LAPS: Computing Loan Default Risk from User Activity, Profile, and Recommendations

1st Wisnu Uriawan  
Informatics Department (INSA de Lyon)  
LIRIS Laboratory CNRS UMR 5205  
Villaurbanne CEDEX, France  
wisnu.uriawan@insa-lyon.fr

2nd Omar Hasan  
Informatics Department (INSA de Lyon)  
LIRIS Laboratory CNRS UMR 5205  
Villaurbanne CEDEX, France  
omar.hasan@insa-lyon.fr

3rd Youakim Badr  
Computer Science and Engineering Department  
The Pennsylvania State University  
Malvern (PA), USA  
yb@psu.edu

4th Lionel Brunie  
Informatics Department (INSA de Lyon)  
LIRIS Laboratory CNRS UMR 5205  
Villaurbanne CEDEX, France  
lionel.brunie@insa-lyon.fr

Abstract—The credit score is one variable in receiving a loan application from a bank or financial institution that provides credit/loan. Many factors determine whether a borrower gets the loan. One of them is through more valuable collateral than the loan that was proposed. However, this is not possible for borrowers to provide it. Personal data, job information, salary amounts, assets owned, and valuable documents are usually required to determine a credit score. We build a personal lending platform model based on the trustworthiness score called LAPS (Loan Risk score, Activity score, Profile score, and Social Recommendation score) borrower trustworthiness score. The borrowers’ trustworthiness is an absolute requirement to ensure they can repay the loans and installments on time. We present the practical ways to select the best features from the Bank Marketing dataset. The feature selection of the dataset applies to blockchain applications. The advantage of LAPS is introducing recommenders’ as guarantors to convince the lenders'/investors’ and minimizes collateral by implementing a LAPS.

Index Terms—Activity, Blockchain, Collateral, Dataset, Features, Lending Platform, Profile, Loan Risk, Recommendation, Trustworthiness.

I. INTRODUCTION

Credit scoring issues will help the bank or financial institution get valid information, and several features describe the eligible borrowers [1]. A personal loan is part of financial services for who person applying for some loan. Traditional mechanisms show weaknesses because it takes time uncertain (it tends to be longer), require many documents, additional costs, etc. Finally, there is no transparency when the borrower is approved or rejected. In the traditional lending application process, persons apply for a loan because they need some funds to support family members, rent a house, buy a car, etc. So they try to find a loan shown in Fig. 1. Many borrowers have been rejected because they do not meet terms and conditions [2].

Today, banks and financial institutions provide loans with terms and conditions that are not easy for borrowers to fulfill these conditions [3]. Banks or Lending marketplaces offer loans and still require collateral to guarantee that borrowers return their loans. Collateral can be in the form of assets that are easier to become money. A guarantor is a person who gives some guarantees to borrowers while applying for some loans.

Type of debt financing and approval percentages are shown in Table I below. Borrowers approval rates are shown in Cash

Fig. 1. A Traditional lending system

Advance Lenders 90% is higher because of fast processing about 1-3 days approval, next followed by Alternative Lenders reach 70% loan processing environs 5-7 days, Traditional Banks about 45%, about 25% time processing about 14-30 days is the last less percentage is Large Banks. Table I shows the scale of ratio and time processing impacts borrowers’ proposal of some loans [4]. Table I describes it is still difficult to obtain some loans from traditional lending systems. The percentage approval was assumed from 100 borrowers. Of the Large Banks, 25 were approved, and 75 are denied in proposed
loans. We provide the LAPS formula as a solution to the

<table>
<thead>
<tr>
<th>Type of Debt</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional Banks</td>
<td>45</td>
</tr>
<tr>
<td>Cash Advance Lenders</td>
<td>90</td>
</tr>
<tr>
<td>Alternative Lenders</td>
<td>70</td>
</tr>
<tr>
<td>Large Banks</td>
<td>25</td>
</tr>
</tbody>
</table>

problem above. It contains the Loan Risk, Activity, Profile, and Social Recommendation score as a borrower trustworthiness score. The borrower has confidence after receiving the trustworthiness score. With this score, the borrower does not need any more collateral. From the lender/investor side, they get assurance that the borrower can repay the loan, and the recommender bridges the gap between the borrower and the lender. With LAPS, all users will get incentives that can be applied safely. In addition, we present the advantages of LAPS when it applies to the Ethereum blockchain-based application [5], which illustrates how to feature selection from UCI Bank Marketing public dataset [6]. This paper’s remainder is structured as follows: Section I introduces the traditional lending system, stakeholders, and background literature. Section II related work. Section III is our proposal. Section IV result and discussion. Section V concludes this paper.

