Trustworthiness Validation for Merchant Onboarding

9-Mar-2019 | Version 3.0

—

Team Name: ai.Orz

Dai Yirui, Dong Meirong, Gu Lijian, Guo Feng, Wong Yoke Keong, Zhang Le

Prepared for NUS-ISS Master of Technology in Intelligent Systems ISY5001 Machine Reasoning Group Project

Table of Contents

1.0 Executive Summary

The aim of our project is to deliver a light-weight, low-cost but highly efficient background screening web service for small-and-medium enterprises (SME) to facilitate fast, frictionless and digital-centric onboarding experiences for their users. In this project, we employ machine reasoning methods like rule-based reasoning and certainty factors to achieve our objectives and accelerate screening processes.

2.0 Business Value

In this section, we elaborate what background screening is and its business value.

2.1 Why background screening is required?

Background Screening is a process commonly employed by businesses, especially financial institutions, to identify the background of their clients and assess any potential risks of illegal intentions for their business relationships. Its main objectives are to detect corruptions, terrorist financings and money launderings and combat illegal transactions.

The process plays a crucial role to the success of a business by lowering its operational risk and boosting its credit reliability. However, conducting the process via traditional manual workflows may create adverse effects. Because time is money in today's world, only businesses that can deliver most efficient and rapid services can gain a more competitive edge and build bigger market share.

Therefore, transforming background screening to a highly-accurate, automated workflow that requires minimum human interventions is only way to ensure clients' backgrounds are well-inspected while seamless user experience is maintained, and business success is achieved.

2.2 Why offering a low-cost service?

Although digitised background screening is important in driving business success, building a comprehensive IT system to accurately verify clients' backgrounds and predict risk scores can be costly.

For example, in 2017, the Monetary Authority of Singapore (MAS) pinpointed background screening as one of the biggest pain points in financial industry and proposed a banking Know Your Customer (KYC) Shared-Service Utility to streamline end-to-end background screening.

However, in 2018, the project was quickly paused due to unexpected high cost, quoted the reason that the proposed solution was going to cost more than the savings that businesses would be going to get out of it.

For big organisations, this decision should not cause very detrimental impacts because they can re-use their original screening models and systems to assess risks or pump in more resources to explore new alternatives. But for SMEs, this cancellation can be quite disappointing because they are in greater demands for a digital platform to improve their screening efficiency, and because they are more susceptible to risks but have limited resources to invest on background screening, let alone full-fledge automated background screenings.

On top of that, although the market offers many platforms for screening, most of them are expensive especially when charged on per report basis. Based on our research, the market rate per report is around USD 39.95. The amount can be a heavy burden to SMEs which require frequent screenings.

To address this problem, we create a light-weighted and scalable screening web service based on publicly available information (e.g. client identification number, address, tax number, etc.) to offer low-cost and reliable preliminary background screenings.

2.3 More about our screening web service

Based on our screening results, SMEs can effortlessly detect high risk clients (e.g. clients with invalid tax number, clients with criminal records) with no financial stress.

If more details are required later, we can always interface user information entered earlier via Application Programming Interface (API) to next stage to obtain more thorough reports with minimum manual entries again.

This powerful web service can be easily modified to suit many applications (e.g. vendor background verification, merchant onboarding screening, personal loan screening). In this report, we demonstrate the use of this web service via a merchant onboarding screening system.

2.4 Merchant Onboarding Screening System

2.4.1 What is Merchant Onboarding?

Merchant onboarding is a credibility review process implemented by banks to ensure incoming merchants can meet their minimum onboarding requirements to use their gateways to accept credit and debit cards.

After onboarding, whenever a customer purchases an item with a credit or debit card, the merchant submits the purchase transaction information to its acquiring bank, which will then submit it through the card association network to the card holder's issuing bank. The issuing bank will approve or decline the charge and bill the cardholder the amount due to the merchant accordingly.

3.0 Knowledge Specification

In this section, we specify how we collect and represent requirements.

