DPU **Dr. D. Y. Patil B-School**
(Program Approved by AICTE, Ministry of Education, Govt. of India)

Blockchain and machine learning based peer-to-peer lending for the post-pandemic economy

S. Gogia[1*] and U. Sharma[2]
Amity University, Noida, India

ABSTRACT

The COVID-19 pandemic has had a two-pronged effect – health and economy. Economies worldwide have been impacted and forced to shut during the lockdowns. In India, SMEs had already been struggling to stay afloat due to absence of financing, creating an inadequate cash flow. With frequent lockdowns, the situation has worsened, making bank loan applications difficult. This paper explores P2P lending in the context of micro-loans using a blockchain-powered solution to provide the necessary funds in a quick, easy, secure and contactless manner. As part of our Proof-of-Concept, a peer-to-peer decentralized Web Application is setup, utilizing the services of the Ethereum blockchain.

KEYWORDS: Blockchain, Lending, Finance, Credit, Machine Learning, Financial Analysis

1. INTRODUCTION

Micro loans are critical for the sustainability of small businesses. Household ventures, SMEs and financially-stricken individuals heavily rely on such loans. Loan accessibility is important, especially for those who might not have perfect loan applications for banks is important to aid them in an uncertain and volatile (Albulescu & Claudiu Tiberiu, 2020) post-pandemic economy.

Our inspirations for this paper and the platform are the financial problems created due to the COVID-19 pandemic, particularly for small businesses (Beauregard, R., et al., 2020) regarding the growing popularity and merits of peer-to-peer lending. The proposed solution is planned to be a decentralized, automated and time-effective framework. It lessens operational dangers and obliges the requirements of our intended interest group – SMEs, new businesses and people who don't have good credit applications. The rising popularity of peer-to-peer lending (Demyanyk, Y., et al. , 2020) among the youth, which makes up for a significant part (Sharma, U. & Manchanda N., 2020) of the entrepreneurship headcount, is a significant factor behind the success of such platforms. The needs of the borrower and

lenders are met in a completely decentralized approach. The perspective requirements of these two sections are:

1.1. Borrower's perspective

Borrowers on the platform should be able to raise loan requests with the following fields – loan amount and a short description of the same. Along with these details, a credit score of the borrower, extracted using a Machine Learning pipeline, discussed in subsequent sections, is attached to the request.

All requests are available to all lenders, and the platform blockchain setup allows for complete transparency. The prevalence of this competition amongst lenders allows for competitive interest rates and reasonable loan terms.

1.2. Lender's perspective

The credit score of the borrower helps the lender in assessing their credibility. Inclusion of a reliable metric to enable lending with strangers over the internet is a key part of our proposed platform because of the role of appearance (Duarte, J. et al., 2020) in peer-to-peer lending.

The decision on submitting a proposal is completely on the mercy of the lenders, as the credit scoring mechanism doesn't follow a binary approach of approved or rejected loan requests, in line with the decentralized nature of the platform. This paper is divided into the following sections: Related Work, Proposed Approach, Experimental Proof-of-Concept, Conclusion and Discussion, and References.

2. LITERATURE REVIEW

In paper (Iftekar, H. et al., 2020) the authors discuss the aggravating effect of the pandemic on the financial sector, exacerbated with the level of government restrictions to tackle the virus' spread, with firms' financial constraints and reliance on debt financing. The authors in paper (Treacy, W. F., & Carey, M., 2020) examine the credit risk rating system used by major US banks.(Zhang et al., 2000) discuss the dramatic impact of the pandemic on financial markets globally. In paper (Michael. C., et al.. 2020), the authors elaborate on the current credit risk models. (AE Khandani et al.) confer consumer credit risk models using Machine Learning. (R Henriquez et al., 2019) explore peer-to-peer lending on a blockchain platform in the context of mortgage financing. (R Wang et al., 2019) discuss the increasing role of blockchain in regards to SME financing. (D West, 2000) investigates the credit scoring accuracy of five neural network models.

3. PROPOSED APPROACH

In this paper, the structure of the proposed peer-to-peer lending platform is laid down, along with an experimental Proof-of-Concept using Ethereum Blockchain, due to its security and popularity (2014).