II. RELATED WORK

We found that most of the existing solutions reviewed still can be improved by detailing this research. This section’s work review is found in the literature-related documents. The research mentioned how the Blockchain-Enabled Social credits System (BLESS) applied in the system leverages the decentralized architecture of the blockchain network, which allows grassroots individuals to participate in the rating process of a social credit system (SCS) and provides tamper-proof of transaction data in the trustless network environment. The anonymity in blockchain records also protects individuals from being targeted in the fight against powerful enterprises. A smart contract-enabled authentication and authorization strategy prevent unauthorized entities from accessing the credit system. The BLESS scheme offers a secure, transparent, and decentralized SCS. However, they have difficulty implementing technology in social aspects such as public acceptance and mass adoption [2].

Their research is developing a credit-scoring model using logistic regression and multivariate discriminant analysis applied in Moroccan Financial Institutions (MFIs). The model combines behavioral and descriptive data related to the borrowers (age, activity, level of education, number of unpaid debts, number of loans, etc.) and (amount of credit, duration of credit, number of concluded loans per portfolio manager, etc.). The weaknesses are required a more extensive data sample, a deep enough history of the behavior of the customer, and also more information about variables related to the client’s activity and its performance to predict the default better [7].

III. OUR PROPOSAL

Features selection is needed to choose high-quality data. The selection process requires a parameter or expected value corresponding to data availability. In particular, personal data, educational background, marital status, family members, financial data, job information, and collateral. The comparative study of several methods like Decision tree [8], statistical and Artificial Intelligent [9], Rough and Tabu search [10] are relevant with research area some author had to deliver the message bring some information for selection feature, and describe the result. Features collection of datasets is tested to indicate suitability for applying the formula/model. Datasets are used from UCI machine learning is a public dataset [6].

A. Dataset Features Selection

Data sources in Fig. 2 are built with some components to be analyzed, features, and fields with a particular purpose. The phase of analysis in Fig. 3 is identifying what needs and being understood, features are about what kind of information needs and more specific, and fields describing an object are analyzed (applicant) [6]. Datasets contain the features collections and row data in Fig. 3, which should be analyzed like the data required approaches in this case. We use the association rules and the weighting approach to select the suitable features for this research. Some features are selected to represent the variable candidate with the high impact factor. The features selection is essential for choosing the best variable to support personal lending [11]. The process will start with feature collection from datasets see in Fig. 3, then continue with features selection by applying a suitable method and testing all the features selected, and the same cycle for the following features until found the highest score of features. In this section, the feature prediction phase predicts the appropriate features for the meaning of borrowers’ candidates. Then, in the features selection methodology, apply some methods to get the best feature corresponding to the borrowers’ profile. After the features selection has finished, we must test all features sample to ensure it is suitable for our research. Finally, we have the best features selection, supporting the credit scoring for selecting the best applicant with the minimum risk.

B. LAPS Splitting Formula

We briefly discuss membership functions variables rules for making a decision and define the trustworthiness score in terms of four variables [5], namely LAPS (Loan Risk,
Activity, Profile, and Social Recommendation) as borrower trustworthiness score [12], see Equation 4. The authors applied the Bank Marketing dataset from UCI public dataset [6] show in Fig. 2:

1) **Loan Risk score** is the component for measure the borrower candidate has the other loan such as housing, car, etc. in Fig. 4 (a), (b), if there is any another loan is risky to allowing get another loan and will decreasing the trustworthiness score, see Equation 1.

\[
\text{Loan Risk score} = \sum_{i=1}^{n} (w_i \ast L_i) \tag{1}
\]

where:
- \( w = \text{Weight for each variable } \{w \in \mathbb{R} \mid w \leq 1\}, \) that able to be defined by user.
- \( i = \text{Sequence of weight and variable}. \)
- \( L = \text{Variables (loan, housing), where } \{L \in \mathbb{Z} \mid L \leq 100\}, \) and scale of values are between 0 to 100.

