (2003) Using neural network rule extraction and decision tables for credit-risk evaluation. *Manag Sci* 49:312–329
56. Marin-de-la-Barcelona A, Marcano-Cedeño A, Jimenez-Trillo J, Piñuela J A, Andina D (2010) Artificial metaplasticity: an approximation to credit scoring modeling. In: *IECON 2010 - 36th annual conference on IEEE industrial electronics society*, pp 2817–2822
57. Tomczak J, Zięba M (2014) Classification restricted boltzmann machine for comprehensible credit scoring model. *Expert Syst Appl* 42:10

58. Zhang Y, Wang D, Chen Y, Zhao Y, Shao P, Meng Q (2017) Credit risk assessment based on flexible neural tree model. In: Cong F, Leung A, Wei Q (eds), *Advances in neural networks - ISNN 2017*, Springer, Cham, pp 215–222
59. Zhaoji Y, Qiang M, Wenjuan W (2010) The application of wn based on pso in bank credit risk assessment. In: 2010 International conference on artificial intelligence and computational intelligence, 3:444–448
60. Fan Q, Yang J (2018) A denoising autoencoder approach for credit risk analysis. In: *Proceedings of the 2018 international conference on computing and artificial intelligence, ICCAI 2018*, New York, NY, pp 62–65
61. Timofeev N, Timofeeva G (2013) Estimation of loan portfolio risk on the basis of markov chain model. In: Hömberg D, Tröltzsch F (eds), *System modeling and optimization*, Springer, Berlin, pp 207–216
62. Farquah MAH, Sriramjee VR, Praveen G (2011) Credit scoring using pca-svm hybrid model. In: Das VV, Stephen J, Chaba Y (eds), *Computer networks and information technologies*, Springer, Berlin
63. Harris T (2015) Credit scoring using the clustered support vector machine. *Expert Syst Appl* 42:741–750, 02
64. Huang S-C (2009) Integrating nonlinear graph based dimensionality reduction schemes with svms for credit rating forecasting. *Expert Syst Appl* 36:7515–7518, 05
65. Li Z (2016) A new method of credit risk assessment of commercial banks. In: 2016 International Conference on Robots Intelligent System (ICRIS), p 34–37
66. Feng W, Zhao Y, Deng J (2009) Application of svm based on principal component analysis to credit risk assessment in commercial banks. In: 2009 WRI Global Congress on Intelligent Systems, 4, p 49–52
67. Lv G, Peng L (2008) Commercial banks' credit risk assessment based on rough sets and svm. In: 2008 4th International Conference on Wireless Communications, Networking and Mobile Computing, p 1–4
68. Yang C-G, Duan X-B (2008) Credit risk assessment in commercial banks based on svm using pca. In: 2008 International Conference on Machine Learning and Cybernetics, volume 2, pp 1207–1211
69. Wei L, Li W, Xiao Q (2016) Credit risk evaluation using: Least squares support vector machine with mixture of kernel. In: 2016 International Conference on Network and Information Systems for Computers (ICNISC), p 237–241
70. Zhu C, Zhan Y, Jia S (2010) Credit risk identification of bank client basing on supporting vector machines. In: 2010 Third International Conference on Business Intelligence and Financial Engineering, p 62–66
71. Ma Y, Liu H (2010) Research of svm applying in the risk of bank's loan to enterprises. In: 2010 2nd International Conference on Information Engineering and Computer Science, p 1–5
72. A least squares fuzzy SVM approach to credit risk assessment, pp. 73–84. Springer, Berlin (2008)
73. Sun W, Yang C (2006) Credit risk assessment in commercial banks based on multi-layer svm classifier. In: Huang D-S, Li K, Irwin GW (eds), *Computational intelligence*, pp 778–785. Springer, Berlin
74. Wei L, Li J, Chen Z-Y (2007) Credit risk evaluation using support vector machine with mixture of kernel. 4488:431–438, 05
75. Lai Kin Keung, Yu Lean, Zhou Ligang, Wang Shouyang (2006) Credit risk evaluation with least square support vector machine. In: Wang Guo-Ying, Peters James F, Skowron Andrzej, Yao Yiyu (eds), *Rough Sets and Knowledge Technology*, p 490–495, Berlin, Heidelberg. Springer Berlin Heidelberg
