PolyFinTech 100 API Hackathon

Peer-To-Peer Lending Category

White Paper Submission for MatchMove

By:
Singapore Polytechnic
Team SafeCredit
(Diploma in Financial Informatics, School of Business)

Date of Submission:
10th July 2020

1. Introduction

1.1. Abstract

SafeCredit aims to provide a more streamlined experience for companies to generate credit scores based on alternative, non-traditional, data sources, and consequently, create a better UI/UX experience for loan applicants.

The software model will be sold as a B2B SaaS business product targeted at P2P lenders, although, theoretically, such a product can also be used not just by P2P lenders, but lenders or companies requiring credit scoring as a solution. The end goal of the software would be to be an invaluable part of the ecosystems of P2P lenders and companies needing credit scoring services.

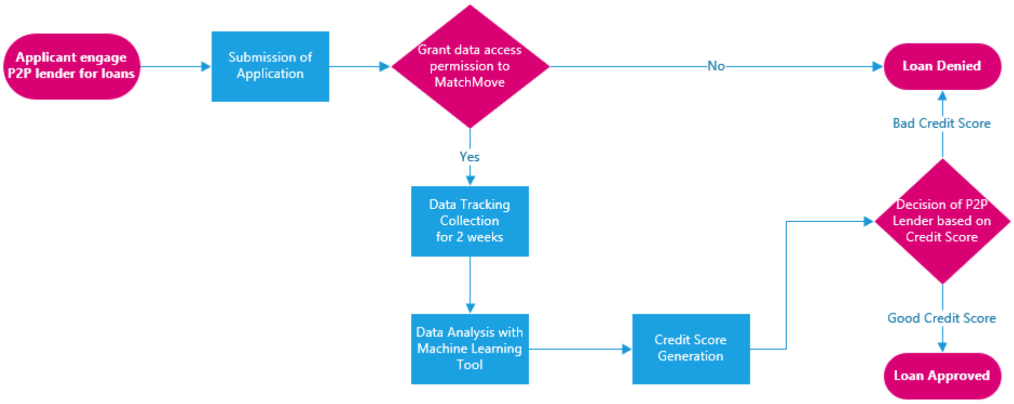
1.2. Motivation

About 8% of the Singapore working residents are freelancers. From taxi drivers to Youtubers, thousands of people live their lives without a monthly paycheck. However, without a regular payslip and CPF contributions, banks usually do not issue loans. As such, there are very limited ways to do alternative credit scoring without the traditional paychecks and employment letter. There are also other marginalized groups in society without access to traditional documents required by established financial institutions, such as the Unbanked, and foreign workers. In Singapore[1], approximately 38% of its population are underbanked, and 2% are unbanked (that is to say, 2% of Singaporeans do not have access to banking services).

Our proposal intends to create a model that can help to do up a credit scoring for individuals and businesses, especially SMEs, using alternative methods. We believe that technological advancements made in the past decade has opened up better, more accessible, ways to calculate credit scores. Our aim is not to provide better credit scores, but to ensure that everyone has access to fair and transparent credit scoring, regardless of his/her position in life.

2. Specifications

2.1. Business Process



Clients who are interested in gathering the credit scores of applicants (such as P2P lenders) will engage our services. From their existing ecosystems and software, clients will redirect applicants to an app to fill up their personal information as well as to verify their identity, doing so by uploading a selfie and uploading their official document's image (e.g. NRIC or passport photo). (Please refer to Figure 1) In addition, applicants have to grant data access permission in order to proceed (Please refer to Figure 2). This app will be monitored by MatchMove. Our software will seamlessly gather the information needed for credit

[1] Low, 2019. *4 in 10 Singaporean adults 'underbanked'; SE Asia ripe for picking by digital financial firms: Report* [online]. Available from: https://www.todayonline.com/singapore/4-10-singaporean-adults-underbanked-south-east-asia-ripe-picking-digital-financial-firms [Accessed 5 July 2020]

scoring without input from the client. After all the information has been gathered, we will generate a credit score and make it available to the client. Should our client desire to disclose the credit score available to the applicant, they would be able to do so.

