Holistic Credit Rating System for Online Microlending Platforms with Blockchain Technology

Yash Mahajan[1(✉)] and Shobhit Srivastava[2]

[1] Department of Computer Science, VIT, Vellore, India
mahajan.yash179@gmail.com
[2] Department of Electronics Engineering,
Gautam Buddha Technical University, Greater Noida, India
simplyshubh@gmail.com

Abstract. The inexorable rise of the Internet has given traditional microlending facilities a new online platform where people from any part of the world can lend their money to those in need of it. With the risk of defaults being tremendously high in comparison to traditional financing, there is an utmost need for the platform to be transparent and trustworthy. Keeping this in mind, we propose a model which is based on the blockchain technology and uses a holistic rating system to rate both the borrower and lender instead of the generally used rating of the Microfinance Institutions (MFIs) or Field Partners (FP), which act as intermediary and have a tie-up with the online peer-to-peer platforms nowadays. This proposed model provides security along with openness in the system and can be integrated into the existing systems for a more efficient system.

Keywords: Microlending · Peer-to-peer · Naive bayes classifier · Blockchain · Hyperledger fabric

1 Introduction

With the growth of the microfinance sector, people have started focussing more and more on the social and economical development rather than just going with the profits which is the case in traditional financial services. Microfinance per se is a broader spectrum of the traditional financial services such as loans, insurance and savings offered to people belonging to the low-income group whereas microcredit or microlending is a part of microfinance which deals with just providing these people with loans at a lower rate. This is typically done to empower the people below poverty line and support them financially for their own endeavours. Traditional finances do not provide loans to such people because it involves a high risk as the borrowers might not be able to repay it back and additionally the people being poor, do not have any collateral in the case of a default. So logically, trust between the lender and the borrower is of utmost importance and hence the whole process should be lucid.

S. M. Thampi et al. (Eds.): SSCC 2018, CCIS 969, pp. 605–616, 2019.
https://doi.org/10.1007/978-981-13-5826-5_47

Extending these services to online platforms, widens the intended audience and literally anyone from any part of the world can lend and borrow money. But this advantage itself can make the system more prone to risks as any borrower can try to dupe the lender and vanish with their money. This is where a blockchain network comes into the picture. The blockchain network validates users and provides confidence to the lenders that the people on the network are genuine and verified, thereby reducing the chances of frauds by a great percentage. In this paper, after research and testing on various blockchain architecture, namely Sawtooth, Corda and Etherium, we propose a system based on Hyperledger Fabric which is a 'permissioned' blockchain architecture developed by IBM and all the transactions happening on it can be monitored and controlled thus making it ideal for being used an architecture for microlending, which requires exclusivity.

Moreover maintaining the integrity is not the only factor for an effective solution. Instilling a sense of trust is equally important. The lender should know that if he is lending the money, he will definitely get it back. This can be done by a credit rating system. Online peer-to-peer microlending platforms can differ widely in their implementation for it [1]. In [1], online platforms are classified into two types. Sites like Kiva, Care International UK and Babylon, tie-up with local MFIs who handle their own business. These sites just offer a platform for these companies to appeal to a wide range of people. Borrowers can go to the nearest MFIs and request for a loan which in turn is posted on site with its requirements, which can be seen by prospective lenders who can get in touch with people involved. In such a system, the MFIs and FPs act as intermediaries and are rated out of 5 based on the result of their past transactions. The lenders can then judge the intermediaries based on their ratings and can decide whether to invest into the listings provided by them. In this approach the lenders do not earn any profit by lending money. Whereas in sites like Zidisha, lenders replace the intermediaries themselves and directly are in contact with the borrowers thereby making the use of a middleman completely unnecessary. In such a model, the lender does get profits based on the interest specified for the loan.

Almost all the platforms online uses a middleman to carry out their work locally, which increases the overhead cost and have an increased rate of interest for the borrowers to compensate for it. This is not ideal in the long run whereas pure peer-to-peer lending sites (Zidisha) do not make use of credit rating system just yet. In this paper, we propose a 'holistic' credit rating system which makes use of the feedback from both the lender and borrower and apply analytics on it to rate both the parties. It takes in all the inputs, analyses for discrepancies within it, and comes up with a rating. This rating can then be used to assess a possible borrower and lender alike. Additionally, we can even tag the user's interest from their past transaction and recommend the lender/borrower with borrower/lender having the same interest thereby increasing the chance of a successful transaction.

