Expert Systems With Applications 198 (2022) 116847

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

PsyCredit: An interpretable deep learning-based credit assessment approach facilitated by psychometric natural language processing

Kai Yang[a], Hui Yuan[b,*], Raymond Y.K. Lau[a]

[a] *Department of Information Systems, College of Business, City University of Hong Kong, Hong Kong*
[b] *AI and Data Science Application Center, School of Business and Management, Shanghai International Studies University, Shanghai 201620, China*

ARTICLE INFO

Keywords:
Credit risk
Personality traits
Social media
Psycholinguistic
Deep learning
Explainable artificial intelligence

ABSTRACT

With the prosperity of the social web, individuals' social media information alleviates the information asymmetry between individuals and online financial institutions, e.g., online lending and has been applied to predict their credit scores. Most existing studies use semantic or sentiment-related information excavated from their textual postings to construct credit evaluation models. However, despite the essential role of borrowers' personalities on their financial decisions, psychological factors, which can also be mined from their personally written text, receive less attention in current literature. It is challenging to apply extant psychometric approaches for online credit assessment tasks. Specifically, under the chaotic social media environment, social media postings published by the borrowers may not be composed by themselves, and therefore their real psychological statuses are difficult to be uncovered through existing approaches. To solve this problem, guided by the design science methodology and grounded on the Systemic Functional Linguistic Theory, we propose a novel IT artifact, named as PsyCredit, which is a deep learning-based online risk assessment framework driven by a novel psychometric approach. Unlike traditional deep learning approaches, which is a black box, results given by PsyCredit are interpretable by leveraging the Layer-wise Relevance Propagation technique, for the sake of high usability. Based on a dataset from a real-world P2P lending company, our experiments verify that, by leveraging the proposed psychometric approach, the credit risk assessment performance gets promotion successfully.

1. Introduction

The growth of social media has raised ample opportunities for business and finance areas (Aral et al., 2013). Scholars have focused on the impact of social media on financial activities at the individual level (Engelberg and Parsons, 2011, Zhu et al., 2012, Goh et al., 2013). In the field of online lending, Burtch et al. (2014) note that information asymmetry is one of the core issues, since borrowers clearly understand their own credit quality, while investors cannot. To alleviate the information asymmetry, extant studies have leveraged social media as an implement information source to describe the behavior characteristics of the borrowers (Zhang et al., 2016). As outlined by Guo et al. (2016), the textual content generated by users is essential when discriminating between good and bad credit users. Existing approaches designed for mining textual data are mainly targeted at sentiment polarities, N-gram features (Guo et al., 2016), and semantic features (Liang & He, 2020). However, the personality of the borrowers, which can be identified through their written language as verified by Mairesse et al. (2007), had yet been explored in the field of online P2P lending, even though borrowers' personality had been proven to be closely associated with their financing behaviors and credit scores (Davey and George, 2011, Bernerth et al., 2012). This paper makes early attempts to answer the following two research questions: 1) how to uncover borrowers' personality through their personal writings in social media; 2) how to incorporate borrowers' personality revealed by their social media postings into the credit assessment task for online P2P lending platforms.

In the field of psycholinguistic, which is the study of the interrelation between linguistic factors and mental aspects (Traxler and Gernsbacher, 2011), researchers have confirmed that individuals' personalities can be revealed by their written language. Specifically, the semantic or syntactic information embedded in the language is the essential cue for discriminating different personality traits, e.g., extraverts use more positive words and show more agreement than introverts (Pennebaker & King, 1999). Accordingly, with the development of natural language processing (NLP) technique recent years, the personality recognition

* Corresponding author.
E-mail address: huiyuan@shisu.edu.cn (H. Yuan).

https://doi.org/10.1016/j.eswa.2022.116847
Received 7 February 2021; Received in revised form 22 November 2021; Accepted 6 March 2022
Available online 12 March 2022

approaches facilitated by automatic NLP techniques, which are also named as psychometric NLP approach by Ahmad et al. (2020), have been confirmed useful by extant studies. Although several studies have made early attempts to detect personality traits through individuals' social media postings (Pratama and Sarno, 2015, Yu and Markov, 2017), to our best knowledge, such psychometric NLP approaches have not yet been introduced to the field of online finance, e.g., online lending. In this paper, we propose a novel psychometric NLP approach that is specifically designed to detect borrowers' personalities through their social media postings for credit assessment.

There are some challenges restricting the wide application of psychometric NLP approaches for credit risk assessment in online background. Firstly, as a restriction of psychometric NLP approaches, only the personalities of the author, but not other people, can be uncovered through textual writings, and thus the originality of the text should be ensured. However, under the chaotic social media environment, this premise does not always hold. Specifically, it is easy for borrowers to copy and paste postings from other users to polish their social media content so as to gain more credit. We have 4 postings sent by the same user as examples to introduce non-original microblogs – 1) "The Mid-Autumn Festival is coming! From now until September 5th, all purchases of XXX will receive a 10% discount. Copy and forward this Tweet to receive a beautiful gift. Come on! What are you waiting for?" 2) "So cute ~~" 3) "Come on, rush forward!" 4) "I want to go on holiday with you". As the example shown in above, apparently, the language style of Microblog 1 is different from that of others (it looks like advertisement), and thus it is likely that Microblog 1 is copied from other places. This problem has been omitted by previous studies, and consequently, the personality detection results would be less accurate when such non-original postings are applied. Secondly, finance-related applications require high reliability and usability, while most of the psychometric NLP approaches are uninterpretable. This restricts the application of such approach in financial applications. To this end, the interpretability of the risk assessment model is important as it helps the financial institutes to understand how the model makes decisions and to involve manual intervention when necessary.

To solve such problems, we propose a novel credit risk assessment framework by incorporating both demographic and psychological information of borrowers into an interpretable risk assessment model, which is named as PsyCredit in this paper. The proposed framework consists of the following components: data collection, personality detector for online borrowers (PDOB), and explainable risk assessment model (ERAM). PDOB is designed guided by the five-factor model (FFM) (Goldberg, 1990), which aims to uncover the big five personality traits of each borrower. Specifically, after collecting data of borrowers, PDOB solves the challenge of the originality of text by applying a novel deep learning-based originality recognition algorithm to filter out non-original postings. As noted by the Systemic Functional Linguistic Theory (SFLT), language is interpersonal and it reflects person-specific characteristics (Halliday, 2004). In other words, the language styles of different individuals should vary a lot, for instance, people have different preferences on choosing words, syntax, or grammar to construct a sentence. Guided by SFLT, the proposed originality recognition algorithm detects the originality of microblogs by extracting both semantic and syntactic features facilitated by a popular pre-trained deep learning-based language model, namely, Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2018). A weighting scheme based on the Earth Mover Distance (Levina & Bickel, 2001) is designed to retain personality-relevant features and discard those irrelevant. After filtering out non-original postings, PDOB applies a supervised machine learning approach to get the scores for different personality traits. The output of PDOB would then be fed to an interpretable risk assessment model, implemented using the Layer-wise Relevance Propagation technique (Bach et al., 2015). It predicts credit scores and produce valid interpretations for each feature of individuals, thereby enhancing the understanding of clients about the underlying mechanisms. Compared with existing studies, our EDNN introduces a unified feature set of both fundamental information (e.g., income) and social media information (i.e., personality traits) and presents the importance of each aforementioned feature. Unlike the traditional deep learning models, EDNN can explain why the individual is classified into the respective label.

Guided by the design science methodology (Hevner et al., 2004), the designed artifact should under rigorous evaluation. Accordingly, we design several sets of experiments to test the effectiveness of each component in the proposed credit risk assessment framework. First, we design experiments to verify the performance of the proposed originality recognition algorithm. Second, the overall performance of PDOB is tested using a unique dataset collected from Sina Weibo users along with their personality labels acquired by an online questionnaire. Then, under another dataset provided by a real-world P2P lending company, several experiments are designed to evaluate the performance of ERAM both quantitatively and qualitatively.

Overall, the contribution of this study is fourfold. First, we propose a novel risk assessment framework, namely PsyCredit, which fills the current research gap by introducing the psychological analysis of borrowers through their social media postings. Second, we propose a novel deep learning-based psychometric NLP approach that solves a common problem of the originality of postings existing in extant studies. Third, we improve the usability of the proposed PsyCredit by leveraging an explainable deep neural network so that the clients could gain more understandings about the decisions made by PsyCredit. Last but not least, we successfully verify the usefulness and the effectiveness of the proposed PsyCredit framework under a large scale dataset provided by a real-world online lending company. Specifically, to our best knowledge, we are among the first to verify the predictive power of borrowers' personality traits on their default behaviors.

This study is presented as follows. The following section reviews the related work about credit risk assessment and personality detection. Then, the proposed framework is presented with the methodology and computational details, followed by the experiments and results. Finally, conclusion and future work are discussed.

2. Related work

2.1. Credit risk assessment

Financial institutions leverage internal or external credit scoring systems to estimate their clients' credit risk, i.e., default or delinquency probability (Serrano-Cinca and Gutiérrez-Nieto, 2016). Many efforts have been made to evaluate the credit risk of individuals for a long time (Fu et al., 2020). Diverse models are involved in this procedure, including two categories of methods, i.e., traditional and advanced statistical models (Abdou and Pointon, 2011). Traditional statistical models include naïve Bayes, logistic regression, and so on, while advanced statistical models usually refer to the data mining methods, e.g., decision trees, neural networks, and support vector machine. Further, different factors for modeling are leveraged. Credit risk assessment models usually incorporate the current financial situation and historical credit information. About 50 or 60 factors were under consideration at the beginning of modeling the scoring systems, and however, only 8–12 variables remained for the optimal prediction based on Fair, Issac and Company, Inc., the largest company of credit scoring (Mester, 1997).