2.2. Data Collection

After applicants have their identity verified, granted access and submitted their applications, data will be collected by MatchMove through Credit Bureau Singapore as part of traditional collection, and through various alternative collection such as tracking shopping, social media, browsing and telco data. The tracking of data will be done for a total of 2 weeks so that we can accurately collect and specify the profile and behaviour of the applicant to come up with a credit score.

2.3. Results

The credit score will be given to the P2P lender, generated by our machine learning model (see 3.4.1). From there, the P2P lender can decide whether to approve or deny the loan based on the credit scoring that has been tested with the model. The P2P lender can choose to disclose the credit score to the clients should they wish, but we do not advise doing so.

However, it is important that the credit scoring itself remains a secret to the applicant. Referencing Goodhart's Law ("*When a measure becomes a target, it ceases to be a good measure*")[2], we recognize that revealing how the credit scoring works will very likely result in tampering and score manipulation. For example, if one of our parameters is how often the applicant opens his/her banking app, the applicant could manipulate his/her credit score by repeatedly opening and closing his banking app. Therefore, we do not encourage clients and applicants to treat the credit score as a target, but rather simply as a measure.

3. Implementation

3.1. Starting Implementation/ Timeframe

The main challenge facing our system is the lack of testing and training data. Therefore, expanding our database with testing and training data will be our first, and most important, priority. The structure of our program is as such that the more it is utilized, the more accurate it becomes over time. At the same time, we will have to assure our early adopters that their investment in our system will not be in vain, that the contributions to our database will enable greater returns in the future for them.

For the first few months during our soft launch, we will work more closely with the clients to increase the pool of reliable data in the database, so as to improve the accuracy of the model. As the model progressively morphs into its desired outcome, we will eventually be less dependent on early adopters. To encourage cooperation with early adopters, especially in our alpha and beta stages, we will not be looking to extract revenue from them. As our machine learning capabilities grow over time, we can start to be less dependent on our early adopters for data.

3.2. Revenue Model

Here, we will only aim to use this revenue model once our software reaches the desired quality, possibly after a stable release. However, due to the low capital requirements in the early stages of the software, engaging in loss leader pricing is an increasingly viable option in order to see deeper market penetration.

The software will be monetized in two ways: In a pay-per-call model and a package model. We recognize that SMEs will require greater flexibility when it comes to cost, and thus, a pay-per-call model would be more attractive to such firms. On the other hand, larger corporations with expanded access to venture capital might require a fixed pricing model that eliminates uncertainty, especially when it comes to the

[2] Manheim, David; Garrabrant, Scott (2018). "Categorizing Variants of Goodhart's Law". arXiv:1803.04585

larger volume of calls they might utilize. The difference between these models is that one acts as a variable cost, whilst the other acts as a fixed cost.

Therefore, a tiered pay-per-call model will be used to attract SMEs. The first tier of 100 will be provided free of charge to entice new subscribers, as well as to expand our subsequent testing data in our data model. In subsequent tiers of approximately 500 calls, we will be charging a fixed amount to be decided in the future, adjusted per the demand for such calls. The price and size of tiers will be tweaked over time depending on the frequency and demand for such calls, to ensure it remains attractive to clients.

Above our pay-per-call model, we will also be providing an all-encompassing price for unlimited calls in a month. This price would entice firms that require vast amount of calls in a month, reducing the price for each call for them. As above, this price will be determined by the demand of such calls.

3.3. Capital Requirements

We do not anticipate high capital requirements exceeding 4 digits. The bulk of the original capital requirement will stem from paying for APIs and services to collect data and to build up our database. Further costs will be accrued from hosting a RapidMiner and SQL server in the cloud, possibly using services like AWS or Microsoft Azure. A single database instance of SQL web hosting in Singapore using Amazon’s RDS platform costs $0.062-$0.416 per hour.