2 Existing Work

When it comes to research, research in the field online microlending platforms is very restricted but we have to derive from existing system and research work from different fields to correlate to our problem statement.

Prior research in the field of microlending platform focus majorly on how different attributes, factors and features play a role in an online microlending environment. In [1], Paruthi et al. characterise different online traits and how they influence lending behaviour - namely (a) Role of Field Partner (intermediary) ratings (b) Role of loan features (c) Role of teams whereas in [5], there was no such strong correlation after accounting for the unobservable effects in the rating. The paper [1] specifies how highly rated field partners (FP) drive more lending activities and how different aspects like gender and other features play a role in lending activities. Moreover, it also outlines that team lending behaviour are willing to take a greater risk than individuals. In this paper we incorporate all the salient features of the paper and build on it for a remedy to the existing problems.

We build a complete rating system and not just for the field partners, so that it is easy to assess all the parties involved equally. While calculating the rating keeping all the important and compared features of the paper [1]. In [4], Desai et al. realized that individual lending behaviour focussed more on the specifics of the borrower and the loan. Additionally in [2], Riggins and Weber delineate their findings that potential investor in the market base their lending activities on personal bias and not on the potential success of the loan. They majorly focus on the gender bias and occupation bias while comparing it in a regression model. But we differ from these finding in the way that, we try to eliminate these biases altogether by being transparent about each and every aspect and provide ratings as a way to compare, subtly intimating the lender to make smart investments.

While calculating the ratings of the borrower, we take into account the feedback of the borrower and rate it either 5 or 0, depending on the credibility of the feedback. In [3], Krishnaveni et al. provides a rating system based on feedback to rate the faculties in the schools and colleges. This is extended on to our proposed system but it is more efficient and whole as it considers all the aspects of the feedback and uses naive bayes classifier to classify the feedback as true or false. This is then used to calculate the rating of that transactions by giving different weights to different features while calculating the rating out of 5.

3 Proposed System

In an online peer-to-peer microlending platform, knowing the credibility of the borrower you are investing in and knowing the compatibility of both the borrower and lender is of utmost importance. In this section, we propose a credit rating system which rates the borrower out of 5 based on their past transactions which helps the lender in assessing the person they are investing in. Figure 1 shows the basic Architecture of the platform which is explained below. Furthermore, the rating of the prospective lender is personalized for each borrower based on your

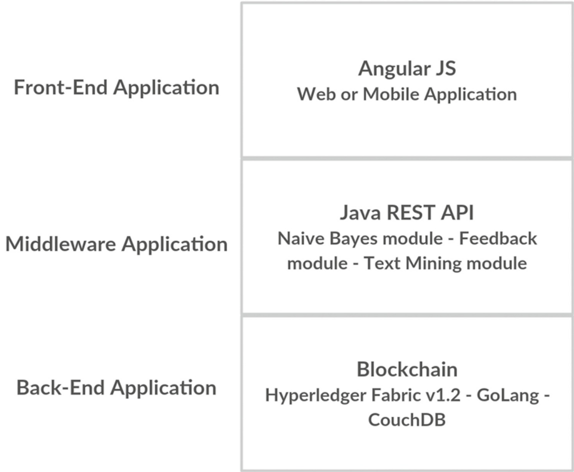

Fig. 1. Architecture of the proposed platform

their loan requirements, with the best match having a higher rating and worst match having the lowest rating and so on.

3.1 Credit Rating for Borrowers

For rating a borrower, there are a lot of factors that we need to consider, but the two main factor which we take in our proposed model is the Amount Repaid R_i and Unforeseen circumstances F_i. Amount repaid is the most important factor while calculating the rating of the borrower as it clearly specifies the capacity of the borrower to repay the money borrowed making him more credible. Unforeseen circumstances are those that are out of the control of the borrower. For example, if a person borrows one thousand rupees from a lender for cultivation of rice in his farm, but due to excessive rain the yield produced is not enough to pay the lender back then this circumstance is also taken into consideration while calculating the credit rating for the borrower. This is done by taking a feedback from the borrower and applying text mining to extract and tokenize the words and then applying naive bayes classification to it to get a score between 1 to 5. Table 1 elucidates the rating and their corresponding meaning cogently.