In order to enhance scoring ability, a lot of effort has been devoted to the improvement of diverse models with standard credit datasets. Some advanced data mining techniques, e.g., neural networks and support vector machine (SVM), are used (West, 2000, Huang et al., 2007, Pławiak et al., 2019). Moreover, some hybrid techniques are also proposed for better scoring results with higher accuracy. For instance, Zhang et al. (2019) proposed a novel multi-stage hybrid model, which is the stacking-based ensemble strategy, for credit scoring. Among these studies, neural network is usually used in the hybrid techniques. For example, Chuang and Huang (2011) designed a two-stage scoring method, combing neural network, and case-based reasoning techniques to reduce Type I error from the first stage, i.e., neural networks. At the meantime, SVM is also used to establish a hybrid model. For example, Zhou et al. (2009) leveraged diverse weighted support vector machine and area under the receiver operating characteristics curve (AUC) maximization for scoring.

Recent years have witnessed the prosperity of online finance, such as online banking, and peer-to-peer lending. Many transactions have been transferred from offline to online. By contract, financial institutions cannot communicate with clients face-to-face. This limitation enhances the importance of credit risk assessment. Further, institutions cannot collect all the relevant information about clients as they do offline. Accordingly, credit risk assessment methods are exploring new measures for better estimation. For instance, in the P2P lending area, one form of online finance, clients are evaluated according to the traditional credit score and other factors, such as the previous activities or social network of the clients (Greiner and Wang, 2009; Shen et al., 2010; Lee and Lee, 2012; Lin et al., 2013; Gonzalez and Loureiro, 2014; Ge et al., 2017). For example, Gonzalez and Loureiro (2014) discussed how the photo of borrowers impacts lenders' decision making through the loan success evaluation and the photo could increase their credit to some extent. In addition, Ge et al. (2017) discussed the impact of social media account disclosure on clients' creditworthiness. Moreover, some machine learning-based online credit assessment approach are proposed. For instance, He et al. (2018) proposed an ensemble method for credit scoring, while Xia et al. (2020) further incorporated a tree-based dynamic heterogeneous ensemble method for credit scoring.

Despite that new features from social media have been extracted for scoring, no study investigates the rich information of the content. Although several universal credit scoring systems like Zhima scores had been widely applied, we still have limited understandings about which types of information embedded in users' social media activities are predictive on their credit status. Our research aims to propose novel features, i.e., personality traits, from social media postings. Though some existing research has claimed the relationships between personality traits and credit scores (Davey and George, 2011, Bernerth et al., 2012), to our best knowledge, no efforts have been devoted to enhancing the scoring with personality traits from readily available social media postings. Indeed, the personality of individuals is closely related to their risk-taking behaviors (Nicholson et al., 2005). Specifically, it is verified that ones' risk propensity is negatively correlated with neuroticism, agreeableness, and conscientiousness and positively correlated with extraversion and openness (Hrazdil et al., 2020). It is plausible that borrowers with high-risk propensity may incur more risk-taking behaviors, which hinders their repaid ability. For instance, they may conduct risky activities like gambling, which puts their personal property in danger. A good credit assessment model should take borrowers' psychological propensity into consideration and produces a comprehensive credit score. Furthermore, machine learning methods have widely applied in the credit risk field and achieved acceptable performance, and however, the lack of explainability has limited the promotion of these methods. Our proposed framework can fully utilize social media data and explore the predictive power of the data from the perspective of personality traits.

2.2. Personality analysis

Personality analysis with text data has been studied for a long time. Machine learning techniques have been widely leveraged to analyze personality traits (Golbeck et al., 2011, Iacobelli et al., 2011). Personality refers to the fundamental characteristics of individuals including their behavior, cognition, and emotional patterns (Corr and Matthews, 2009). In the field of psychology, Big Five model, also known as the five-factor model (FFM), is widely applied to evaluate the personality traits of individuals with respect to five pre-defined dimensions, namely, extraversion (EXT for short), neuroticism (NEU for short), agreeableness (AGR for short), conscientiousness (CON for short) and openness (OPN for short) (Goldberg, 1990).

There are several existing methods to measure the big five personality traits, e.g., International Personality Item Pool (IPIP), NEO-PI-R (Costa and McCrae, 2008), Self-descriptive sentence questionnaires (Fruyt et al., 2004), Self-report questionnaires (Donaldson and Grant-Vallone, 2002), etc. In general, these methods utilize pre-defined personality inventories or questionnaires to measure ones' disposition. However, such types of methodology are relatedly time-consuming and costly, since sometimes, it is hard to get a prompt response from the respondents. To alleviate such a problem, some researchers turn to explore the psychological value of ones' second-hand information. Researchers in the field of both psychology and linguistic disciplines have verified that individuals' writings or spoken dialogues reflect their personalities (Mehl et al., 2006). Specifically, some linguistic cues, e.g., syntax or grammar styles, emotional polarity, or even the use of punctuations, are proved to be useful to evaluate ones' personality traits (Mairesse et al., 2007). These findings prompt the development of automatic personality detection approaches. On the one hand, some psychology lexicons like LIWC (Linguistic Inquiry and Word Count) (Pennebaker et al., 2001) and MRC (Research Council Psycholinguistics Database) have been widely leveraged in the automatic personality detection tasks (Farnadi et al., 2013). On the other hand, the bag-of-word model is also commonly applied to vectorize the unstructured textual data, and accordingly, to capture the linguistic patterns for personality detection (Farnadi et al., 2013).

With the rapid rise of user-generated content (UGC) in social media, people compose social media postings or blogs everyday. Since these self-disclosed data carry some linguistic cues that identify their own personality traits, it is feasible to uncover ones' disposition through this novel data source. Tausczik and Pennebaker (2010) summarize the usage of the lexicon-based method, namely LIWC, to address the automatic personality detection tasks. However, one of the most prominent drawbacks for the lexicon-based methods is that it is difficult for them to analyze social media data where the language environment changes dynamically every day. Specifically, lexicons like LIWC only include some pre-defined words, and they lack the ability to identify some new words that emerged in social media.

To address this problem, some recent studies seek ways to apply machine learning methods to detect individuals' personalities through their social media postings or blogs. For instance, Pratama and Sarno (2015) reported the usefulness of machine learning methods like Support-Vector Machines (SVM), K-nearest neighbors algorithm (KNN), and Naive Bayes (NB). However, most of the existing methods leverage the bag-of-word model to transform unstructured language into fix-length vectors, which neglects the sequential information embedded in documents. To alleviate such a problem, Wright and Chin (2014) utilized an N-gram model to learn the short-term dependency between words. Recently, with the rapid development of deep learning, it has been widely applied in social media analysis (Abdi et al., 2019, Kumar et al., 2020, Beskow et al., 2020). As for personality detection,which is a hot topic within the field of affective computing, deep learning has also been proven to be effective (Mehta et al., 2019, Li et al., 2020). For instance, Yu and Markov (2017) apply deep neural networks like CNN or GRU to capture longer-term dependency information to enhance the performance. Moreover, Li et al. (2020) leveraged a deep learning network to uncover personality from multimodal datasets. However, as an obvious drawback of such deep learning-based approaches, a lot of labeled samples are needed in the training process, while the samples with personality labels are costly to collect. How to make trade-offs between using deeper networks and, at the same time, using less labeled samples is, to our best knowledge, remained to be an unsolved problem in the existing automatic personality detection literature. Additionally, with so much attempt for automatic personality detection in English context, we surprisingly find that there is very limited literature dealing with this problem in the Chinese context.

3. Deep learning framework for credit risk assessment

3.1. Overview of PsyCredit

The study proposes an interpretable deep learning-based credit assessment approach, i.e., PsyCredit. Fig. 1 shows the deep learning framework for credit risk assessment. The main modules consist of data collection, personality mining, assessment model, and output visualization.

Two categories of information about clients are collected: basic and social media. Social media information refers to the postings published by corresponding clients. Further, these postings are processed through deep learning techniques to mine the personality traits of these clients. In this module, an external source of information is used to train the personality mining model.

These extracted personality traits together with traditional basic information are input into one explainable assessment module, which achieves good credit risk assessment with the two categories of information and explainability of the assessment results. Hence, the output includes credit risk and corresponding explainable summary about the risk result. Each module is discussed in detail.

3.2. Data collection

The first module is data collection, consisting of basic information of clients provided by one financial institution and the corresponding social media postings. Basic information refers to the traditional basic information used in credit risk assessment. This study includes age, income, employment year, mobile verification, Zhima score[1], loan amount, and a referrer relationship with the applicants, revealing the clients' fundamental conditions. For social media information, the latest 500 postings of each client are collected through the program, and if the number of one client's postings is less than 500, all the valid postings are collected. Through this module, two categories of clients' information are stored.

3.3. Personality detector for online borrowers

3.3.1. Originality detection

In social media platforms like Twitter or Weibo, the authorship of the Tweets are unclear, i.e., we are not sure whether the Tweets were composed by the users themselves. For instance, users may copy and paste postings from others. This would bring a lot of difficulty for the personality detection tasks via users' social media postings since the core assumption of the existing machine learning-based psychometric NLP approaches is that the text should come from the authors themselves (Mairesse et al., 2007). Based on the Systemic Functional Linguistic Theory (SFLT), language can reveal person-specific characteristics (Halliday, 2004). Specifically, language styles vary a lot across individuals, and thereby, individuals' usage of preference words, syntax and grammar styles can be seen as the fingerprint of themselves that includes rich identity information.