As the business grows, we would expect marketing costs to grow the userbase, as well as potential personnel costs should we desire to expand or refine our system. As the needs of the software expands, we would expect burn rate to increase.

3.4. System Architecture

The system will be designed in 3-Tier architecture which consists of the Presentation layer, the Business layer, and the Data Access layer. (Please refer to Figure 3)

3.4.1 Machine Learning Process (Modelling)

To come up with a credit scoring for the applicant, a machine learning model will be used. Past and existing borrowers’ data will be used to train the model to predict a score for new applicants.

We aim to train at least 4 models to ensure an acceptable level of results:

1) Decision Trees
2) Random Forest
3) Neural Networks
4) Deep Learning

Models will be evaluated via their accuracy, which can be seen in the lift charts, ROC, and confusion matrix. Such data models will be generated and held in a machine learning server (i.e. RapidMiner Server) in the cloud to be continuously updated with new data. As the system expands with new data, models can be created and migrated using python and libraries like PyTorch and Google’s TensorFlow that give us greater flexibility and control over the specifics of the models. As part of this process, we have to account for the eventuality of Type 1 and Type 2 errors occurring.

As we recognize that predictive trends will change over time, we will aim to continuously update and evaluate our models to ensure that the models remain relevant. To that end, we will continuously update the database with training data gathered from our clients, thus creating a feedback loop in which the accuracy of the models will be improved over time. After the new applicant’s data has been fit into the model, a score will be shown and then given to the P2P lender.