![Fig. 4. Loan and Housing Dataset (a) and List of Loan and Housing (b)](image)

2) **Activity score** describing the borrower activity in occupation such as job or business activity in Fig. 5 (a), (b), to measure the ability to pay and considering the credit plafond or credit limit that correspond with their activity, if borrower candidate has a good occupation they will get the highest value of Activity score, see Equation 2.

\[
\text{Activity score} = \sum (A) \tag{2}
\]

where:
- \( A = \text{Variable (Job activity), where } \{A \in \mathbb{Z} \mid A \leq 100\}, \)
- and scale of values are between 0 to 100.

![Fig. 5. Job Dataset (a) and List of Job (b)](image)

3) **Profile score** is the personal data of borrower candidates such as age, education level, and marital status in Fig. 8. These variables support to trustworthiness score. For example, the borrower should be older than 18 years old and 88 years old maximum age show in Fig. 6 (a), (b), [13], have an education level in Fig. 7 (a) to measure the economy and activity in industry or entrepreneur, and have marital in Fig. 7 (b) status to consider the family dependent. All the variables summarise in one variable as **Profile score**. The formula to get the Profile score is shown in Equation 3:

\[
\text{Profile score} = \sum_{i=1}^{n} (w_i \ast P_i) \tag{3}
\]

where:
- \( w = \text{Weight for each variable } \{w \in \mathbb{R} \mid w \leq 1\}, \) that able to be defined by user.
- \( i = \text{Sequence of weight and variable}. \)
- \( P = \text{Variables (age, education, marital) are } \{P \in \mathbb{Z} \mid P \leq 100\}, \) where scale of values is between 0 to 100.

![Fig. 6. Range of Age (a) and Diffusion of Age (b)](image)
4) The social recommendation score is the primary variable for borrowers to get support directly from other users to add recommendation value. This value serves as a guarantor for borrowers to get some loan from lenders/investors through the lending platform as shown in Equation 4 and 5.

Social Recommendation score = variables \( S \) (Social Recommendation) are \{ \( S \in \mathbb{Z} \mid S \leq 100 \}, \) where scale of values is between 0 to 100.

IV. RESULT AND DISCUSSION

The experiment is a sequence of the dataset obtained as shown in Fig. 9. The features that have been selected as results are 6 (six), including age, marital, education, job, housing, and loan. LAPS captured from each feature and then converted to number value as shown in Fig. 13. The borrowers get value (Loan Risk, Activity, and Profile score) from example dataset. The Social Recommendation score will get by the other user as Recommenders.

A. Result

This section describes the process splitting formula according to the dataset to obtain the selected suitable features.

1) Age and Marital status features selection as shown in Fig. 10 (a), (b), have impact to ability to pay back the loan. Age and Marital status show the borrowers’ family members and conditions. Dataset shows distribution range of age with percentage indicated in Fig. 10 (a). Age is the one important variable that also shows the productivity of the borrower. The Marital status dataset shows the list of status borrowers in Fig. 10 (b). Borrowers of productivity ages and single have the ability to pay installments on time.

2) Education and Job features selection is shown in Fig. 11 (a), (b), indicating the borrower candidate with a higher education level will get a good job position opportunity. The impact of education on salary changes significantly. The most substantial effect of education is also expressed at the highest level of education. Even under other factors, the dominant role of education on salary. The borrowers have higher education returns, particularly above the high school level. In general, we find that the higher the education in line with their salary growth.

3) Housing and Loan features selection are at greater risk of housing instability compared to homeowners in Fig.
12 (a), (b). Among the factors contributing to that risk are financial situations. Debt housing is riskier and far more likely than homeowners to pay more than 30 percent of their income in housing costs.

(a)  (b)  

Fig. 12. Percentage of Housing (a) and Percentage of the other Loan (b)

4) Borrower trustworthiness score presented the eligible borrowers after they have the trustworthiness score. The borrowers’ trustworthiness score gives the borrowers scores after registering with a default value for the first time. It will increase the borrowers’ activity in the lending process and activity in the payment process (on simulation). The recommenders can give excellent recommendations to borrowers who propose a loan. An essential part of the personal lending simulation is a recommendation that aims to reduce dependence on collateral.

\[
Trustworthiness\ Score = Loan\ Risk\ score + \text{Activity score} + \text{Profile score} + \text{Social Recommendation score}
\]  

(4)

with:

- **Trustworthiness Score**: Borrower trustworthiness score.
- **Loan Risk score**: Information of the record from another loan of Borrower.
- **Activity score**: Business activity or job information of Borrower.
- **Profile score**: Personal information of Borrower.
- **Social Recommendation score**: The recommendation value of Borrowers from Recommender.