Guided by the SFLT, in this paper, we propose a novel unsupervised learning approach to identify postings originally composed by the users themselves. Specifically, a popular pretrained deep learning-based language model, namely, BERT (Devlin et al., 2018), is used for extracting features revealing individuals' language styles. Although BERT is a deep learning model with low interpretability, recent studies had confirmed that syntactic or grammar-related features would be captured in middle or lower layers, while higher layers encode semantic features. As stated by the SFLT, the syntactic preference of individuals, which is formed by their long-term experience, would serve as their personal identification. Accordingly, we apply the outputs of the intermediate layer in BERT as the textual embeddings representing each posting.

After the vectorization of the social media postings, an unsupervised approach based on the K-Means algorithm is designed to leverage such linguistic cues for authorship detection. The overall algorithm is shown in Algorithm 1. Since the intermedia layers' outputs of BERT are high-dimensional and un-interpretable, some dimensions may be useless for authorship detection and generate noises in the traditional K-Means clustering approach. Since it is apparent that features with similar distributions under different clusters would have less ability to discriminate different clusters, those with similar distributions across clusters should be assigned lower weight during the iterations. Thus, intuitively, we intend to measure the distance between the distributions of the certain feature under multiple clusters and assign the weights to each feature when calculating the center points. Instead of choosing the Jensen-Shannon (JS) divergence as the distance measurement approach, we apply the Earth Mover Distance (Levina & Bickel, 2001), also known as Wasserstein distance, in the proposed approach. As noted by Arjovsky et al. (2017), JS divergence cannot provide a usable gradient when distributions are supported on non-overlapping domains, while the Earth Mover Distance can overcome this problem.

Accordingly, Earth Mover Distance (Levina & Bickel, 2001), which measures the distance between two probability distributions, is applied to calculate the weight for each feature. This distance can be regarded as the minimum amount of work required to transform u into v. The Earth Mover Distance is formally defined as follows:.

$$W(u,\ v) = \inf_{\pi\in\Gamma(u,v)} \int_{\mathbb{R}\times\mathbb{R}} |x-y| d\pi(x,y) \tag{1}$$

Where u and v are two probability distribution and $\Gamma(u,v)$ is the set of probability distributions whose marginals are u and v on the first and second factors respectively. For two probability distributions p and q with their sample sets $\boldsymbol{S_p}$ and $\boldsymbol{S_q}$ respectively, the Earth Mover Distance w.r.t m^{th} feature on the given embeddings can be approximated as follows:

$$\boldsymbol{w_m}(\boldsymbol{S_p}, \boldsymbol{S_q}) = \frac{\sum_{x^{(i)}\in \boldsymbol{S_p}} \sum_{x^{(j)}\in \boldsymbol{S_q}} |x_m^{(i)} - x_m^{(j)}|}{\sum_{x^{(i)}\in \boldsymbol{S_p}} \sum_{x^{(j)}\in \boldsymbol{S_q}} 1} \tag{2}$$

Where $x_m^{(i)}$ is value of m^{th} dimension of$x^{(i)}$, which is the i^{th} samples in$\boldsymbol{S_p}$, and $x_m^{(j)}$ is value of m^{th} dimension of$x^{(j)}$, which is the j^{th} samples in$\boldsymbol{S_q}$.

Algorithm 1 reports the overall process of the proposed K-Means-based approach. We assume that users compose more postings by themselves than by copy and paste. Then, we separate the postings into several clusters using an improved K-Means approach. Those within the largest cluster would be retained and regarded as the original postings of the users. As shown in Algorithm 1, K is the number of clusters prescribed in advanced. We first randomly initialize K clusters. From Lines 3–4, Equation (2) is applied to calculate the distance between each clusters for each dimension of the BERT outputs, and d_m represents the overall inter-clusters distance. In Lines 6–9, the weight of m^{th} dimension is applied when reassigning each sample to new clusters. Lines 10–11 is identical to the traditional K-Means, which aims to recalculate the

[1] https://help.aliyun.com/knowledge_detail/36003.html?spm=5176.22414175.sslink.1.17db4f5cIt60hO.

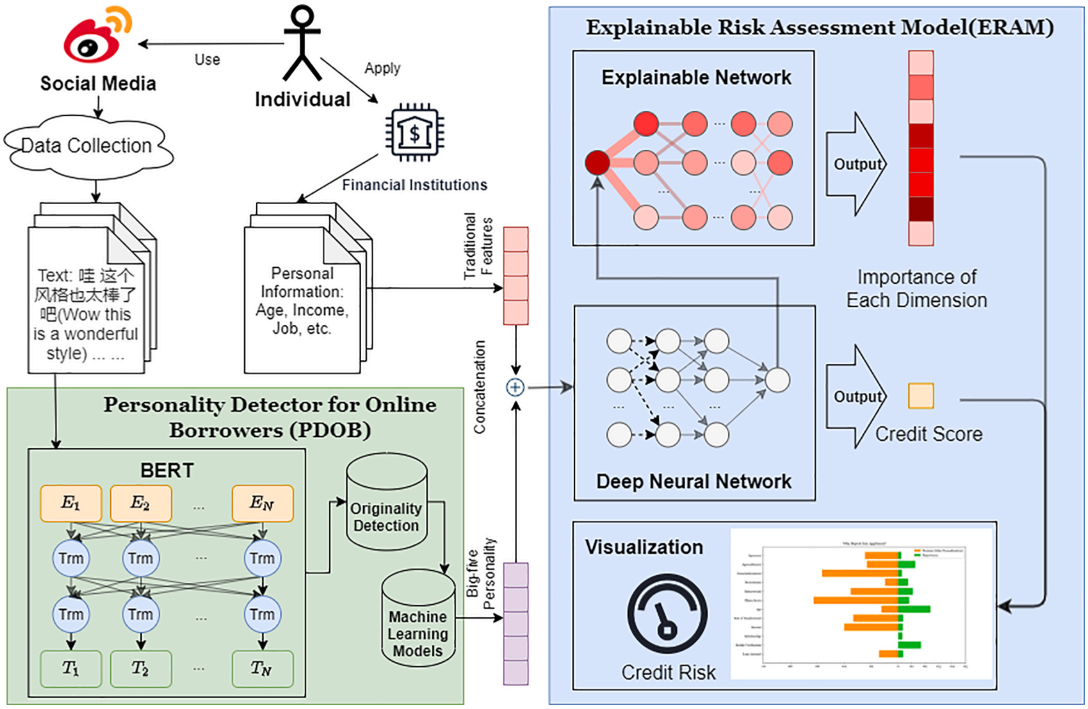


Fig. 1. PsyCredit: A Framework of credit risk assessment.

cluster centroids.

Algorithm 1 An Improved K-Means for *Authorship Detection*

```
Select K cluster centroids$\mu_1,\mu_2, \dots, \mu_K \in \mathbb{R}^n$ from K initial clusters $S_1,S_2, \dots,S_K$.
WHILE$S_1,S_2, \dots, S_K$ have not converged DO
    FOR $m \in [1, M]$ DO
        $d_m \leftarrow \sum_{p\in[1,K]} \sum_{q\in[1,K]} w_m(S_p, S_q)$
    END FOR
    FOR each $x^{(i)} \in X$ DO
        $c_i \leftarrow \underset{j}{\mathrm{argmin}} \sum_{m\in[1,M]} d_m (x^{(i)} - \mu_j)^2, j \in [1, K]$
        Assign $x^{(i)}$ to cluster $S_{c_i}$
    END FOR
    FOR $j \in [1, K]$ DO
        Update$\mu_j \leftarrow \frac{\sum_{i\in[1,N]} D(i,j) x^{(i)}}{\sum_{i\in[1,N]} D(i,j)}$, where $D(i,j) = \begin{cases} 1, c_i = j \\ 0, c_i \neq j \end{cases}$
    END FOR
END WHILE
```

3.3.2. *Personality mining*

The personality mining module is evoked to mine clients' personality traits from their social media postings. Based on the Big-Five model, the designed approach can detect the following five personality traits: neuroticism, conscientiousness, extraversion, agreeableness, and openness. (1) Neuroticism is a tendency to experience negative emotions. Individuals who score high neuroticism cannot handle their life well and their impulsive life is linked to high credit risk; (2) Conscientiousness is a tendency to show self-discipline. High-conscientious individuals realize their responsibilities and they are disciplined in fulfilling them. Applying to financial activities, these individuals can demonstrate self-control in paying their loan; (3) Extraversion is defined as the state of enjoyment external to oneself (e.g., social activities). Individuals with high extraversion engage in social-based consumption, while they are also involved in high-risk financial activities (e.g., impulse buying). Both result in their high credit risk; (4) Agreeableness refers to the individual differences in social harmony. As the definition suggests, individuals with high agreeableness follow the repayment rule, while those with low agreeableness may break promises (i.e., paying the loan on time); (5) Openness indicates a preference for a variety of experience. Few studies have examined the impact of openness on financial behavior. Individuals with high openness are involved in new ideas and approaches, resulting in irrational financial decisions.

Specifically, a popular pretrained deep learning-based language model, namely, BERT, is used for personality detection. One critical problem in this module is the training of the personality mining model. Existing studies have used the survey method to evaluate individual's personality traits. However, responding to a series of questions is time-consuming. Though recent studies have investigated how to mine personality traits from postings, little attention has been paid to the Chinese language. In this module, a survey is conducted to mine the personality traits and investigate their postings in social media. Then, the personality results and the postings are used to train the personality mining model.