Table 1. Ratings and their significance

Rating	Meaning
$x \leq 1$	Frequent defaulter
$1 < x \leq 2$	High risk
$2 < x \leq 3$	Moderate risk
$3 < x \leq 4$	Trusted
$4 < x \leq 5$	Highly trusted

To calculate the credit rating for the borrower, our system gives the amount transferred more weightage than the feedback in the ratio of 1:4. The following formula, designed originally represents the rating calculations for a borrower.

if $\mathrm{R_i} \neq \mathrm{L_i}$ **then**

$$CreditRating(C_i) = 0.8 * ((\frac{R_i}{L_i}) * 5) + 0.2 * (F_i) \tag{1}$$

else

$$CreditRating(C_i) = (\frac{R_i}{L_i}){*}5) \tag{2}$$

end if

where R_i is the amount repaid by the $borrower_i$, L_i is the amount to be paid by the $borrower_i$ and F_i is the rating of feedback provided by the $borrower_i$.

For example, if the borrower gets 100 from the lender and because of the rains is unable to pay him back and only pays back 80 out of the 100 to be paid and his reason is genuine then the credit rating will be calculated in the following way:

$$C_i = 0.8 * \frac{80}{100} * 5 + 0.2 * 5 \tag{3}$$

Which comes up to 4.2 rating for this particular transaction. This value can then be averaged out with the past transaction to arrive at the cumulative rating using the given average formula:

$$FinalRating(C_{\mathrm{F}}) = \frac{\sum_{i=1}^{N} x_i}{N} \tag{4}$$

where C_F is the Final Cumulative Rating of the $borrower_i$, x_i is the rating received by the borrower for the *ith* transaction and N is the total number of transactions.

3.2 Credit Rating for Lender

Similar metrics and parameters cannot be used to rate the lender and hence we use a different algorithm for calculating a personalized rating for lenders from the borrowers point of view.

The four main parameters that we have taken for judging a lender are (a) Rate of Interest, (b) Duration of the loan, (c) Amount of loan. The borrower enters these field for the particular loan he/she is requesting and then the average of (a), (b) and (c) is calculated for each lender separately. These values are then added to a vector <a, b, c>. Each attribute is given a specific weight based on it importance in the rating calculation. The vector of the borrower is then compared with each and every lender's vector to calculate the similarity and

Table 2. Weightage for attributes a, b and c

Attribute	Weightage
W_a	0.4
W_b	0.3
W_c	0.3

correlation. Based on the correlation, the lender is assigned a rating out of 5. This creates a rating for each lender based on the loan requested.
To calculate the similarity, we make use of the given originally derived formula -

$$C_R = (\sum i \in a, b, c) \frac{W_i * \frac{((abs(X - X_i)))}{max(X, X_i)}}{|X|} \tag{5}$$

where C_R is the Credit Rating for the Lender, W_i is the corresponding weightage of the attribute under consideration taken from Table 2, X is the vector of the *borrower* and X_i is the vector of the $Lender_i$.

Based on this formula if the borrower requests for a loan interest of 10%, duration of loan in months is 12 and amount of loan is 1000 whereas if one lender has an average loan interest (based on past transactions) of 8%, average duration of loan as 20 months and amount of average loan as 2000 then the correlation value will be calculated as 0.367 which is moderately similar. The rating then is calculated by $(1 - C_i) * 5$ which gives an accurate rating out of 5 for the given lender. In this case, it is 3.165.

3.3 Rating for the Feedback of the Borrower

For calculating the genuinity of the feedback we make use of the given data set which is prepopulated and apply Naive Bayes Classifier to it. There are namely two classes - Yes or No and two attributes which we take into consideration - circumstance, which mentions the reason why there was a failure in repayment, and purpose, to conclude whether the given reason does make sense for purpose he was using the money for. It does not make sense to put draught as a reason for the failure for a technological start-up. Using this data set we then compare the feedback to classify the feedback as genuine or not. This happens in three steps:

(1) Collect Feedback. The feedback is collected from the borrower from the frontend application made on HTML, CSS and AngularJS, only if he is not able to repay the amount he has borrowed, the feedback is stored on to the database.