3.1 Knowledge Representation

Our team used these approaches to gather requirements for system designs:

- Interviewed a domain expert (please refer to appendix A) to better understand the screening processes;
- Analysed relevant academic researches and industrial reports to better study the risk evaluation processes;
- Retrieved industrial standard risk estimates for system developments. These estimates were derived from supervised machine learnings driven by production data. To protect business information, we will not reveal the sources of these estimates in this report.

After requirements gathering, we carefully filtered the information collected and extracted only meaning results for knowledge representation.

3.2 Knowledge Representation

At this stage, requirements are represented in different formats to:

- abstract insights out of the interview scripts from domain experts;
- convert the insights into facts and rules and represented by Knowledge is Everything (KIE) guided data objects, forms, guided rules and decision tables.

In table 1 below, we summarise our requirements by categories.

S/N	Areas	Insights of areas	Knowledge Acquisition Technique
1	Financial	To indicate a merchant's financial status such as: • Net Income to show profitability • Total Deposit to show financial stability	Elicitation of tacit knowledge through the conduct of interview with domain experts
2	Operational	To indicate a merchant's operational status such as: • Unique Entity Number (UEN) to check registration status (cross-check with information from ACRA) • Employee size to show operational scale. • Founding Year to check operational history • Registration location to check and minimise operational risk such as potential money laundering	Web service call to obtain publicly available/documented information from organisations like the Accounting and Corporate Regulatory Authority (ACRA)
3	Management	To show a merchant's Management status such as: • CEO IC will be used to do compliance check	Elicitation of tacit knowledge through the conduct of interview with compliance officers

Table 1: Requirements Summary

In business flow diagram 1, we show the decision rules used to determine if a merchant fulfills the onboarding criteria.