5) The LAPS formula is a commitment between borrowers, lenders/investors, and recommenders set by the smart contracts management so that all parties understand each other’s obligations and risks that will be accepted. The variables include Loan Risk, Activity, Profile, and Social Recommendation. All data will be assessed as a borrowers’ trustworthiness score (LAPS). LAPS formula, we add positive weight for each variable, in equation 5:

\[
LAPS = (w_l \times \text{Loan Risk score}) + (w_a \times \text{Activity score}) + (w_p \times \text{Profile score}) + (w_s \times \text{Social Recommendation score})
\]  

(5)

where \(\{w \in \mathbb{R} \mid w \leq 1\}\), and \(w_l, w_a, w_p, \text{and } w_s\) are positive weights of the trustworthiness parameters such that \(w_l + w_a + w_p + w_s = 1\). The weights of the trustworthiness attributes are predetermined based on their priority value that can modify by consensus. For example, \(w_l = 0.25, w_a = 0.2, w_p = 0.25, w_s = 0.3\). In this example, the social recommendation is given the highest percentage, and activity is given the lowest value it shows that the social recommendation is the priority to measure the eligible borrower candidate. Equation 5 is the complete formula for trustworthiness score after weight added is supportive to imprecise conclusions. After they get the score (see in Fig. 13 and Fig. 14), they can propose some loans with their borrowers’ trustworthiness score and determine the maximum loan. The borrowers’ trustworthiness scores will increase alongside the track record of borrowers’ payments. After converting the selection dataset, then grouping the features in line with the LAPS variables formula seen in Fig. 14, the features are mentioned in the splitting formula. We obtained the selection feature grouping. The experiment applied these features selected to a personal lending prototype. The first experimental results were seen in Fig. 15 show the borrowers trustworthiness score, and the first user obtains 79, following the other users, after the recommender gives the value in the prototype. The LAPS formula will compute the trustworthiness score.
The second experiment in Fig. 16 shows with different data tests (variables changes values), the result of borrower trustworthiness score is positive values. The prototype succeeded in computing the selected features from the LAPS formula. The system will automatically increase the value of the borrowers’ trustworthiness score (LAPS). The borrower will be able to propose a more significant amount than before if their score rises. The borrower with a high trustworthiness score will be easier to propose loans with increasing loan plan limits in the next cycle. Smart contracts management at borrowers, lenders/investors, and recommenders sides will handle each functionality from the available services on the Ethereum blockchain. With the limitation of the available digital wallet account tests, the test runs per each account with the equal method for all datasets.

B. Discussion

Features selection is a part of choosing the best variable for supporting personal lending, and many features could be selected. Credit Scoring is the most important for selecting the best applicant with the minimum risk. Some features will help describe the conditions for the borrower to pay back the loan. On the other hand, lenders/investors cover their risk with these features, and their money can not return. First, features data will be compiled with the standard method, and the result will be analyzed combined with another method. Secondly, all features try to connect with others and choose the relevant features. The result will show how to allocate the relation between two features. Next, try to rank each feature with the highest score and follow the following score until finished.

The presented the LAPS formula the challenges and open problems previously discussed. The formula covers minimizing collateral when borrowers propose a loan. All variables (Loan risk, Activity, Profile, and Social recommendation score) support borrowers to get a loan from lenders/investors. The LAPS formula is well adapted to the personal lending platform to accommodate the recommenders and lenders/investors to decide.

V. Conclusion

This research aims to compute the trustworthiness score (LAPS) to provide a reliable borrower. The result obtained the selected features from UCI Bank Marketing public dataset with the highest impact factor weight. The variables in this context are the categorical type features converted to quantitative. The LAPS formula shows all borrowers’ activity by referring to the personal lending prototype, which directly interacts among borrowers, recommenders, and lenders/investors. The LAPS model describes the scoring of trustworthiness that has been successfully applied to the personal lending prototype. Lenders can use the trustworthiness score to decide the eligible borrowers’ candidates.

Acknowledgment

The first author wishes to acknowledge the MORA Scholarship from the Indonesian government and INSA de Lyon LIRIS laboratory UMR 5205 CNRS, France which partially supports and funds this research work.

References