76. Li Jianping, Liu Jingli, Xu Weixuan, Shi Yong (2004) Support vector machines approach to credit assessment. In: Bubak Marian, van Albada Geert Dick, Sloot Peter M A, Dongarra Jack, (eds), *Computational Science - ICCS 2004*, p 892–899, Berlin, Heidelberg. Springer Berlin Heidelberg
77. Gestel Tony Van, Baesens Bart, Suykens Johan A K, Van den Poel Dirk, Baestaens Dirk-Emma, Willekens Marleen (2006) Bayesian kernel based classification for financial distress detection. *European Journal of Operational Research*, 172:979–1003
78. Martens D, Baesens B, Van Gestel T, Vanthienen J (2007) Comprehensive credit scoring models using rule extraction from support vector machines. *Eur J Oper Res* 183:1466–1476
79. Van Gestel Tony, Baesens Bart, Garcia Joao, Van Dijke Peter (2003) A support vector machine approach to credit scoring
80. Vedala R, Kumar B R (2012) An application of naive bayes classification for credit scoring in e-lending platform. In: 2012 International Conference on Data Science Engineering (ICDSE), p 81–84
81. Okesola Olatunji, Okokpujie Kennedy, Adewale Adeyinka, John Samuel, Omoruyi Osemwegie (2017) An improved bank credit scoring model: A naïve bayesian approach. p 228–233, 12
82. Benyacoub B, El Bernoussi S, Zoglat A (2014) Building classification models for customer credit scoring. In: 2014 International Conference on Logistics Operations Management, p 107–111
83. Petropoulos Anastasios, Chatzis Sotirios, Xanthopoulos S (2016) A novel corporate credit rating system based on student's-t hidden markov models. *Expert Systems with Applications*, 53, 01
84. Vieira Armando, Duarte João, Ribeiro Bernardete, Neves Joao Carvalho (2009) Improving personal credit scoring with hlvq-c. In: Köppen Mario, Kasabov Nikola, Coghill George, (eds), *Advances in Neuro-Information Processing*, p 97–103, Berlin, Heidelberg. Springer Berlin Heidelberg
85. Wei G, Yingjie S, Mu Y X (2015) Commercial bank credit risk evaluation method based on decision tree algorithm. In: 2015 Seventh International Conference on Measuring Technology and Mechatronics Automation, p 285–288
86. Lang J, Sun J (2014) Sensitivity of decision tree algorithm to class-imbalanced bank credit risk early warning. In: 2014 Seventh International Joint Conference on Computational Sciences and Optimization, p 539–543
87. Szwabe Andrzej, Misiorek Pawel (2018) Decision trees as interpretable bank credit scoring models. In: Kozielski Stanisław, Mrozek Dariusz, Kasprowski Paweł, Małysiak-Mrozek Bożena, Kostrzewa Daniel, (eds), *Beyond Databases, Architectures and Structures. Facing the Challenges of Data Proliferation and Growing Variety*, p 207–219, Cham. Springer International Publishing
88. Xia Y, Liu C, Li YY, Liu N (2017) A boosted decision tree approach using bayesian hyper-parameter optimization for credit scoring. *Expert Syst Appl* 78:02
89. Fu Hui, Liu Xiaoyong (2011) A hybrid model for credit evaluation problem. In: Tan Ying, Shi Yuhui, Chai Yi, Wang Guoyin, (eds), *Advances in Swarm Intelligence*, pages 626–634, Berlin, Heidelberg. Springer Berlin Heidelberg
90. Ruiz Saulo, Gomes Pedro, Rodrigues Luís, Gama João (2017) Credit scoring in microfinance using non-traditional data. In: Oliveira Eugénio, Gama João, Vale Zita, Lopes Cardoso Henrique, (eds), *Progress in Artificial Intelligence*, p 447–458, Cham, (2017). Springer International Publishing
91. Cheng-Lung H, Mu-Chen C, Chieh-Jen W (2007) Credit scoring with a data mining approach based on support vector machines. *Expert Syst Appl* 33:847–856, 11
92. Oreski S, Oreski D, Oreški G (2012) Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment. *Expert Syst Appl* 39:12605–12617, 11
93. Taremian HR, Naeini MP (2011) Hybrid intelligent decision support system for credit risk assessment. In: 2011 7th International

- Conference on Information Assurance and Security (IAS), p 167–172
94. Rodan Ali, Faris Hossam (2016) Credit risk evaluation using cycle reservoir neural networks with support vector machines readout. In: Nguyen Ngoc Thanh, Trawiński Bogdan, Fujita Hamido, Hong Tzung-Pei, (eds), *Intelligent Information and Database Systems*, p 595–604, Berlin, Heidelberg. Springer Berlin Heidelberg