We use Machine Learning and techniques like Weight of Evidence and Information Value for credit scorecard generation, which plays a vital role in the overall functioning of the platform. Logistic Regression algorithm is used for the modelling due to its wide acceptance (Gang, D., 2010) in the development of such scorecards.
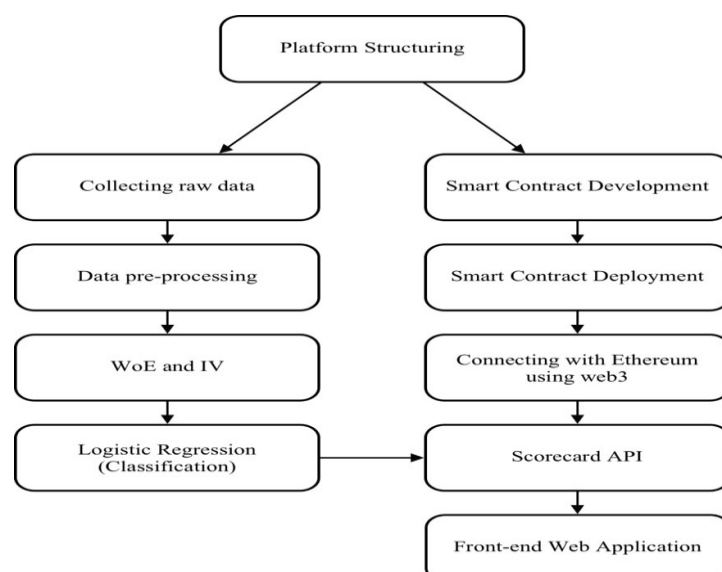

Fig. 1: Generalized Framework for the POC

4. EXPERIMENTAL PROOF-OF-CONCEPT

Tools used

For the Proof-of-Concept Web Application, React is used, offering a highly scalable and responsive platform. Leveraging the web3.js library, the application connects to the Ethereum Blockchain. Leveraging the perks of the powerful Ethereum ecosystem, core logic of the platform is implemented in Solidity Smart Contracts. Python is used for the credit scorecard model, which is further formulated as a Flask API to facilitate communications with the Web Application. Python and React were selected due to the availability of well-maintained popular libraries.

Dataset, cleaning, and pre-processing

For the development of the credit risk scorecard, we used a dataset (www.kaggle.com) made available on Kaggle by the Lending Club, a US-based P2P lender. This dataset relates to the loans sanctioned between 2007 and 2014 by the said provider.

The dataset has over 75 features, ranging from loan status to customer information. In the first step of the pipeline, attributes with more than 50% missing values were omitted. Certain platform-specific attributes were also removed. Examples of these are the id of customers, membership identification, etc.

The status of loans was transformed into binary values of positive and negative one-off instances using domain knowledge. Charged off, default and late payments were categorized in the later. Some other columns were also cleaned and processed for easy processing and to increase readability.

Credit Scorecard

The cleaned and processed dataset was split into training and test sets with an 80:20 ratio. On the training set, we used k-Fold testing to validate model performance. Weight of Evidence (WoE) and Information Value (IV) were subsequently used for feature engineering. The former was used to measure the predictive power of the variable in relation to the loan outcome. IV is used to rank a distinctive variable based on their comparative significance. The mathematical formulas for calculating these two parameters are as follows:

$$WoE = \ln\left(\frac{\% \ of \ positive \ outcomes}{\% \ of \ negative \ outcomes}\right)$$

$$IV = \sum (\% \ of \ positive \ outcomes - \% \ of \ negative \ outcomes) \times WoE$$

Following the standard practice [18], the IV of variables are interpreted as:

Table 1: Information Value Predictive Power	
Value	**Predictiveness**
< 0.02	None
0.02 – 0.1	Low
0.1 – 0.3	Medium
0.3 – 0.5	High
> 0.5	Suspiciously High

After feature engineering and transforming our data with WoE and IV, Logistic Regression was used with K-fold Cross Validation, maintaining the class imbalance.The untouched test set (20% of the original data set) is used for prediction and model benchmarking. The model is evaluated with ROC curve.