Given that the collection of the labeled samples is expensive, the complexity of the designed model is restricted by the size of the dataset. To address this problem and comprehensively use the advanced deep learning approach to improve the performance, this study uses BERT, a pretrained deep learning approach. With the development of deep learning, a recent trend in the field of natural language processing (NLP) is to use a pretrained language model to encode the information embedded in language, for example, ELMo (Peters et al., 2018), ULMFiT (Howard and Ruder, 2018), and BERT (Devlin et al., 2018). Inspired by the idea of transfer learning (Torrey and Shavlik, 2010), these language models are highly complex and pretrained through a large domain-general corpus. BERT is one of the most advanced language models and has been proven to be useful in different downstream NLP tasks. This study is among the first to introduce the pretrained language model in the field of automatic personality detection. This study uses the Pytorch[2] version of BERT[3] released by Google AI, as well as the pretrained model[4] for the Chinese language. On initialization by the pretrained parameters, BERT can represent a document by a fix-length document embedding.

Fig. 1 shows that after vectorization by BERT, different machine

[2] https://pytorch.org/.

[3] https://github.com/codertimo/BERT-pytorch.

[4] https://github.com/google-research/bert.

learning methods can be applied to the downstream task (i.e., the personality detection task). Based on earlier studies (Pratama and Sarno, 2015), we model it as a classification problem. Specifically, as each trait in Big Five has two poles, the task of detecting each trait is modeled as an independent binary classification task. Therefore, different classification methods could be leveraged. We do not specify a classifier here because the performance of different classification methods is compared under our experiments.

3.4. Explainable risk assessment module

As Fig. 1 shows, the explainable risk assessment module (ERAM) leverages deep neural networks, followed by one explainable component. With regard to deep neural network, it contains several fully connected layers and each fully connected layer is followed by a dropout layer. Given that we model risk assessment as a classification problem, the cross entropy loss is applied as the loss function (Murphy and Kevin, 2012):.

$$L(\theta) = -[l\bullet\log(\mathrm{p}(\mathrm{y}=1|\theta)) + (1-\mathrm{l})\bullet\log(1-\mathrm{p}(\mathrm{y}=1|\theta))]$$

where θ is the parameter in the neural network, l is the actual label, and $\mathrm{p}(\mathrm{y}=1\backslash\theta)$ is the output of the model representing the probability of a given sample to be true. The optimizer here is Stochastic Gradient Descent with momentum (SGD) (Ferguson, 1982). Two parameters are involved in this optimizer, that is, learning rate and the momentum. Specifically, in each training iteration, the parameter θ is updated:

$$v_t = \beta v_{t-1} + \alpha\nabla_\theta L(\theta)$$

$$\theta \leftarrow \theta - v_t$$

where α and β represent the learning rate and the momentum correspondingly. The standard neural network contains one input layer, one output layer, and one hidden layer, and deep neural network is one kind of neural networks with more than one hidden layer with better learning capabilities. However, despite the performance, the deep neural network lacks explainability. Therefore, the clients may not be convinced by the black-box assessment. For example, the client is labeled as a high credit risk label and then the loan application is rejected and a pure deep neural network cannot interpret the assessment result if asked about the reason. As introduced by Bach et al. (2015), Layer-wise Relevance Propagation (LRP) is a method that explains how the deep neural network makes its decisions. Specifically, it identifies important input dimensions through a backward pass in the neural network. Besides, Montavon et al. (2017) find its mathematical foundation in Deep Taylor Decomposition (DTD). The rule of LRP is defined as:

$$R_i^{(l)} = \sum_j \frac{x_i w_{ij}}{\sum_i x_i w_{ij}} R_j^{(l+1)}$$

where $R_i^{(l+1)}$ is the relevance value at layer $(l+1)$ and neuron$j$. LRP performs a proportional decomposition that uses an upper layer relevance value $R_i^{(l+1)}$ to calculate the lower layer relevance value$R_i^{(l)}$. x_i indicates the activation of neuron$i$, and w_{ij} is the learned weight parameters from neuron i to$j$.

Compared with the extant deep learning techniques, our ERAM achieves good assessment performance and feasible explainability of the importance of each characteristic of the clients. In addition to the accurate credit risk assessment, the output module provides explanations through one visualization component.

4. Experiments and results

To verify the usefulness of the design artifact, we design three experiments to test the performance of different components, i.e., originality detection, personality mining, and credit risk assessment, respectively. The following subsection would elaborate on these experiments in detail.

4.1. Experiment 1: Performance of the proposed originality detection approach

To verify that the proposed approach can distinguish the authorship of postings from different social media users, we design an experiment under an open dataset, namely microblogPCU[5], which was collected from Sina Weibo, with over 40,000 Weibo Postings from 432 Weibo Users. For each user, we mix his/her postings with those composed by other users under the proportion of 5:1. We aim to test whether the proposed approach can distinguish the users' original postings from those composed by others. The largest cluster would be selected as the original postings of the users, and the accuracy of the postings (belonging to the users) under this largest cluster is applied to the proxy performance of the proposed approach. As the baseline method, the traditional K-Means is tested under our experiment. Furthermore, apart from the Earth Mover Distance (EM) distance, we test the performance of the proposed approach using another distance measuring metric, namely, Jensen-Shannon (JC) distance. Moreover, since different types of information are captured in different intermedia-layers of BERT, we test the performance of the proposed approach utilizing embeddings constructed from different BERT intermedia-layers. We adopt the pre-trained checkpoint of BERT which is specifically for Chinese language.[6] The output of different layers of BERT is tested so that we could verify which layers are more important for the originality detection task. The experimental results are shown in Fig. 2. We find that the proposed approach using EM distance as a weighting scheme outperforms others. We also find that BERT embeddings from shallow layers perform better than that from deep layers. Specifically, the information collected from the second layer works best. The reason is that the information collected by the shallow layers are more related to syntax or grammar of the language, which can reflect ones' language identity.

To evaluate the proposed method qualitatively, we design a case study. We mix the Weibo Postings from two users, named as A and B respectively, and apply the proposed approach to separate them from two parts. The postings are shown in Table 1 of Appendix A on online supplements. We choose the second layer of BERT to construct the embeddings of each posting. We visualize the distribution of their postings in several selected dimensions in Fig. 3. In Fig. 3 (a), the features are with the highest Earth Mover distance measured by our proposed approach, while Fig. 3 (b) reports features with the lowest Earth Mover distance. We find that the two users (with red or blue color) can be easily separated under the BERT features with the highest Earth Mover distance, but are mixed with lowest Earth Mover distance. This verifies that the proposed approach can successfully select the important features to enhance the performance.

4.2. Experiment 2: Performance of the BERT-based personality detector

We first conduct a survey to acquire the Big Five personalities of Weibo users. To obtain the personality labels of Chinese Weibo users, we use the short version of the Revised NEO Personality Inventory (NEO PI-R) (Costa and McCrae, 2008), namely, NEO-FFI-R, in our experiment. NEO-FFI-R provides a systematic assessment of Big-five personality traits, and consists of 60 items. Since its reliability had been proven by previous studies (Aluja et al., 2005; Aluja et al., 2007), it had been widely applied for psychological testings. Specifically, we following the methodology adopted by Celli et al. (2013) to construct our dataset. We

[5] https://pgram.com/dataset/microblogpcu/.

[6] https://storage.googleapis.com/bert_models/2018_11_03/chinese_L-12_H-768_A-12.zip.

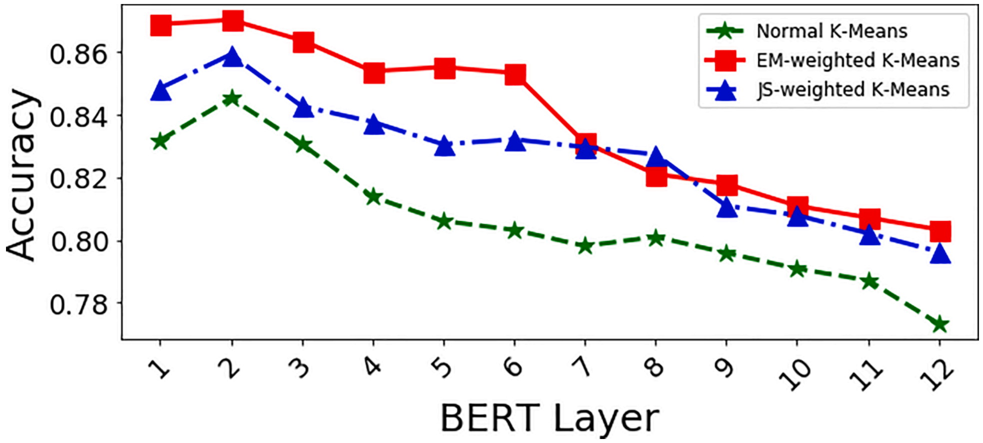


Fig. 2. Performance of the Originality Detection Approaches w.r.t Different BERT Layers as Input.

Table 1
Descriptive Statistic of Personality Traits of Weibo Users.