  95. Weidong Huang, Xiangwei Zhu (2010) SuQingling. Research on application of personal credit scoring based on bp-logistic hybrid algorithm. In: 2010 International Conference on Computer Application and System Modeling (ICASM 2010), 4:V4–735–V4–739
  96. Djemaiel Yacine, Labidi Nadia, Boudriga Nouredine (2016) A dynamic hybrid rbf/elman neural networks for credit scoring using big data. In: Abramowicz Witold, Alt Rainer, Franczyk Bogdan, (eds), *Business Information Systems*, p 102–113, Cham. Springer International Publishing
  97. Huang Y, Tian C (2008) Research on credit risk assessment model of commercial banks based on fuzzy probabilistic neural network. In: 2008 International Conference on Risk Management Engineering Management, p 482–486
  98. Huang Bo, Zhang Qing-Pu, Hu Yun-Quan (2005) Research on credit risk management of the state-owned commercial bank. In: 2005 International Conference on Machine Learning and Cybernetics, volume 7, pages 4038–4043 Vol. 7
  99. Zhou J, Bai T (2008) Credit risk assessment using rough set theory and ga-based svm. In: 2008 The 3rd International Conference on Grid and Pervasive Computing - Workshops, p 320–325
  100. Jiang Ming-hui, Yuan Xu-chuan (2007) Construction and application of pso-svm model for personal credit scoring. In Yong Shi, Geert Dick van Albada, Jack Dongarra, and Peter M. A. Sloot, editors, *Computational Science – ICCS 2007*, pages 158–161, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg
  101. Hao Yanyou, Chi Zhongxian, Yan Deqin, Yue Xun (2007) An improved fuzzy support vector machine for credit rating. In: Li Keqiu, Jesshope Chris, Jin Hai, Gaudiot Jean-Luc, (eds), *Network and Parallel Computing*, pages 495–505, Berlin, Heidelberg. Springer Berlin Heidelberg
  102. Van Gestel Tony, Baesens Bart, Van Dijcke Peter, Suykens Johan A K, Garcia Joao (2005) Linear and non-linear credit scoring by combining logistic regression and support vector machines
  103. Alaraj M, Abbod M (2016) Classifiers consensus system approach for credit scoring. *Knowl-Based Syst* 104:04
  104. Zhen W, Wenjuan S (2016) Commercial bank credit risk assessment method based on improved svm. In: 2016 International Conference on Intelligent Transportation, Big Data Smart City (ICITBS), p 353–356
  105. Zięba Maciej, Świątek Jerzy (2012) Ensemble classifier for solving credit scoring problems. In Camarinha-Matos Luis M, Shahamatnia Ehsan, Nunes Gonçalo, (eds), *Technological Innovation for Value Creation*, p 59–66, Berlin, Heidelberg. Springer Berlin Heidelberg
  106. Lai Kin Keung, Yu Lean, Wang Shouyang, Zhou Ligang (2006) Credit risk analysis using a reliability-based neural network ensemble model. In: Kollias Stefanos, Stafylopatis Andreas, Duch Włodzisław, Oja Erkki, (eds), *Artificial Neural Networks – ICANN 2006*, p 682–690, Berlin, Heidelberg. Springer Berlin Heidelberg
  107. Hsieh Nan-Chen, Hung Lun-Ping, Ho Chia-Ling (2009) A data driven ensemble classifier for credit scoring analysis. In: Theeramunkong Thanaruk, Kijssirikul Boonserm, Cercone Nick, Ho Tu-Bao, (eds), *Advances in Knowledge Discovery and Data Mining*, pages 351–362, Berlin, Heidelberg. Springer Berlin Heidelberg
  108. Chen H, Jiang M, Wang X (2017) Bayesian ensemble assessment for credit scoring. In: 2017 4th International Conference on Industrial Economics System and Industrial Security Engineering (IEIS), p 1–5
  109. Chou T (2007) A novel prediction model for credit card risk management. In: *Second International Conference on Innovative Computing, Informatio and Control (ICICIC 2007)*, p 211–211
  110. Makrygianni Ira I, Markopoulos Angelos P (2016) Loan evaluation applying artificial neural networks. In: *Proceedings of the SouthEast European Design Automation, Computer Engineering, Computer Networks and Social Media Conference, SEEDA-CECNSM '16*, pages 124–128, New York, NY, USA. ACM
  111. Zhang Z (2011) Research of default risk of commercial bank's personal loan based on rough sets and neural network. In: 2011 3rd International Workshop on Intelligent Systems and Applications, p 1–4