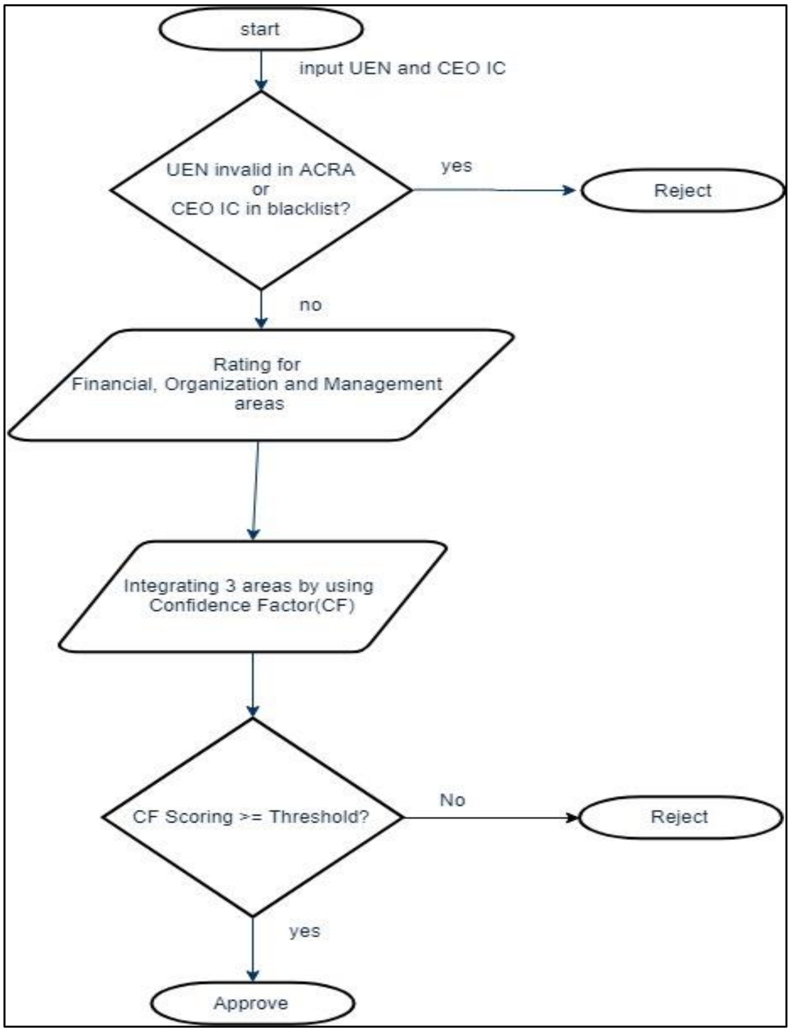

Figure 1: Business Flow

There are a few important “one vote veto” rules in this business flow. The rules are processed at the beginning to simplify the overall business processes. For example, the UEN verification, which is executed immediately after Python web front receives an input. Applicants who fail this verification will be rejected straight away to reduce the need for further manual inputs.

After an application passes through the first stage, KIE guided rules and decision tables will be triggered. These rule engines, each carries a confident factor value, are activated to calculate the final risk score.

4.0 Solution

Our solution composes of a front-end user interface (UI) and a back-end rule-based assessment engine which are developed in:

- Python for web platform and;
- iBPM for rule engines.

The web platform performs some basic validation on user inputs, whereas the backend rule engine fires certain rules when premises are met. Furthermore, the system interfaces with RESTful API calls to call or transfer data to meet overall business objectives and bring the best user experience to our clients.

4.1 System Architecture

The system architecture diagram shown below illustrates how the front-end web platform has been integrated with the back-end system and configurable rule base.

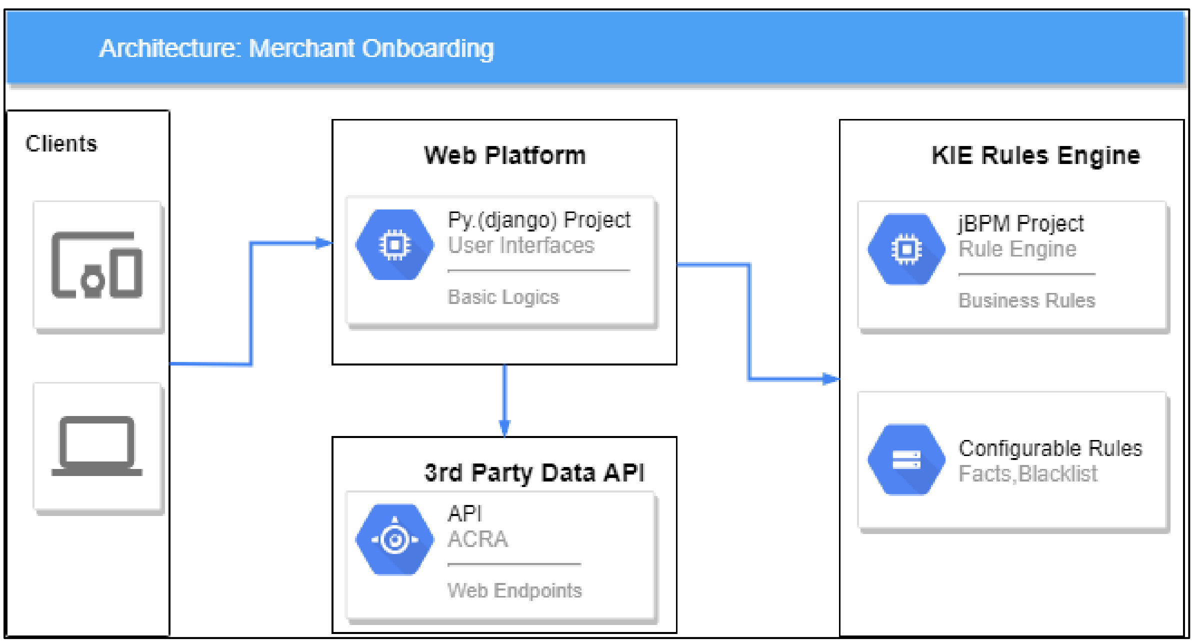

Figure 2: System Architecture

4.2 Project Scope and Assumptions

While users have the freedom to configure data sources per their own preferences in our system, these factors must be taken into consideration:

- ACRA information;
- Configured CEO IC blacklist;
- Financial Action Task Force (FATF) list of high-risk country.