**Fig. 2: ROC Curve for the Logistic Regression**
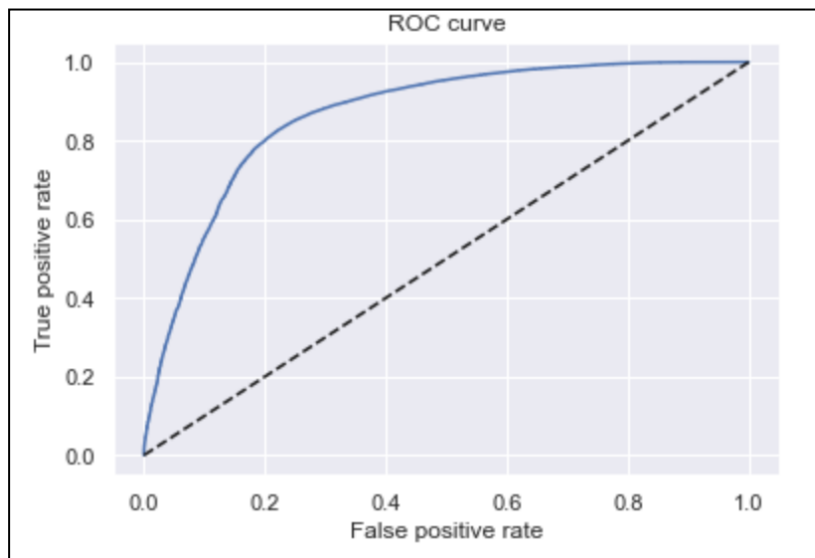
An AUROC score of 0.87 and a Gini score of 0.74 were achieved on the test set. Using this classification model and the FICO range of 300-850, the credit scorecard model was developed.

Smart Contract and Web Application

Making use of Solidity programming language for the Proof-of-Concept, the intended structure of the platform is laid down in the framework of the Smart Contract (www.github.com), which acts like the backbone of the complete platform. The overall structure of the Smart Contract to achieve the desired structure of the platform can be conceived as follows:

Fig. 3 Smart Contract Architecture
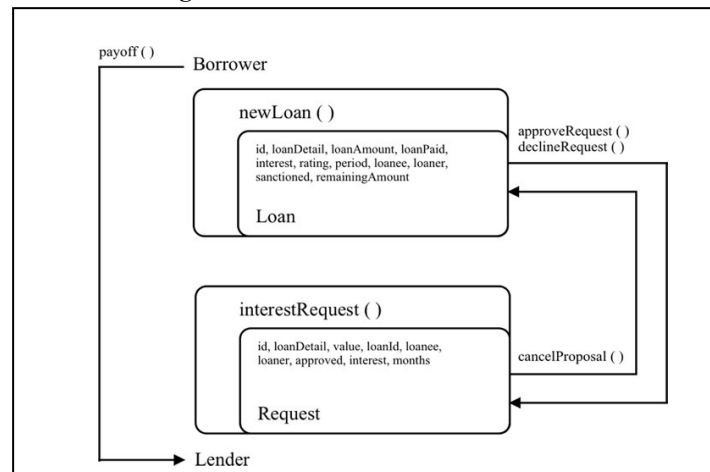
Lenders can browse all loan applications – sending their terms for the loan agreement. After submitting these proposals, creditors can withdraw them and claim their reimbursement till the time a debtor authorizes the application. After the validation from the debtor, the loan amount is relocated to the debtor's cryptocurrency wallet and an EMI scheme is started, notifying them about their total, and per-loan EMI, which are paid directly to the lender.

For the POC, a React based Web Application is developed which utilizes the Smart Contract, communicates with the Ethereum public blockchain and provides an interface for

the end user. The credit scorecard model, which was deployed as an independent API, is utilized in the Web Application using HTTP requests.

5. CONCLUSION AND DISCUSSION

Embracing the startup ecosystem and promoting SMEs is the need of the hour to strengthen the economy. Micro loans with a decentralized and transparent backbone architecture are ought to improve the financial conditions for those involved.

The accuracy attained by the Logistic Regression model, which was used for developing the credit scorecard, is acceptable. The Credit Risk model can therefore be treated as a means for judging loan candidates by the lenders. The said platform allows for collateral-free loan applications, making it a lucrative option for small scale businesses and startups. Blockchain-based implementation enables paperless and completely digital secure transactions.

Offering loans at a reduced risk with multiple lenders allows for adaptable reimbursements at negotiable loan costs with no trust issue between various moneylenders. Elimination of middlemen would not only give the transaction independence to two parties involved but also hasten the process. The transparency of the credit scoring model and the decentralized nature of loan applications and their approvals eliminate bias.