  112. Miglionico Maria Cristina, Parillo Fernando (2012) An application in bank credit risk management system employing a bp neural network based on sfloat24 custom math library using a low cost fpga device. In: Salvatore Greco, Bernadette Bouchon-Meunier, Giulianella Coletti, Mario Fedrizzi, Benedetto Matarazzo, and Ronald R. Yager, editors, *Advances in Computational Intelligence*, pages 84–93, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg
  113. Chen Ya-Qi, Zhang Jianjun, Ng Wing (2018) Loan default prediction using diversified sensitivity undersampling. 240–245, 07
  114. Feki A, Ishak A, Feki S (2012) Feature selection using bayesian and multiclass support vector machines approaches: Application to bank risk prediction. *Expert Syst Appl* 39:3087–3099, 02
  115. Ribeiro B, Silva C, Chen N, Vieira A, Carvalho das Neves J (2012) Enhanced default risk models with svm+. *Expert Syst Appl* 39:10140–10152, 09
  116. Oguz H T, Gurgun F S (2008) Credit risk analysis using hidden markov model. In: 2008 23rd International Symposium on Computer and Information Sciences, 1–5
  117. Ni Weijian, Liu Tong, Zeng Qingtian, Zhang Xianke, Duan Hua, Xie Nengfu (2018) Robust factorization machines for credit default prediction. In: Geng Xin, Kang Byeong-Ho, (eds), *PRICAI 2018: Trends in Artificial Intelligence*, pages 941–953, Cham, 2018. Springer International Publishing
  118. Masmoudi K, Abid L, Masmoudi A (2019) Credit risk modeling using bayesian network with a latent variable. *Expert Syst Appl* 127:03
  119. Zhao Zhenyu, Zhang Wei, Zhou Yayue (2011) National student loans credit risk assessment based on gabp algorithm of neural network. In: 2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), 2196–2199
  120. Su J, Zhang Y (2017) Application of bp neural network optimization algorithm based on genetic algorithm in credit risk early-warning of commercial bank. In: 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), p 487–491
  121. Yao Ping, Wu Chong, Yao Minghui (2009) Credit risk assessment model of commercial banks based on fuzzy neural network. In: Yu Wen, He Haibo, Zhang Nian (eds), *Advances in Neural Networks – ISNN 2009*, p 976–985, Berlin, Heidelberg. Springer Berlin Heidelberg
  122. Mubarek A M, Adali E (2017) Multilayer perceptron neural network technique for fraud detection. In: 2017 International Conference on Computer Science and Engineering (UBMK), p 383–387
  123. Patil P S, Dharwadkar N V (2017) Analysis of banking data using machine learning. In: 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), p 876–881
  124. Ghobadi F, Rohani M (2016) Cost sensitive modeling of credit card fraud using neural network strategy. In: 2016 2nd

- International Conference of Signal Processing and Intelligent Systems (ICSPIS), pages 1–5
125. Behera T K, Panigrahi S (2015) Credit card fraud detection: A hybrid approach using fuzzy clustering and neural network. In: 2015 Second International Conference on Advances in Computing and Communication Engineering, pages 494–499
  126. Zhan Qing, Yin Hang (2018) A loan application fraud detection method based on knowledge graph and neural network. In: Proceedings of the 2Nd International Conference on Innovation in Artificial Intelligence, ICIAI '18, pages 111–115, New York, NY, USA. ACM
  127. Kazemi Z, Zarrabi H (2017) Using deep networks for fraud detection in the credit card transactions. In: 2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI), pages 0630–0633
  128. Zamini Mohamad, Montazer Gholamali (2018) Credit card fraud detection using autoencoder based clustering. pages 486–491, 12
  129. Liu Ou, Ma Jian, Poon Pak-Lok, Zhang Jun (2009) On an ant colony-based approach for business fraud detection. In: Huang De-Shuang, Jo Kang-Hyun, Lee Hong-Hee, Kang Hee-Jun, Bevilacqua Vitoantonio, (eds), Emerging Intelligent Computing Technology and Applications, pages 1104–1111, Berlin, Heidelberg. Springer Berlin Heidelberg