Furthermore, due to time and project scope constraints, we use only Singapore as an example to demonstrate some basic functionalities in this project.

The project scope is restricted by geographical areas because:

- most of the benchmarks used during the assessment process, particularly UEN, vary from country to country.
- certainty factors including the financial certainty factor and the operational certainty factor are obtained through machine learning process that takes only regional raw data and domain standards. The factors may not be applicable in other countries.

4.4 System Features

Despite many limitations, our team has developed a very practical solution leveraging on different technologies, which can substantially add value to our users, mainly SMEs.

4.4.1 System Intelligence

- Most of our assessment criteria are configurable. Additionally, our system is able to accept the most recent and relevant benchmarks calculated from machine learning process. This means our system is able to cope with the latest industrial standards. This is very critical to the success of our product since the determining factors of credibility assessments change constantly.

- Our system is able to assess a merchant from many different perspectives, for example, financial health, organization status and credibility. Each is supported by observable or inferable data as shown below.

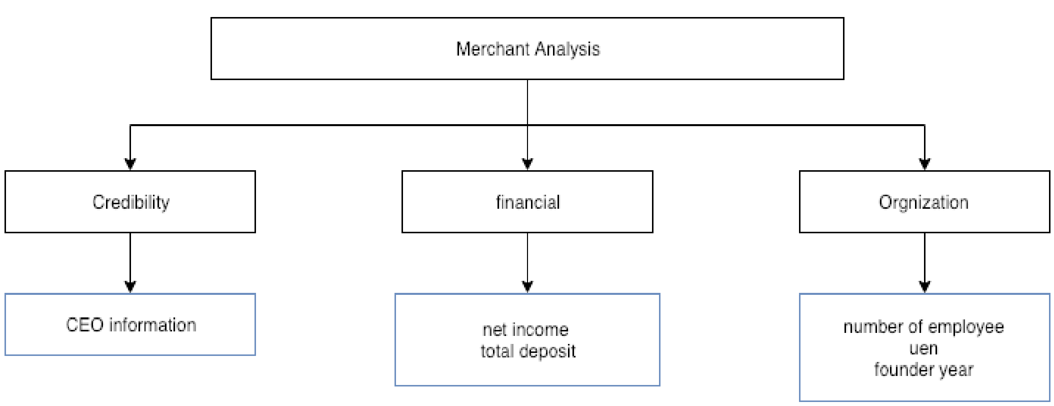

Figure 3:Merchant Analysis Components

4.4.2 Back-end Reasoning

Background screenings, or more specifically, risk estimations, have many uncertainties (e.g. the net income can vary each month). These uncertainties cannot be represented by a simple “yes” or “no” logic.

Since boolean approaches do not address uncertainties, we decided to integrate a heuristic method into our system for the quantification of uncertainties.

Out of all the methods (e.g. Fuzzy Logic, Bayesian Approach) available, we chose certainty factor theory for below reasons:

- It is formalized. Certainty factors theory was designed in an attempt to formalise heuristic approaches to reasoning which provides a simple explanation to how conclusion is reached.
- It is experience-based. This method encourages the use of risk estimates gauged and adjusted by domain experts based on their professional experiences for assessments.
- It is highly relevant in our scenario. Because we have one hypothesis (e.g. the merchant is credible) in our mind already, we just need a degree of confirmation (or disconfirmation) of the hypothesis with concrete evidences.

How do we calculate risk estimates?