(2) Text Mining. The stored feedback is then split and tokenized using the Splitter class from the NLTK (Natural Language Toolkit) module. The feedback is split into its constituent words and stored in a list with each entry having an

Table 3. Part of the sample dataset for genuine reasons

Reason	Valid
Flood	Yes
Draught	Yes
Hailstorm	Yes
Light rainfall	No
Earthquake	Yes

independent word. For example, if the borrower write a feedback saying,"Could not repay because of floods", this will be converted into a list -

["Could", "not", "repay", "because", "of", "floods"]

This list will then be compared against Table 3 for similarity and only matching words will be stored in the Final List (Fl).

(3) Naive Bayes Classifier. The Naive Bayes is collection of classifier based on the Bayes Theorem which is stated mathematically as follows:

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)} \tag{6}$$

Where B stands for the matches between the input and the sample dataset (Table 4), A is the total number of rows belonging to a particular class (Yes or No) and P(A) is the total number of rows under a particular class divided by the total number of rows in the data set.

Table 4. Part of sample dataset for valid/invalid based on two parameters

Circumstances	Purpose	Genuine?
Draught	Agriculture	Yes
Flood	Agriculture	Yes
Hurricane	Agriculture	Yes
Draught	Export business	No
Flood	Technical start-up	No
Hurricane	Gym	No
Flood	Transport service	Yes

On applying the Naive Bayes classification, it compares the value of the input to that of the sample data set and calculates a value for each class (Yes and No). The class which produces the maximum value when the algorithm is executed is the class to which the input belongs. If the input belongs to the class Yes then a rating of 5 is given to it otherwise a rating of 0 is given to it. This simplifies our approach to checking the credibility of the feedback provided and this value can then be used in the credit rating calculation of the borrower.

3.4 Keeping the System Profitable

For the smooth functioning of our online peer-to-peer microlending platform, the profitability of our platform needs to be considered at all times. Microlending is in itself a high risk business wherein you invest your money into a business/start-up which you think might yield success if given the proper resources. The business/start-up has equal probability of success and failure and the risk of defaults is higher than any other traditional financial services. To maintain the system in profit, the individuals investing should be encouraged to make smart investments. Hence to keep our microlending platform in a profitable state is very difficult.

For example, if investor(I) wants to lend 100, he/she can do it in two ways - (1) invest the whole amount into a single enterprise or (2) divide the investment between different borrowers. (1) is a high risk trade because if the borrower defaults, the whole amount invested is lost incurring a huge loss to the investor whereas in (2) the risk is divided along with the money between the borrowers and the threat of losing all the invested money is minimized to a great extent. This can be seen when calculating the probability of all the investments defaulting. If P(X) is the probability of event X failing then

$$P(X_{\mathrm{i}}\, and\, X_{\mathrm{j}}\, and...X_{\mathrm{n}}) = P(X_{\mathrm{i}}) * P(X_{\mathrm{j}}).. * P(X_{\mathrm{n}}) \tag{7}$$

Which will fall close to zero as the value of 'n' increases. Hence to minimise the risk involved in any transaction, it is better to hedge the investments than to put all the 'eggs in one basket'.

Moreover, dividing the investments does not solve our problem of maintaining profitability of the whole microlending platform. Investing the same amount between a list of borrowers merely minimises the amount of loss incurred to a investor but does not guarantee profit. For example, if an Lender (L) decides to invest 100 between two borrowers B_1 and B_2 such that borrower (B_1) gets 20 and borrower (B_2) gets 80 at interest rates say X% and Y% respectively. If borrower (B_1) fails to repay the amount he was supposed to, the lenders incurs a loss of 20 and to make up for it, even without interest, he would have to increase the percentage of interest for borrower (B_2) by 25% which is a lot when thought about. This percentage increases even more if borrower (B_2) defaults.

So our solution to minimising such losses and increasing the profits is to remove the frequent defaulters altogether. According to our rating system, if the feedback by the borrower is genuine, they will earn a rating of 1 at least, irrespective of whether they are able to pay any money back or no. Using a strict and no-tolerance approach, we remove all the borrowers who earn a rating of less than one (<1) immediately from the microlending platform. Also if the average rating of the borrower for its last two transaction is less than 1.5 (<1.5), we remove the borrower from the system as he contributes to the net loss incurred by the platform. Using this approach, the frequent defaulters are constantly removed from the system and profitability is maintained by reducing the risks of defaults.