	EXT	NEU	CON	AGR	OPN	# of Weibo	# of Remained
Count	187	187	187	187	187	187	187
Mean	0.46	0.49	0.48	0.60	0.47	83.3	73.3
Std	0.16	0.17	0.17	0.16	0.17	118.6	108.6
Min	0	0	0	0	0	3	2
Max	1	1	1	1	1	419	417

distributed the personality inventory through social media, especially Weibo, and collected 187 valid responses. Then, based on the authorization given by the subjects, their Weibo postings were collected to construct a dataset. Specifically, 15,571 postings are collected. After the originality detection step, we 13,709 postings are remained. For each of the user, in average, 108.6 origin postings are generated. We standardized the obtained Big Five personality labels into zero and one, and report the data description in Table 1. After filtering out those non-original postings using the originality detection approach, the Tweets then went into several machine learning models to learn the patterns between borrowers' linguistic styles and their personality traits.

Moreover, to evaluate the performance of the proposed framework, four common metrics are used: Precision, Recall, F1, and accuracy. Precision refers to the percentage of individuals classified as positive (i.e., high risk) who are actually positive and Recall is the percentage of actual high-risk individuals classified as positive. F_1 is the harmonic mean of the aforementioned Precision and Recall. Accuracy is the average classification performance.

$$precision = TP/(TP+FP)$$

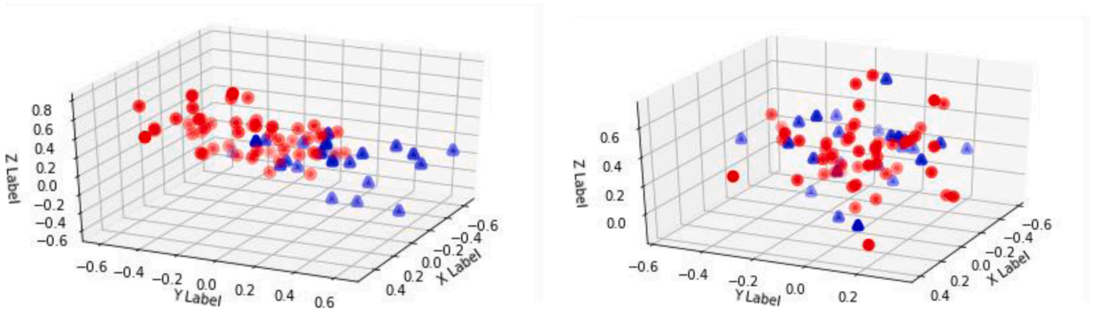
$$recall = TP/(TP+FN)$$

$$\frac{F_1 = 2\bullet precision\bullet recall}{precision + recall}$$

$$\text{Accuracy} = (TP+FN)/(TP+TN+FP+FN)$$

where TP is the true positive counts, TN is the true negative counts, FP is the false positive counts, and FN is the false negative counts.

We apply various popular machine learning methods in the experiments, such as decision tree (Tree), AdaBoost classifier, Linear Discriminant Analysis (LDA), Bagging classifier, and random forest classifier. Besides, we also test the performance of a fully connection network (FC) in the experiment. Moreover, to verity the usefulness of the proposed originality detection approach, we conduct an ablation experiment by excluding the originality recognition process. Table 2 shows the overall experimental results. The performance of each classifier is measured by F1, recall, and precision, which are reported in Figs. 4–6. We find that FC has the highest performance in terms of all metrics, with an average of 0.6671 w.r.t F1 score. Additionally, by comparing the performance with and without the originality recognition, we find that the proposed originality detection approach largely increases the performance of the personality detection process by 13.9%. Given these experimental findings, we selected the fully connection network to train a personality detector for further analysis.



(a) Dimensions with largest EM distance (b) Dimensions with lowest EM distance

Fig. 3. Visualization of Postings from Two Individuals w.r.t Different BERT Dimensions. Notes: the red points represents the postings written by A, while the blue ones are by B; Dimensions X, Y and Z in (a) are those with largest EM distance, while that in (b) are those with lowest EM distance.

Table 2
Overall Comparison between Different Machine Learning Methods.

	FC	FC (no OR)	Tree	Ada	LDA	Bagging	RFR
Accuracy	0.6671	0.5855	0.5674	0.6007	0.6288	0.6012	0.6257
F1	0.6394	0.5547	0.5399	0.5684	0.6009	0.5280	0.5704
Precision	0.6570	0.5639	0.5433	0.5817	0.6113	0.6001	0.6245
Recall	0.6301	0.5465	0.5372	0.5570	0.5918	0.4723	0.5270

Note: This table presents the average scores over Big Five traits w.r.t different evaluation metrics. ‘FC (no OR)’ in the second column means to exclude the originality recognition process before personality detection.

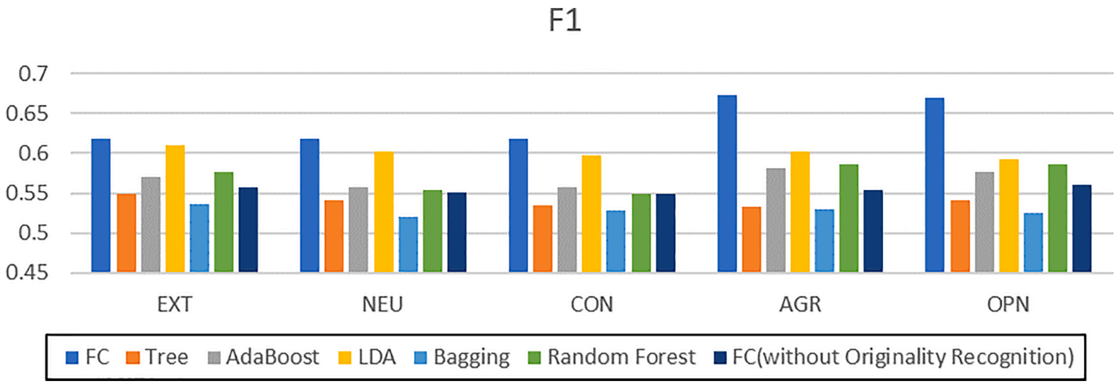


Fig. 4. Performance Comparison w.r.t F1.

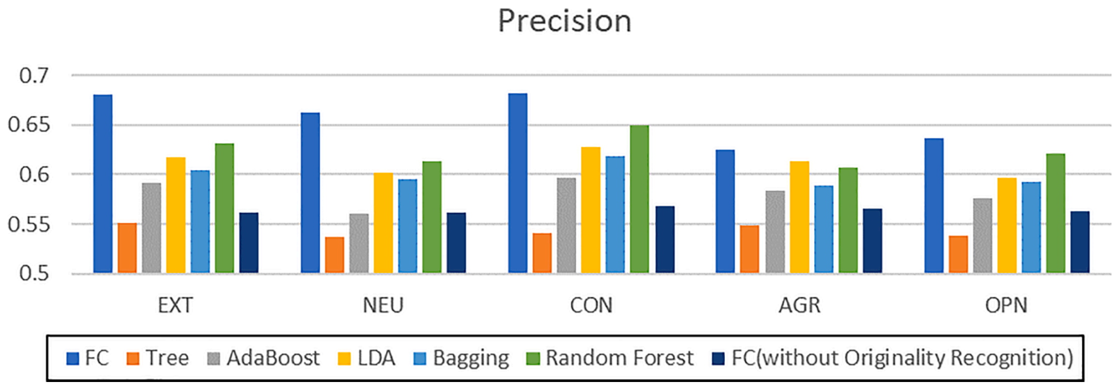


Fig. 5. Performance Comparison w.r.t Precision.

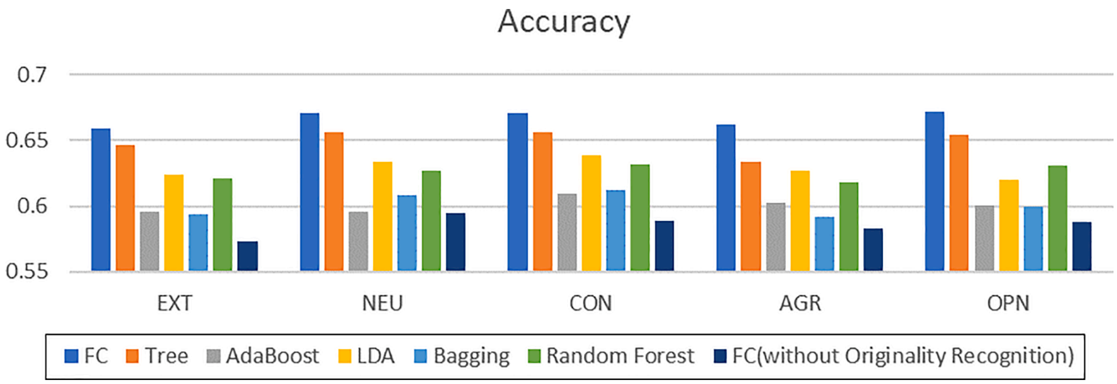


Fig. 6. Performance Comparison w.r.t Accuracy.

4.3. Experiment 3: Evaluation of the automatic risk assessment approach with detected personalities

In this experiment, the deep neural network was employed to model the relationships between our proposed personality measures and individual credit risk. Four common metrics (i.e., Precision, Recall, F1, and Accuracy) are used to evaluate the performance of classification.

If the applicant's loan is at least 30 days overdue, the applicant is labeled as "high credit risk"; otherwise, the applicant is of "low credit risk". As mentioned above, one of the P2P lending companies in China provided the relevant dataset. The original dataset contains 1209 clients, with 1146 low-risk applicants and 63 high-risk applicants, which is reasonable in the real world because only a few applicants would default. Risk is denoted as the label of each client. The value of 1 refers to high risk, whereas the value of 0 means low risk. There are seven fundamental variables and five personality traits for each client:.