1. Set prior probability of hypothesis *H* being true (assuming x samples/ total samples for a particular condition will succeed)

2. Find evidence *E* to support probability of hypothesis (calculate x samples/total samples for a particular condition will succeed with actual data)

3. Calculate the probability *P(H|E)* that hypothesis *H* is true given the evidence *E*

4. Apply the models below to calculate risk estimates

$$MB(H, E) = \begin{cases} 1, & if\ P(H) = 1 \\ \dfrac{\max\{P(H \mid E), P(H)\} - P(H)}{1 - P(H)}, & if\ P(H) < 1 \end{cases}$$

$$MD(H, E) = \begin{cases} 1, & if\ P(H) = 0 \\ \dfrac{\min\{P(H \mid E), P(H)\} - P(H)}{0 - P(H)}, & if\ P(H) > 0 \end{cases}$$

How do we calculate final risk score?

We have three main conditions (e.g. financial, operational and management) contributing to the final risk score, supporting the hypothesis that merchant is credible.

Under each condition, if the sub-conditions have OR relationship. Then, the maximum certainty factor is chosen as the final certainty factor for that condition.

$$\mathrm{CF(H)} = \max\{\mathrm{CF(E1)}, \mathrm{CF(E2)}, \ldots, \mathrm{CF(En)}\} * \mathrm{CF(Rule)}$$

If the sub-conditions have AND relationship. Then, the minimum certainty factor is chosen as the final certainty factor for that condition.

$$CF(H) = \begin{cases} \min\{CF(E1\}, CF(E2), \ldots, CF(En)\} * CF(Rule), & if\ \ CF(Ei) > 0,\ i = 1, 2, \ldots, n \\ 0, & otherwise \end{cases}$$

Finally, we apply the formula below to combine these certainty factors CF (condition) into one score.

$$CF(H, E1 \wedge E2) = \begin{cases} CF(E1) + CF(E2)(1 - CF(E1)), & if\ \ CF(E1), CF(E2) > 0 \\ CF(E1) + CF(E2)(1 + CF(E1)), & if\ \ CF(E1), CF(E2) < 0 \\ \dfrac{CF(E1) + CF(E2)}{1 - \min\{|CF(E1)|, |CF(E2)|}, & if\ \ sign(CF(E1)) \neq sign(CF(E2)) \end{cases}$$

4.4.3 Certainty Factors Implementation

Financial Certainty Factor Table

CF(financial) = Financial Certainty Factor

Financial Certainty Factor	Net Income				
Total Deposit	Infi-1000000	1000000-500000	500000-0	0-(-500000)	(-500000)-Infi
Infi-3000000	0.9	0.8	0.7	0.2	-0.2
3000000-1000000	0.8	0.6	0.4	-0.2	-0.5
1000000-0	0.6	0.4	0.2	-0.5	-0.8

Table 2:Financial Certainty Factor

Operational Certainty Factor Table

CF(operational) = Operational Certainty Factor

If the registration location is founded in the list of high risk country then CF(operational) = -1.0
Otherwise, we will calculate certainty based on the table below.

Number of Employees / Founding year	Infi-10000	10000-1000	1000-0
Before 2000	0.8	0.6	0.4
2000-2019	0.6	0.4	0.2

Table 3:Operational Certainty Factor

Credibility Certainty Factor

CF(credibility) = Credibility Certainty Factor
If ceo_ic in found in blacklist then CF(credibility) = -1.0

Combined Final Certainty Factor

```
if cf_credibility == -1.0 or cf_operational == -1.0:
        final_cf = -1.0
elif cf_operational + cf_financial == 0.0:
        final_cf = 0.0
elif cf_financial < 0.0 and cf_operational < 0.0:
        final_cf = cf_financial + cf_operational + (cf_financial * cf_operational)
elif cf_financial >= 0.0 and cf_operational >= 0.0:
        final_cf = cf_financial + cf_operational - (cf_financial * cf_operational)
elif cf_financial * cf_operational < 0.0 and abs(cf_financial) > abs(cf_operational):
        final_cf = (cf_financial + cf_operational) / (1 - abs(cf_operational))
elif cf_financial * cf_operational < 0.0 and abs(cf_financial) < abs(cf_operational):
        final_cf = (cf_financial + cf_operational) / (1 - abs(cf_financial))
```

Interpretation of Results

The scale below maps risk scores to degree of certainty. A score of 1.0 shows that onboarding is strongly recommended while -1.0 shows strongly not recommended.