  130. Charleonnan A (Oct 2016) Credit card fraud detection using rus and mrn algorithms. In: 2016 Management and Innovation Technology International Conference (MITicon), pages MIT-73–MIT-76
  131. Bouchti A E, Chakroun A, Abbar H, Okar C (2017) Fraud detection in banking using deep reinforcement learning. In: 2017 Seventh International Conference on Innovative Computing Technology (INTECH), pages 58–63
  132. Abakarim Youness, Lahby Mohamed, Attiou Abdelbaki (2018) An efficient real time model for credit card fraud detection based on deep learning. In: Proceedings of the 12th International Conference on Intelligent Systems: Theories and Applications, SITA'18, pages 30:1–30:7, New York, NY, USA. ACM
  133. Karlos Stamatis, Kostopoulos Georgios, Kotsiantis Sotiris, Tampakas Vassilis (2017) Using active learning methods for predicting fraudulent financial statements. In Giacomo Boracchi, Lazaros Iliadis, Chrisina Jayne, and Aristidis Likas, editors, Engineering Applications of Neural Networks, pages 351–362, Cham. Springer International Publishing
  134. Jiang C, Song J, Liu G, Zheng L, Luan W (2018) Credit card fraud detection: A novel approach using aggregation strategy and feedback mechanism. IEEE Internet of Things Journal 5(5):3637–3647
  135. Rahmawati D, Sarno R, Fatichah C, Sunaryono D (2017) Fraud detection on event log of bank financial credit business process using hidden markov model algorithm. In: 2017 3rd International Conference on Science in Information Technology (ICSITech), pages 35–40
  136. Khan A, Singh T, Sinhal A (2012) Implement credit card fraudulent detection system using observation probabilistic in hidden markov model. In: 2012 Nirma University International Conference on Engineering (NUiCONE), pages 1–6
  137. Gyamfi N K, Abdulai J (2018) Bank fraud detection using support vector machine. In: 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), pages 37–41
  138. Kotsiantis Sotiris, Koumanakos Euaggelos, Tzelepis Dimitris, Tampakas Vasilis (2006) Predicting fraudulent financial statements with machine learning techniques. In: Antoniou Grigoris, Potamias George, Spyropoulos Costas, Plexousakis Dimitris, (eds), Advances in Artificial Intelligence, pages 538–542, Berlin, Heidelberg. Springer Berlin Heidelberg
  139. Pang Sulin, Yuan Jimeng (2018) Wt model & applications in loan platform customer default prediction based on decision tree algorithms. In: Huang De-Shuang, Bevilacqua Vitoantonio, Premaratne Prashan, Gupta Phalguni, (eds), Intelligent Computing Theories and Application, pages 359–371, Cham. Springer International Publishing
  140. Mareeswari V, Gunasekaran G (2016) Prevention of credit card fraud detection based on hsvm. In: 2016 International Conference on Information Communication and Embedded Systems (ICICES), pages 1–4
  141. Montini Denis, Matuck Gustavo, de Avila Montini Alessandra, Cunha Adilson, Ribeiro Alexandre, Dias Luiz (2013) A sampling diagnostics model for neural system training optimization. Proceedings of the 2013 10th International Conference on Information Technology: New Generations, ITNG 2013, 04
  142. Kamaruddin Sk, Ravi Vadlamani (2016) Credit card fraud detection using big data analytics: Use of psaoann based one-class classification. In: Proceedings of the International Conference on Informatics and Analytics, ICIA-16, pages 33:1–33:8, New York, NY, USA. ACM
  143. Malhotra Rashmi, Malhotra DK (2003) Evaluating consumer loans using neural networks. Omega, 31:83–96, 04
  144. Huang S, Day M (2013) A comparative study of data mining techniques for credit scoring in banking. In: 2013 IEEE 14th International Conference on Information Reuse Integration (IRI), pages 684–691
  145. Khemakhem Sihem, Boujelbene Younes (2017) Artificial intelligence for credit risk assessment: Artificial neural network and support vector machines. ACRN Oxford Journal of Finance and Risk Perspectives, 6:1–17, 01