In addition to the peer-to-peer lending application, the credit scoring model can be used for various applications (Lyn, T., 2017) like Risk Modelling, Cluster analysis for credit control, etc. Transparency of transactions on the blockchain make way for Financial Analysts and Researchers to further study the correlation between user attributes and loan defaults to foster the Credit Scorecard model. The source code and generated datasets for the Credit Scorecard model and the POC Web Application are open-sourced (www.github.com).

6. REFERENCES

- Albulescu, C. T. (2021). COVID-19 and the United States financial markets' volatility. *Finance Research Letters*, *38*, 101699.
- Bachmann, A., Becker, A., Buerckner, D., Hilker, M., Kock, F., Lehmann, M., ... & Funk, B. (2011). Online peer-to-peer lending-a literature review. *Journal of Internet Banking and Commerce*, *16*(2), 1.
- Beauregard, R., Lopez, J. A., & Spiegel, M. M. (2020). Small Business Lending during COVID-19. *Small*, *2020*, 35.
- Crouhy, M., Galai, D., & Mark, R. (2000). A comparative analysis of current credit risk models. *Journal of Banking & Finance*, *24*(1-2), 59-117.
- Demyanyk, Y., & Kolliner, D. (2014). Peer-to-peer lending is poised to grow. *Economic Trends*.
- Dong, G., Lai, K. K., & Yen, J. (2010). Credit scorecard based on logistic regression with random coefficients. *Procedia Computer Science*, *1*(1), 2463-2468.
- George, N. (2017). *All Lending Club loan data* [Data set]. https://www.kaggle.com/wordsforthewise/lending-club/metadata
- Gogia, S. (2020). *QuickMoney.sol at main · simmsss/QuickMoney*.
  https://github.com/simmsss/QuickMoney
- Hasan, I., Politsidis, P., & Sharma, Z. (2020). Bank lending during the COVID-19 pandemic.
- Henriquez, R., Bittan, N., & Tulbassiyev, K. (2019). Blockchain and business model innovation: Designing a P2P mortgage lending system. *Netanel and Tulbassiyev, Kanat, Blockchain and Business Model Innovation: Designing a P2P Mortgage Lending System (April 14, 2019)*.
- Henriquez, R., Bittan, N., & Tulbassiyev, K. (2019). Blockchain and business model innovation: Designing a P2P mortgage lending system. *Netanel and Tulbassiyev,*

Kanat, Blockchain and Business Model Innovation: Designing a P2P Mortgage Lending System (April 14, 2019).

- Jefferson, D., Stephan, S., & Lance, Y. (2012). Trust and credit: The role of appearance in peer-to-peer lending. *Review of Financial Studies*, *25*(8), 2455-2483.
- Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, *34*(11), 2767-2787.
- Kumar, A. (2019). A Study of Significant Characteristics of E-payment Regime in India. *MERC Global's International Journal of Management*. Vol. 7, Special Issue 1, pp. 109-113.
- Kumar, A. and Brar, V. (2018). "Digital Marketing and Role of Blockchain in Digital Marketing Industry". *International Journal of All Research Education and Scientific Methods*. Vol. 6, Issue 12, December, pp. 23-26.
- Kumar, A., Gawande, A. and Brar, V. (2020). "Impact of Social Distancing on Marketing Communication". *Vidyabharati International Interdisciplinary Research Journal*. Vol. 11, Issue 02, December, pp. 267-270
- Sharma, U., & Manchanda, N. (2020, January). Predicting and Improving Entrepreneurial Competency in University Students using Machine Learning Algorithms. In *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 305-309). IEEE.
- Thomas, L., Crook, J., & Edelman, D. (2017). *Credit scoring and its applications*. Society for industrial and Applied Mathematics.
- Treacy, W. F., & Carey, M. (2000). Credit risk rating systems at large US banks. *Journal of Banking & Finance*, *24*(1-2), 167-201.
- West, D. (2000). Neural network credit scoring models. *Computers & operations research*, *27*(11-12), 1131-1152.
- Wood, G. (2014). Ethereum: A secure decentralised generalised transaction ledger. *Ethereum project yellow paper*, *151*(2014), 1-32.
- Zhang, D., Hu, M., & Ji, Q. (2020). Financial markets under the global pandemic of COVID-19. *Finance Research Letters*, *36*, 101528.