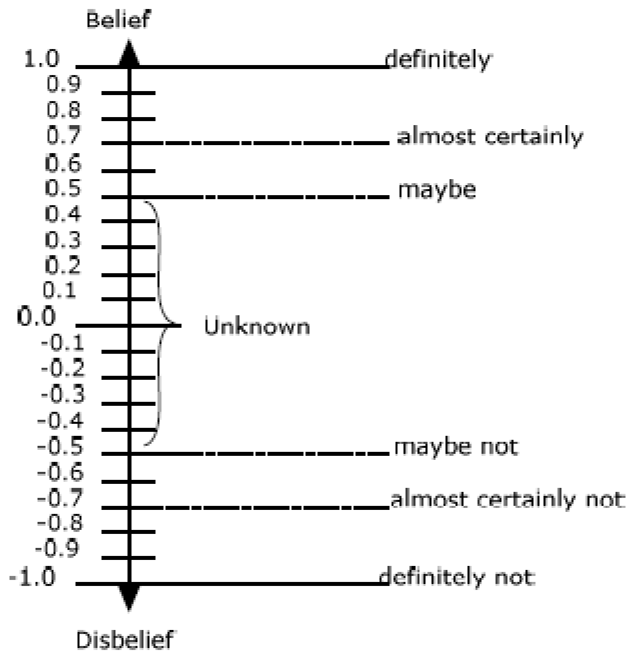

Figure 4:Risk Score Scale

4.4.4 Ease of Assess

The system front-end is a web platform. Running the front-end requires only one command line. User all over Singapore will be able to access this service from home or office, instead of queueing in front of bank.

4.4.5 Security

After properly deployed, the front-end and back-end component works as blackbox. Although front user from internet can access to web platform but the calculating and assessing process is completely blind to them. The RESTful API communication between front-end and back-end systems are secured by basic authentication.

4.4.6 Scalability

The system is developed to cater for users from different regions.Users have the authority to configure data source and assessment benchmark base on their own needs.

4.5 Limitations

The entire decision process in suggesting a merchant for band whether it is health or not. Despite the information from government website as well as the consultation with the subject matter expert, it is inevitable that the system will fail to precisely assess every single merchant from each aspect. This group of information assert on intangible factors such as company image, customer loyalty and management efficiency. While these factors were not part of the decision matrix, it is certainly possible to study and incorporate them during further enhancement of the system to provide a more comprehensive result.

Successfully onboarded or rejected merchant data are usually very sensitive and has been kept confidentially by system users such as banks. To provide the most comprehensive assessment, the system requires precise assessing criteria which are generated from machine learning process. The preciseness of the assessment result is directly related to the data availability and pool size.

4.6 Future Enhancements

1) **Integration with publicly available data**
   At current state, the system integrates with ACRA data because ACRA provides free request API to public. If more funding provided, there are actually various channels of data sources we can refer to.

2) **Enhancement with Machine Learning Models**
   At current state, the system uses certainty factors that have been provided by domain expert. Then our team manually configured those parameter into the system. To further enhance the system efficiency, a dynamic integration between rule engine and machine learning system is required.

3) **Validation-as-a-Service**
   The system has the potentiality to run on cloud and provide even wider accessibility to public. Because not only big organisations rely on validation services for assessments, many SMEs also require similar type of services. Furthermore, we can expand our business to meet requirements of clients in other industries.

4) **Board member characteristics by fuzzy logic**
   Give sufficient time, we can add the analytics of board characteristics to our system to improve results accuracy by adapting and applying Rim Boussaada’s paper published in European Central Bank’s economic bulletin in 2018.

   The paper applies fuzzy logic to study how different combinations of governance mechanisms influence credit risks.