- Loan Amount (Loan): The loan amount (in thousand) the applicant applies.
- Age (Age): The age of the applicant.
- Monthly Income (MIncome): Monthly income of the applicant. Categorical variable, 0 if below 3000, 1 if between 3000 and 5000, 2 if between 5000 and 10,000, 3 if between 10,000 and 20,000, and 4 if above 20,0000.
- Years of Employment (YEmployment): Years of current employment. Categorical variable, 0 if less than 1 year, 1 if between 1 and 5 years, 2 if between 6 and 10 years and 3 if more than 10 years.
- Zhima Score (Zhima): Zhima score provide by Ant Financial.
- Mobile Verification (MVerification): Indicator whether the registration telephone number is consistent to the telephone number connected with the provided bank card. 1 if consistent and 0 otherwise.
- Contact Relationship (CRelationship): The relationship of the guarantor with the applicant. Categorical variable. 1 = father, 2 = spouse, 3 = friend, 4 = brother/sister, 5 = other relative, 6 = colleague, and otherwise 0.
- Extraversion (EXT): The value of the extraversion personality.
- Neuroticism (NEU): The value of the neuroticism personalityConscientiousness (CON): The value of the conscientiousness personality.
- Agreeableness (AGR): The value of the agreeableness personality.
- Openness (OPN): The value of the openness personality.

The data description is presented in Table 3.

However, it is an imbalanced dataset. Traditionally, the classifiers label all the applicants as low risk with such kind of imbalanced dataset. The accuracy criterion for training classifiers is not feasible in this condition. Thus, before the dataset is input into the deep learning model, we need to address the imbalanced dataset to reduce the bias between majority and minority classes. This study leverages the over-sampling strategy from the data perspective. Over-sampling refers to supplementing the data in the minority class. This study uses a common over-sampling technique such as Synthetic Minority Over-sampling Technique (SMOTE). This technique is an extension of random over-sampling. It creates the synthetic data points in terms of the current data points in the minority class. To avoid randomness, SMOTE is adopted for five times and then five datasets are generated accordingly. Balanced datasets are achieved through SMOTE. Each new dataset contains 1146 low-risk and 1146 high-risk applicants. Furthermore, to fully exploit the dataset, a 10-fold cross-validation process is used.

Table 3
Descriptive Statistics of Clients' Information.

Variable	Obs	Mean	Std. Dev.	Min	Max
Loan	1209	26.9338	16.8313	1	50
MVerification	1209	0.6286	0.4834	0	1
CRelationship	1209	2.1423	2.1122	0	6
MIncome	1209	2.0066	0.9481	0	4
YEmployment	1209	1.5327	0.7757	0	3
Age	1209	29.1803	5.7308	21	54
Zhima	1209	686.8453	45.5762	350	807
EXT	1209	0.4717	0.2453	0	1
NEU	1209	0.4967	0.2463	0	1
CON	1209	0.3937	0.2384	0	1
AGR	1209	0.4933	0.2460	0	1
OPN	1209	0.3967	0.2405	0	1
Risk	1209	0.5724	0.5910	0	1

The deep neural network contains eight fully connected layers, and each fully connected layer is followed by a dropout layer. The number of iteration is set as 2000. The optimizer here is Stochastic Gradient Descent (SGD). Two parameters are involved in this optimizer: learning rate and momentum. A parameter selection process is conducted before feeding the data into the model. Traditionally, each generated dataset is randomly partitioned into 10 subsamples and every 9 subsamples are considered a training dataset and one as a testing dataset. In this parameter selection procedure, the original training dataset is further divided into a training and validation dataset. The former is used to train the model and the latter is for parameter validation. Thus, the optimal parameters to further verify the effectiveness of the proposed method in the testing dataset are obtained.

In this stage, Fig. 7 illustrates the parameter selection in one dataset. It shows that in this dataset, the model with learning rate as 0.01 and the momentum as 0.9 can achieve the best performance.

Table 4 presents the experimental results through the selected parameters. The accuracy can achieve 87.57% and F1 is 0.8830 in testing datasets. In addition, AUC (Area under the ROC Curve) is used to measure the performance of our classification models. The average AUC score is 0.9365.

We then test the performance of the credit assessment approach without using personality traits. Specifically, the following experiments only uses the basic features commonly used in credit scoring, for example, age and income. The deep neural networks with the same structure and parameters are used here. Table 5 reports the experimental results with only basic features. Compared with the results in Table 4, the results in Table 5 present that the models with personality traits outperform those without personality traits. In other words, personality traits can be effective indicators in predicting the credit risk of applicants. Besides, the average AUC score here is 0.7947, which also reveals the superiority of our proposed personality traits.

Further, we compare the performance when using other common classification models: (1) Random: Each applicant is randomly assigned with a label, that is, high risk or low risk; (2) linear discrimination analysis (LDA); (3) Naïve Bayes (NB); (4) Support Vector Machine (SVM); (5) MLP: Standard Neural Network; and (6) Logit: Logistic Regression.

Table 6 shows the experimental results with diverse classification models. The results reveal that deep learning technique outperforms other common classification models in terms of the four metrics.

4.4. Visualization of the risk assessment results

To intuitively demonstrate the explainability of our proposed credit risk assessment framework, first, a random client is selected with his or her information (including personality traits and basic information). Each feature is attached with respective importance, revealing why the client is classified into "high risk" or "low risk.".

Fig. 8 visualizes the value of each feature (after normalization) and the importance of this client's features. It shows that the feature of the applicant's age is the most important in determining the "high risk." This figure reveals that this applicant is rejected largely because of his/her age and invalid mobile verification. Among the five personality traits, his/her high openness and neuroticism are the most important features determining the high-risk label. To our surprise, Zhima score is considered to be less important by our model, which implies that it could

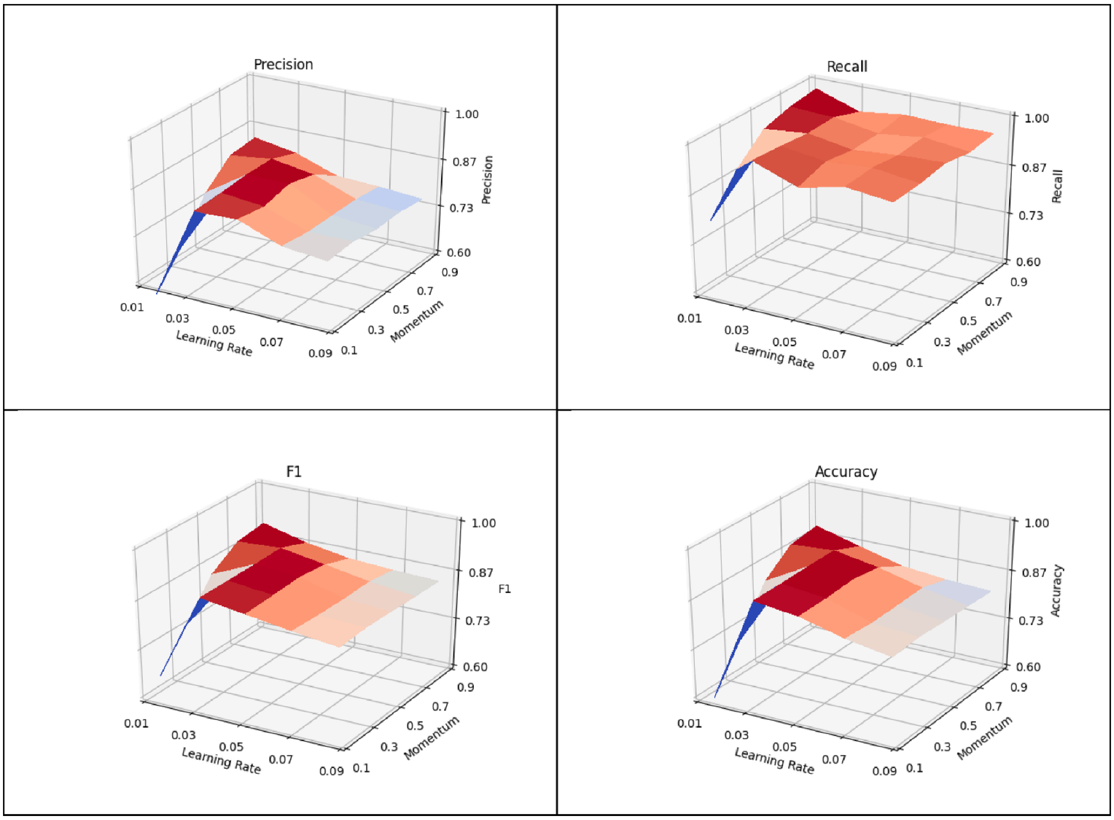


Fig. 7. Parameter Selection in One Dataset.

Table 4
Experimental Results.

Dataset	1	2	3	4	5	Average
Precision	0.8225	0.8572	0.8253	0.8374	0.8199	0.8325
Recall	0.9095	0.9600	0.9674	0.9614	0.9204	0.9437
F1	0.8616	0.9049	0.8901	0.8937	0.8645	0.8830
Accuracy	0.8564	0.8992	0.8809	0.8857	0.8564	0.8757

Table 5
Experimental Results without Personality Traits.

Dataset	1	2	3	4	5	Average
Precision	0.7290	0.7534	0.7183	0.7453	0.7466	0.7385
Recall	0.9084	0.9384	0.9477	0.9280	0.9139	0.9273
F1	0.8041	0.8344	0.8139	0.8223	0.8175	0.8184
Accuracy	0.7805	0.8150	0.7827	0.8019	0.7979	0.7956

be further improved if more social media-related features are added.