3.5 Blockchain Implementation

When it comes to a system which needs trust to be a profitable, there should be a way to authenticate the users and that is where the blockchain network comes into the picture. In our model blockchain network is used to validate a user and only then allow him to join the network. To verify the user and his credibility we have used the accepted verification for the particular country the individual belongs to. After that, all the existing members of the network needs to approve the user's request to join the network and only upon that will the user will be allowed to enter our microlending platform. This way the fake profiles or accounts are completely removed and there is no room for dubious activities and accounts.

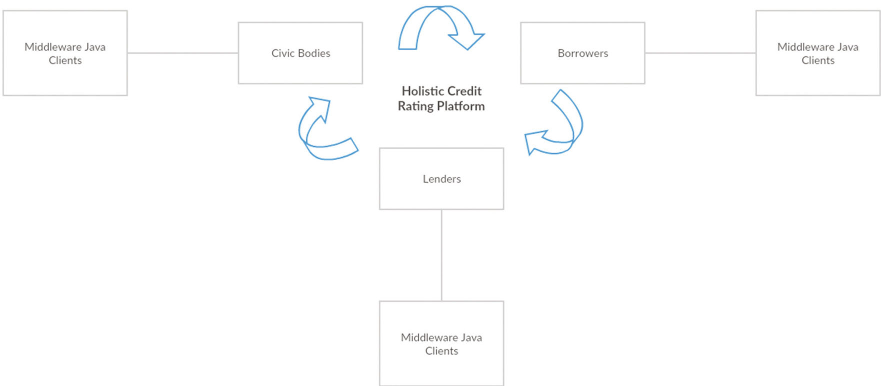

Fig. 2. Blockchain architecture for the system

There are vast array of options available for implementing a blockchain network - Hyperledger Fabric, Hyperledger Sawtooth, Ethereum and Corda being the most famous amongst the lot. But the features and functionality provided by Hyperledger Fabric perfectly suits the microlending platform which we intend to propose. Hyperledger fabric is a framework for 'permissioned' membership, where all participants have known identities so it serves our purpose as all members on the network need to know that everyone on the microlending platform are genuine and if they are investing in a borrower then they can rest assured that the person they are lending to is a genuine person and not someone virtual entity intended the dupe the investors. Using the concept of channels, we provide a secure and private environment for the transaction after the borrower and the lender have been matched. The implementation provides us with a transparent, trustworthy and a secure network for the proper functioning of the microlending platform.

In our implemented microlending platform, our architecture (crypto-config.yaml) looks like this - one organization - with three peers to it - namely one for the borrowers, one for the lenders and one for the intermediaries (Field Partners and MFIs). Figure 2 shows the blockchain architecture for our system. There is an additional peer for an auditor to review all the transaction on the network for the accuracy, integrity and fairness. There exists a common channel for all to see the current transactions that are going on in the microlending site. When there exists a local body governing the transaction (FPs and MFIs) then, it will be made the endorser to check for the validity of the transactions. The network will be established with chaincode functions like - createLender(), createBorrower(), calculateBorrowerRating() and so on. A blockchain ledger consists of two types of records: individual transactions and blocks. The first block consists of a header and data that pertains to transactions taking place within a set time period. The block's timestamp is used to help create an alphanumeric string called a hash. Each block contains a cryptographic hash of the previous block, a timestamp, and transaction data (generally represented as a merkle tree root hash). By design, a blockchain is resistant to modification of the data. For use as a distributed ledger, a blockchain is typically managed by a peer-to-peer network collectively adhering to a protocol for inter-node communication and validating new blocks. Once recorded, the data in any given block cannot be altered without alteration of all subsequent blocks, which requires consensus of the network majority. We make use of this concept for keeping the system only with any valid and genuine person

3.6 Assumptions

There are assumptions that we have made while deriving the formulas given above. Firstly we assume that the borrower is honest and genuine while writing the feedback and is not involved in wrongful activities. The borrower can deceive by putting in wrong information which might not even be the reason behind its loan getting defaulted which might produce unexpected results so to ensure that the system is consistent, we make assume that the borrower does not enter wrong information. Secondly we assume, there is no discrepancies between the observed data and the expected data.

4 Results and Future Work

Our proposed model of an online microlending platform based on the blockchain network with smart credit rating system betters the current implementation and eliminates a chunk of difficulties faced by microlending platforms both online and offline.