Appendix

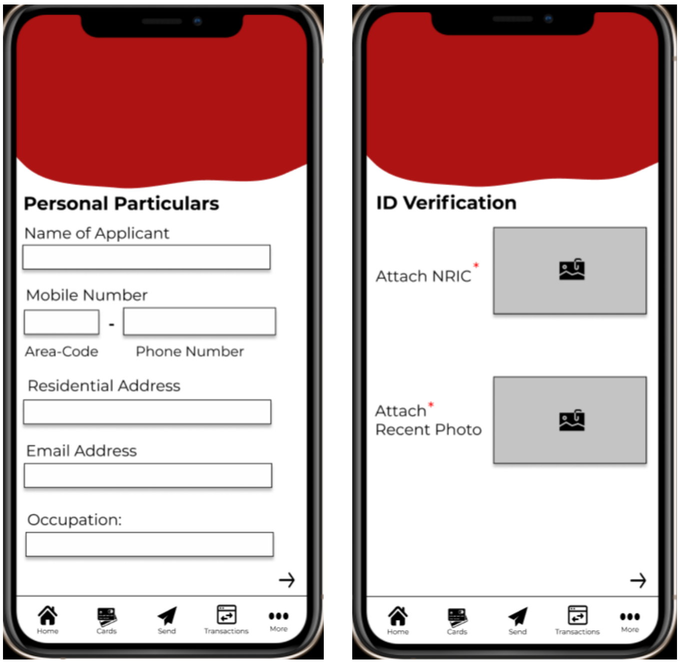


Figure 1: Applicant's Personal Particulars UI

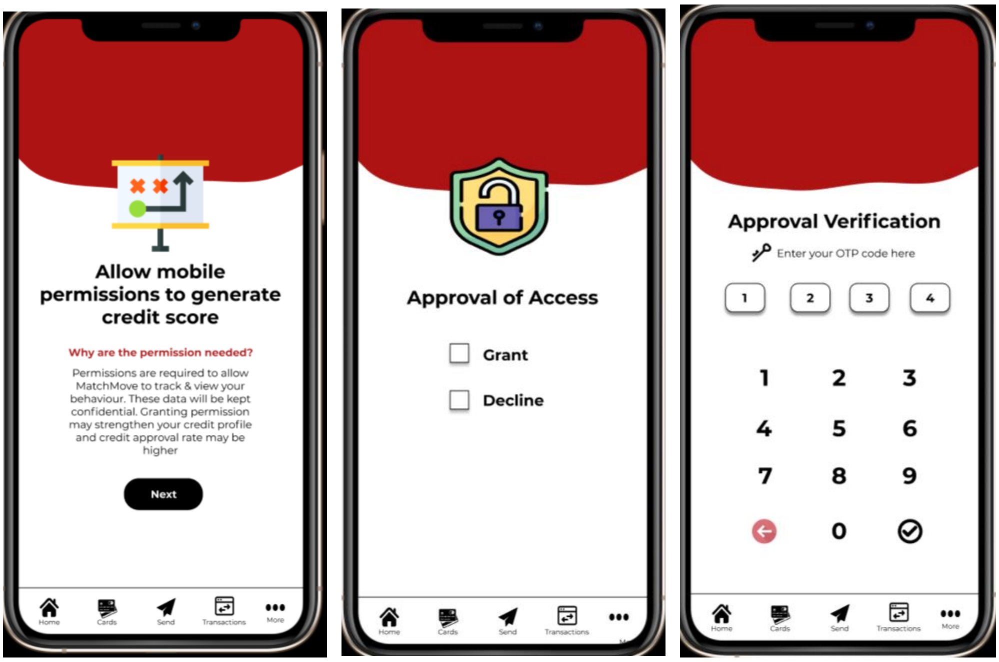


Figure 2: Data Access UI

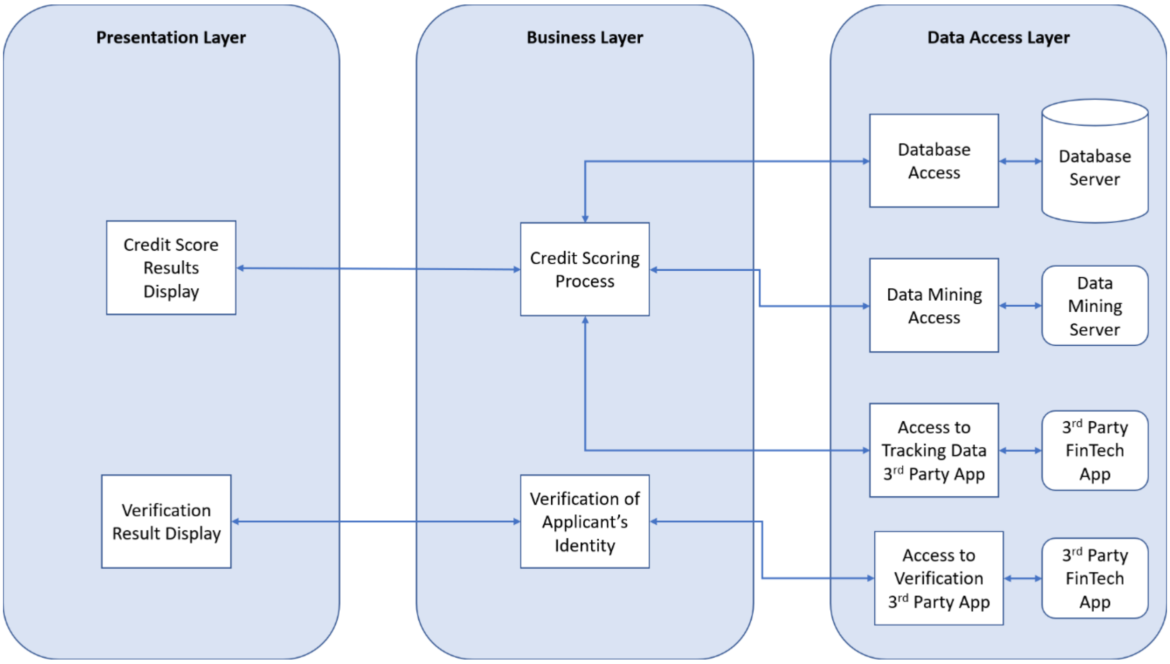


Figure 3: 3-Tier Software Architecture Diagram of Credit Scoring System