  146. Nwulu Nnamdi I, Oroja Shola, Ilkan Mustafa (2011) Credit scoring using soft computing schemes: A comparison between support vector machines and artificial neural networks. In: Ariwa Ezendu, El-Qawasmeh Eyas, (eds), Digital Enterprise and Information Systems, pages 275–286, Berlin, Heidelberg. Springer Berlin Heidelberg
  147. FISHER R A The use of multiple measurements in taxonomic problems. Annals of Eugenics, 7(2):179–188
  148. Wiginton JC (1980) A note on the comparison of logit and discriminant models of consumer credit behavior. Journal of Financial and Quantitative Analysis 15(3):757–770
  149. Grablowsky Bernie J, Talley Wayne K (1981) Probit and discriminant functions for classifying credit applicants: a comparison
  150. Glover Fred Improved linear programming models for discriminant analysis\*. Decision Sciences, 21(4):771–785
  151. Mangasarian OL (1965) Linear and nonlinear separation of patterns by linear programming. Oper Res 13(3):444–452
  152. Henley WE, Hand DJ (1996) A k-nearest-neighbour classifier for assessing consumer credit risk. The Statistician 45(1):77
  153. Mays E (2001) Handbook of Credit Scoring. Global Professional Publishing, Business Series
  154. Li S-T, Shiue W, Huang M-H (2006) The evaluation of consumer loans using support vector machines. Expert Syst Appl 30:772–782, 05
  155. Hand D J, Henley W E Statistical classification methods in consumer credit scoring: a review. Journal of the Royal Statistical Society: Series A (Statistics in Society), 160(3):523–541
  156. Pandey T N, Jagadev A K, Mohapatra S K, Dehuri S (2017) Credit risk analysis using machine learning classifiers. In: 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), pages 1850–1854
  157. Singh P (2017) Comparative study of individual and ensemble methods of classification for credit scoring. In: 2017 International Conference on Inventive Computing and Informatics (ICICI), pages 968–972

158. Ravisankar P, Ravi V, Raghava Rao G, Bose I (2011) Detection of financial statement fraud and feature selection using data mining techniques. *Decis Support Syst* 50(2):491–500
159. Why SVM is quadratic. <https://tinyurl.com/y2xre5zn>. Accessed: 2019-08-13
160. Baesens B, Setiono R, Mues C, Vanthienen J (2003) Using neural network rule extraction and decision tables for credit-risk evaluation. *Manage Sci* 49(3):312–329
161. Crouhy M, Galai D, Mark R (2000) A comparative analysis of current credit risk models. *J Bank Finance* 24:59–117, 01
162. Lessmann S, Baesens B, Seow H-V, Thomas LC (2015) Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *Eur J Oper Res* 247:124–136
163. Universities working in the area of credit risk evaluation. <https://tinyurl.com/CreditRiskUniversities>. Accessed: 2019-08-13
164. Wolpert DH, Macready WG (1997) No free lunch theorems for optimization. *IEEE Trans Evol Comput* 1(1):67–82
165. Hung C, Chen J-H (2009) A selective ensemble based on expected probabilities for bankruptcy prediction. *Expert Syst Appl* 36:5297–5303, 04
166. Yu L, Wang S, Lai KK (2008) Credit risk assessment with a multistage neural network ensemble learning approach. *Expert Syst Appl* 34(2):1434–1444
167. Wang G, Hao J, Ma J, Jiang H (2011) A comparative assessment of ensemble learning for credit scoring. *Expert Syst Appl* 38(1):223–230
168. Piramuthu S (2006) On preprocessing data for financial credit risk evaluation. *Expert Syst Appl* 30:489–497, 04
169. Krishna, Gutha Jaya and Ravi, Vadlamani. Feature Subset Selection Using Adaptive Differential Evolution: An Application to Banking. *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data*, 157–163, 2019
170. Popat R R, Chaudhary J (2018) A Survey on Credit Card Fraud Detection Using Machine Learning 2nd International Conference on Trends in Electronics and Informatics (ICOEI), 1120–1125
171. Dastile Xolani, Celik Turgay, Potsane Moshe (2020) Statistical and machine learning models in credit scoring: A systematic literature survey <http://www.sciencedirect.com/science/article/pii/S1568494620302039>, 106263
172. Bakshi S (2018) Credit Card Fraud Detection : A classification analysis 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), 152–156