Lack of Trust

The major reason why the microlending platforms and agencies exists is because the banks consider necessitous people too much of a risk to invest in and so these platforms and agencies acts as middlemen connecting the lenders to the borrowers with a high interest rate and an equally high risk of defaults. Most of the times, the lenders do not even know where the money they have lent is going to and this brings certain uncertainty and doubt in the lender which might lead to a restricted approach. Moreover with the current implementation, lenders can only 'hope' that the investment they have made turns out to be a smart one and yields results and does not lead to defaults. The model that we propose tries to eliminate this 'lack of trust' completely. The proposed model is based on the hyperledger fabric network, as it was explained earlier, which is a permissioned network so not anyone can join the network. To join the network, a new user will require the approval of all the nodes in the network. This will bring genuinity and trust to the platform as the lenders can rest assured that all the members (borrowers) of the platform are legitimate and can be trusted. With the fabric network, we implement our own credit rating system to rate both the borrowers and lenders smartly. With the introduction of smart credit rating, both the borrower and lender can feel a sense of trust and confidence between them - lender because the credit of the borrower provides an equivalent probability of loan being repaid and borrower because of the trust the lender has shown in him being giving him the loan readily.

More Investment and More Profit

The lack of trust, makes the intermediaries (FPs and MFIs) an important entity in moderating the transactions and maintaining the trust. But the downside to this is that due to the existence of these intermediaries, lenders are unable to make any profits in the current implementation of microlending platforms. In our proposed model, we eliminate the need for the intermediaries (in contrast to the classical microlending platform) by directly connecting the people in need with the people willing to invest from any part of the world. This way the lenders gets all the profit earned through their investment and pushes them to invest more. Moreover trust is built with the system and according to [6], Bottazzi et al. predict that there is a positive relationship between trust and investment and that earlier stage investment requires high trust in venture capital and this applies to microfinance investment too. So basically increase in trust leads to increase in the number of investment any lender makes and consequently increase in the overall profit of the system. This is not only beneficial to the platform but also to the prospective borrowers who are in dire need of money as this increases the chances of them getting a loan from a lender.

Smart Investment

The investor needs to invest smartly to make profits, increase the profits and in turn increase the profit of the whole system for the smooth functioning of it. Sometimes, the lender invests randomly or let's emotion come in the way while investing instead of investing smartly which increases the overall risk factor. We help the investor in making smart investments by suggesting the investor to divide his investments amongst a number of borrowers so that the overall risk decreases and the overall profits of the investor increases as there is a very less chance of all the investments failing to give its returns, whenever necessary. Furthermore, we also remove the frequent defaulters from the system to keep the system in a profitable state overall. This way, all the stakeholders in the system remain happy - gaining more investors for the borrowers, reducing the risk involved in the investment for the investor and keeping the system in high profits.

With the rise of technology, in the future we can improve the efficiency of the system and find a better way to increase the genuinity in the system overall but more importantly in the validating the feedback provided by the borrower for his default.

5 Conclusions

In this paper, we talk about the different problems existing in the current microlending platforms and we suggest a solution to it by proposing a system which is based on the blockchain technology namely - hyperledger fabric. Additionally we provide an unique rating system to rate both the parties involved in a transaction - lender and borrower to maintain a clarity between the two. Successful implementation of this has yielded definite results till now.

References

1. Paruthi, F.G., Frias-Martinez, S.E., Frias-Martinez, T.V.: Peer-to-Peer microlending platforms: characterization of online traits. In: 2016 IEEE International Conference on Big Data (2016)
2. Riggins, F.J., Weber, D.M.: Information asymmetries and identification bias in P2P social microlending. In: 2015 48th Hawaii International Conference on System Sciences (2015)
3. Desai, R.M., Kharas, H.: Do philanthropic citizens behave like governments? In: Internet-Based Platforms and the Diffusion of International Private Aid (2009)
4. Krishnaveni, K.S., Pai, R.R., Iyer, V.: Faculty rating system based on student feedbacks using sentimental analysis. In: 2017 International Conference on Advances in Computing, Communications and Informatics (2017)
5. Hartarska, V., Nadolnyak, D.: Does rating help microfinance institutions raise funds? Cross-country evidence. Int. Rev. Econ. Finan. **17**(4), 558 (2008)
6. Bottazzi, L., Da Rin, M., Hellmann, T.F.: The importance of trust for investment: evidence from venture capital. In: Forthcoming, Review of Financial Studies, p. 571 (2016)