To further report the effectiveness of the explanation module, we first rank all the features according to the importance of each feature in one dataset. Thus, we compare the performance of experiments by removing ranked features gradually. The performance is seen to decrease with the exclusion of important features. Meantime, the least important features are removed one by one. Though the performance declines gradually, the changes are smaller compared to the changes after removing the most important features, which indicate the generated importance resorts in this explainable module. Further, random features are gradually removed, and Fig. 9 also illustrates that the experiments of random removal generally outperform those of removing the most important features.

Table 6
Experimental Results with Diverse Models.

Dataset	Random	LDA	NB	SVM	MLP	Logit
Precision	0.5191	0.6061	0.6330	0.6042	0.8177	0.6059
Recall	0.4468	0.5574	0.6589	0.5575	0.8693	0.5756
F1	0.4257	0.5646	0.6419	0.5638	0.8407	0.5757
Accuracy	0.5082	0.5935	0.6379	0.5921	0.8357	0.5976

5. Conclusion and future work

This study proposes an explainable credit risk assessment method based on deep learning techniques, that is, PsyCredit, to evaluate the clients' risk conditions. A personality mining model is built with the deep text mining model, that is, PDOB. Besides, one integrated credit risk assessment framework with explainability, that is, ERAM, is proposed based on the personality mining model. Experiments conducted on a real-world dataset have revealed that personality traits mined from social media postings are conducive to credit risk assessment. The results demonstrated the superiority of our proposed framework in the performance of evaluating clients' credit risk. Moreover, compared with the existing studies on improving accuracy, our novel framework introduces the LRP technique to attach each feature with their respective contribution of the final assessment result. This study provides several theoretical and managerial implications.

5.1. *Theoretical implications*

This study contributes to the literature on information systems and finance. There has been rapid advancement in electronic finance, and credit risk assessment is critical to the finance area. Online information for credit risk assessment has attracted much academic attention in both finance and information systems. Though earlier studies have claimed

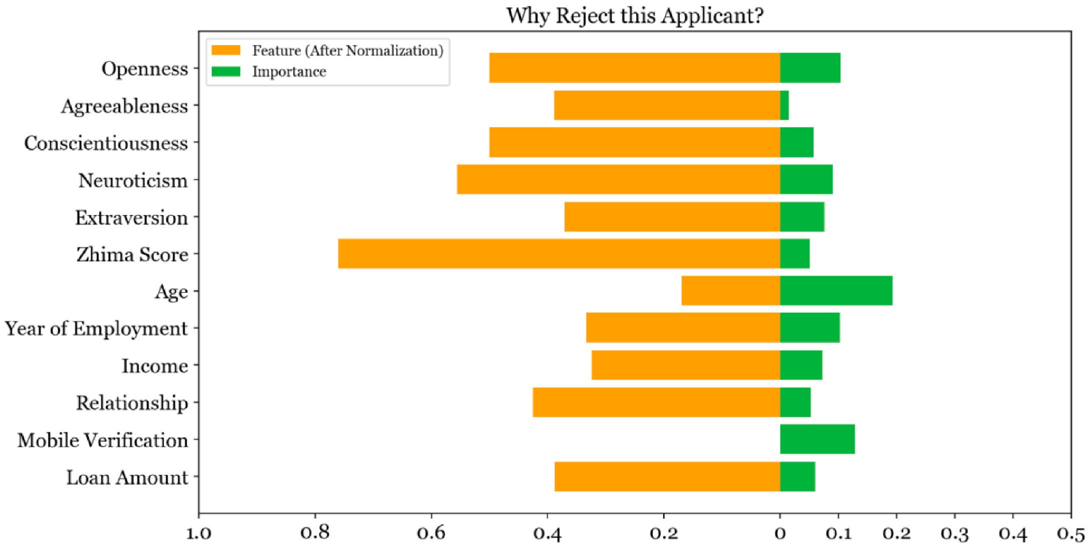


Fig. 8. Importance Visualization. *Notes: 'Feature' shows the values of each feature; 'Importance' shows how important the corresponding features are when the ERAM makes decision.*

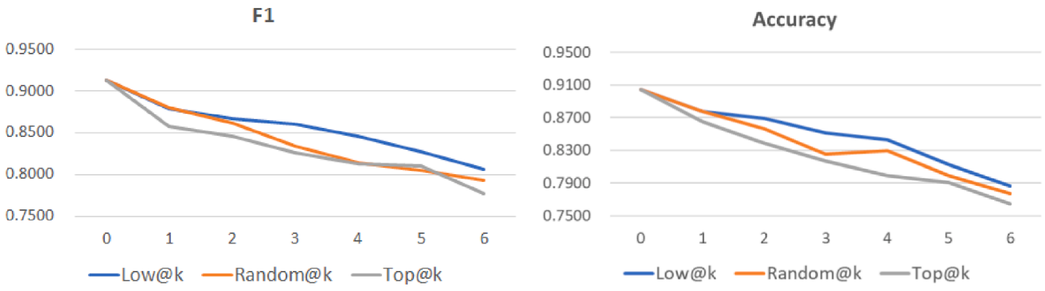


Fig. 9. Performance after removing features.

the validity of personality traits in risk evaluation, no study has explored the personality traits in social media via machine learning and applied the extracted traits in the risk assessment process.

Firstly, grounded on Systemic Functional Linguistic Theory, we propose a novel psychometric approach to uncover individuals' personalities through their social media postings. We contribute to the field of psycholinguistic by solving the problem caused by the originality of social media postings, which is ignored by previous studies. Further, this approach is based on Chinese context. Studies so far have only focused on English-based models to explore the personality traits through text, and omitted the source of text in one's personal page. This is the first study to investigate personality traits with the Chinese language through the text and the originality of the text.

We also propose an integrated framework based on the personality mining model, i.e., PsyCredit. This study contributes toward explainability of the framework. Traditional deep learning models, for example, deep neural networks, have been proved to be effective in many areas. However, they lack transparency. The output cannot present the underlying mechanism. We apply LRP after a deep learning structure to pass back the output and thus generate the contribution of each feature about the client. This is the first study to include personalities into explainable deep learning.

5.2. Managerial implications

This study also provides implications for the credit market. Good credit risk assessment models can avoid losses for financial institutions and promote a favorable financial market.

The personality mining module (PDOB) in the proposed framework can extract the personality traits of clients by incorporating the factor of originality of social media postings. Specifically, the module focuses on the Chinese language to process a large volume of social postings. Therefore, financial institutions can understand their clients from the perspective of personality traits within a short time and less cost (e.g., surveys).

Furthermore, compared with the traditional deep neural network, ERAM incorporates the LRP technique to pass back the output to generate the contribution of each feature about the clients. The framework can generate explanations about the underlying mechanism of the assessment results, solving the black-box problem of traditional deep learning techniques. Hence, financial institutions can apply it for better assessment performance and provide explanations about the assessment results. Through assessment with high accuracy, the framework helps save time and cost and facilitates the respective decisions based on the assessments (e.g., loan applications).

5.3. Future directions

Further research can focus on performance improvement. First, this study used the BERT; thus, future studies should extend the model for better evaluating the personality traits of applicants. Second, currently, only 187 valid questionnaires were collected and used for training and testing the personality mining model. Future studies should increase the number of questionnaires and the sample.

CRediT authorship contribution statement

Kai Yang: Data curation, Writing – original draft. **Hui Yuan:**

Conceptualization, Methodology, Software. **Raymond Y.K. Lau:** Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Yuan’s work was supported by the National Natural Science Foundation of China (Grant No. 72001144), and Innovative Research Team of Shanghai International Studies University (No. 2020114044). Besides, Lau’s work was supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project: CityU 11507219), and City University of Hong Kong SRG (Project: 7005780).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eswa.2022.116847.

References

Abdi, A., Shamsuddin, S. M., Hasan, S., & Piran, J. (2019). Deep learning-based sentiment classification of evaluative text based on Multi-feature fusion. *Information Processing & Management, 56*(4), 1245–1259.

Abdou, H. A., & Pointon, J. (2011). Credit scoring, statistical techniques and evaluation criteria: A review of the literature. *Intelligent Systems in Accounting, Finance and Management, 18*(2–3), 59–88.

Ahmad, F., Abbasi, A., Li, J., Dobolyi, D. G., Netemeyer, R. G., Clifford, G. D., & Chen, H. (2020). A Deep Learning Architecture for Psychometric Natural Language Processing. *ACM Transactions on Information Systems (TOIS), 38*(1), 1–29.

Aluja, A., Garcıa, O., Rossier, J., & Garcıa, L. F. (2005). Comparison of the NEO-FFI, the NEO-FFI-R and an alternative short version of the NEO-PI-R (NEO-60) in Swiss and Spanish samples. *Personality and Individual Differences, 38*(3), 591–604.

Aluja, A., García, L. F., Cuevas, L., & García, O. (2007). The MCMI-III personality disorders scores predicted by the NEO-FFI-R and the ZKPQ-50-CC: A comparative study. *Journal of Personality Disorders, 21*(1), 58–71.

Aral, S., Dellarocas, C., & Godes, D. (2013). Introduction to the special issue—social media and business transformation: A framework for research. *Information Systems Research, 24*(1), 3–13.

Arjovsky, M., Chintala, S. & Bottou, L. (2017). Wasserstein generative adversarial networks. *In Proceedings of the 34th International Conference on Machine Learning*, 70, 214–223.

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K. R., & Samek, W. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS ONE, 10*(7), Article e0130140.

Bernerth, J. B., Taylor, S. G., Walker, H. J., & Whitman, D. S. (2012). An empirical investigation of dispositional antecedents and performance-related outcomes of credit scores. *Journal of Applied Psychology, 97*(2), 469.

Beskow, D. M., Kumar, S., & Carley, K. M. (2020). The evolution of political memes: Detecting and characterizing internet memes with multi-modal deep learning. *Information Processing & Management, 57*(2), Article 102170.

Burtch, G., Ghose, A., & Wattal, S. (2014). Cultural differences and geography as determinants of online prosocial lending. *Mis Quarterly, 38*(3), 773–794.

Celli, F., Pianesi, F., Stillwell, D., & Kosinski, M. (2013, June). Workshop on computational personality recognition: Shared task. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 7, No. 1).

Chuang, C. L., & Huang, S. T. (2011). A hybrid neural network approach for credit scoring. *Expert Systems, 28*(2), 185–196.

Corr, P. J., & Matthews, G. (Eds.). (2009). *The Cambridge handbook of personality psychology* (pp. 748–763). Cambridge: Cambridge University Press.

Costa, P. T., Jr, & McCrae, R. R. (2008). *The revised NEO personality inventory (NEO-PI-R)*. Sage Publications Inc.

Davey, J., & George, C. (2011). Personality and finance: The effects of personality on financial attitudes and behaviour. *International Journal of Interdisciplinary Social Sciences, 5*(9), 275–294.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint* arXiv: 1810.04805.

Donaldson, S. I., & Grant-Vallone, E. J. (2002). Understanding self-report bias in organizational behavior research. *Journal of Business and Psychology, 17*(2), 245–260.

Engelberg, J. E., & Parsons, C. A. (2011). The causal impact of media in financial markets. *The Journal of Finance, 66*(1), 67–97.

Farnadi, G., Zoghbi, S., Moens, M. F., & De Cock, M. (2013). Recognising personality traits using facebook status updates. In *Proceedings of the Workshop on Computational Personality Recognition (WCPR13) at the 7th International AAAI Conference on Weblogs and Social Media (ICWSM13)* (pp. 14–18). AAAI.

Ferguson, T. S. (1982). An inconsistent maximum likelihood estimate. *Journal of the American Statistical Association, 77*(380), 831–834.

Fruyt, F. D., McCrae, R. R., Szirmák, Z., & Nagy, J. (2004). The five-factor personality inventory as a measure of the five-factor model: Belgian, American, and Hungarian comparisons with the NEO-PI-R. *Assessment, 11*(3), 207–215.

Fu, X., Ouyang, T., Chen, J., & Luo, X. (2020). Listening to the investors: A novel framework for online lending default prediction using deep learning neural networks. *Information Processing & Management, 57*(4), Article 102236.

Ge, R., Feng, J., Gu, B., & Zhang, P. (2017). Predicting and deterring default with social media information in peer-to-peer lending. *Journal of Management Information Systems, 34*(2), 401–424.

Goh, K. Y., Heng, C. S., & Lin, Z. (2013). Social media brand community and consumer behavior: Quantifying the relative impact of user-and marketer-generated content. *Information Systems Research, 24*(1), 88–107.

Golbeck, J., Robles, C., & Turner, K. (2011). Predicting personality with social media. *In CHI'11 Extended Abstracts on Human Factors in Computing Systems* (pp. 253–262).

Goldberg, L. R. (1990). An alternative“ description of personality”: The big-five factor structure. *Journal of Personality and Social Psychology, 59*(6), 1216.

Gonzalez, L., & Loureiro, Y. K. (2014). When can a photo increase credit? The impact of lender and borrower profiles on online peer-to-peer loans. *Journal of Behavioral and Experimental Finance, 2*, 44–58.

Greiner, M. E., & Wang, H. (2009). The role of social capital in people-to-people lending marketplaces. *ICIS 2009 proceedings*, 29.

Guo, G., Zhu, F., Chen, E., Liu, Q., Wu, L., & Guan, C. (2016). From footprint to evidence: An exploratory study of mining social data for credit scoring. *ACM Transactions on the Web (TWEB), 10*(4), 1–38.

Halliday, M. A. K., & Hasan, R. (2004). An introduction to functional grammar. (3rd ed., revised by C. Matthiessen). *Londres*: Edward Arnold.

He, H., Zhang, W., & Zhang, S. (2018). A novel ensemble method for credit scoring: Adaption of different imbalance ratios. *Expert Systems with Applications, 98*, 105–117.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 75–105.

Howard, J., & Ruder, S. (2018). Fine-tuned language models for text classification. *arXiv preprint* arXiv:1801.06146, 194.

Hrazdil, K., Novak, J., Rogo, R., Wiedman, C., & Zhang, R. (2020). Measuring executive personality using machine-learning algorithms: A new approach and audit fee-based validation tests. *Journal of Business Finance & Accounting, 47*(3–4), 519–544.

Huang, C. L., Chen, M. C., & Wang, C. J. (2007). Credit scoring with a data mining approach based on support vector machines. *Expert Systems with Applications, 33*(4), 847–856.

Iacobelli, F., Gill, A. J., Nowson, S., & Oberlander, J. (2011). Large scale personality classification of bloggers. In *International conference on affective computing and intelligent interaction* (pp. 568–577). Berlin, Heidelberg: Springer.

Kumar, A., Srinivasan, K., Cheng, W. H., & Zomaya, A. Y. (2020). Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data. *Information Processing & Management, 57*(1), Article 102141.

Lee, E., & Lee, B. (2012). Herding behavior in online P2P lending: An empirical investigation. *Electronic Commerce Research and Applications, 11*(5), 495–503.

Levina, E. & Bickel, P. (2001). The earth mover's distance is the mallows distance: Some insights from statistics. *In Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, 2, 251-256.

Li, Y., Wan, J., Miao, Q., Escalera, S., Fang, H., Chen, H., … Guo, G. (2020). CR-Net: A deep classification-regression network for multimodal apparent personality analysis. *International Journal of Computer Vision*, 1–18.

Liang, K., & He, J. (2020). Analyzing credit risk among Chinese P2P-lending businesses by integrating text-related soft information. *Electronic Commerce Research and Applications, 40*, Article 100947.

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

Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research, 30*, 457–500.

Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology, 90*(5), 862.

Mehta, Y., Majumder, N., Gelbukh, A., & Cambria, E. (2019). Recent trends in deep learning based personality detection. *Artificial Intelligence Review*, 1–27.

Mester, L. J. (1997). What's the point of credit scoring? *Business Review, 3*(Sep/Oct), 3–16.

Montavon, G., Lapuschkin, S., Binder, A., Samek, W., & Müller, K. R. (2017). Explaining nonlinear classification decisions with deep taylor decomposition. *Pattern Recognition, 65*, 211–222.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology, 77*(6), 1296.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001), 2001.

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. In *Proceedings of NAACL-HLT* (pp. 2227–2237).

Pławiak, P., Abdar, M., & Acharya, U. R. (2019). Application of new deep genetic cascade ensemble of SVM classifiers to predict the Australian credit scoring. *Applied Soft Computing, 84*, Article 105740.

Pratama, B. Y., & Sarno, R. (2015). Personality classification based on Twitter text using Naive Bayes, KNN and SVM. In *2015 International Conference on Data and Software Engineering (ICoDSE)* (pp. 170–174). IEEE.

Serrano-Cinca, C., & Gutiérrez-Nieto, B. (2016). The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending. *Decision Support Systems, 89*, 113–122.

Shen, D., Krumme, C., & Lippman, A. (2010). Follow the profit or the herd? Exploring social effects in peer-to-peer lending. In *2010 IEEE Second International Conference on Social Computing* (pp. 137–144). IEEE.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*(1), 24–54.

Torrey, L., & Shavlik, J. (2010). Transfer learning. *In Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques* (pp. 242–264). IGI global.

Traxler, M., & Gernsbacher, M. A. (Eds.). (2011). *Handbook of psycholinguistics*. Elsevier.

West, D. (2000). Neural network credit scoring models. *Computers & Operations Research, 27*(11–12), 1131–1152.

Wright, W. R., & Chin, D. N. (2014). Personality profiling from text: introducing part-of-speech N-grams. *In International Conference on User Modeling, Adaptation, and Personalization* (pp. 243-253). Springer, Cham.

Xia, Y., Zhao, J., He, L., Li, Y., & Niu, M. (2020). A novel tree-based dynamic heterogeneous ensemble method for credit scoring. *Expert Systems with Applications, 159*, Article 113615.

Yu, J., & Markov, K. (2017). Deep learning based personality recognition from facebook status updates. In *2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST)* (pp. 383–387). IEEE.

Zhang, W., He, H., & Zhang, S. (2019). A novel multi-stage hybrid model with enhanced multi-population niche genetic algorithm: An application in credit scoring. *Expert Systems with Applications, 121*, 221–232.

Zhang, Y., Jia, H., Diao, Y., Hai, M., & Li, H. (2016). Research on credit scoring by fusing social media information in online peer-to-peer lending. *Procedia Computer Science, 91*, 168–174.

Zhou, L., Lai, K. K., & Yen, J. (2009). Credit scoring models with AUC maximization based on weighted SVM. *International Journal of Information Technology & Decision Making, 8*(04), 677–696.

Zhu, R., Dholakia, U. M., Chen, X., & Algesheimer, R. (2012). Does online community participation foster risky financial behavior? *Journal of Marketing Research, 49*(3), 394–407.