5.0 Conclusion

In this report we have provided an overview of how machine reasoning can automate the background screening process by implementing business rules in a Business Process Management System like jBPM. We also demonstrated the value of the web service solution in the context of a Merchant Onboarding Use Case that is very common in the Financial Services industry.

In future enhancements we hope to integrate with more data sources and work with APIs provided by data exchanges. We also hope to enhance the system with Machine Learning Models, thereby making the solution even more robust and resilient, with a Validation-as-a-Service business model as a future state.

6.0 Bibliography

Data.gov.sg (2019, March 9). ACRA Information on Corporate Entities: https://data.gov.sg/dataset/acra-information-on-corporate-entities

Bizfile, Business Filing Portal Of ACRA (2019, March 9). Business Profile: https://www.tis.bizfile.gov.sg/ngbtisinternet/faces/oracle/webcenter/portalapp/pages/staticpages/BusinessProfile_Static.jspx

Financial Action Task Force (2019, March 9). High-risk and other monitored jurisdictions: http://www.fatf-gafi.org/countries/#high-risk

Experian (2019, March 9). Business Credit Report: https://www.experian.com/small-business/business-credit-reports.jsp

Red Hat, JBOSS jBPM (2019, March 9). jBPM Documentation: https://docs.jboss.org/jbpm/release/7.12.0.Final/jbpm-docs/html_single/

Boussaada, R. (2018). Board Characteristics and MENA Banks' Credit Risk: A fuzzy-set. *Economics Bulletin*.

DBS. (2019, March 03). *Credit Card Acceptance for Retail Businesses.* From DBS : https://www.dbs.com.sg/sme/day-to-day/collections/credit-card-acceptance/for-retail-businesses

Grosan, C., & Abraham, A. (2011). *Intelligent System: A Modern Approach.* Springer-Verlag Berlin Heidelberg.

Menon, R. (2017, November). *Singapore FinTech Journey 2.0.* From Monetary Authority of Singapore: http://www.mas.gov.sg/News-and-Publications/Speeches-and-Monetary-Policy-Statements/Speeches/2017/Singapore-FinTech-Journey-2.aspx

Straits Times. (2018, November 12). *MAS to Shelve 'know-your customer' Project Due to High Costs, Work on SME Innovation Platform: Ravi Menon.* From Straits Times: https://www.straitstimes.com/business/banking/mas-to-shelve-know-your-customer-utility-project-due-to-unexpected-high-costs-ravi

UOB. (2019, March 3). *Merchant Services.* From UOB: https://www.uob.com.sg/personal/cards/merchant-services.page

Visa. (2017). *Effective Merchant Onboarding to Drive Portfolio Growth.* From Visa: https://usa.visa.com/dam/VCOM/global/partner-with-us/documents/effective-merchant-onboarding-to-drive-portfolio-growth.pdf

7.0 Appendix A - Interview

We have interviewed one account manager (about 5 years of experience) from financial sector to to better understand the requirements.

7.1 Context

As shared earlier, we are collecting the information to build a mini risk prediction system for organisations to assess the credibility of their clients.

One sample use case:

1. A Bank wants to check merchants' credibility before their onboarding
2. Merchants to enter application information
3. The bank can use our system to conduct a first level screening based on the available information
4. Our system will first eliminate applications that do not meet basic requirements (e.g. UEN not valid)
5. Our system will proceed to calculate the risk (e.g. <0.5 = low risk; >0.5 = high risk) to accept a particular application
6. Bank to make informed decision based on the assessment results provided

7.2 Transcript

Q1: What are the key metrics used for credit scoring / client credibility assessment in your organisation?
A1: It depends very much on your business nature and clients' industry.
In general, we first assess financial stability (e.g. debt to equity ratio, burn rate, credit history, percentage of deferred revenue and market outlook).
Then, factors like company reputation, industry, longevity, stakeholders' credit worthiness and billing method (e.g.prepaid or postpaid) will also be taken into considerations.

Q2: How would you prioritize/rank these metrics?
A2: The main purpose of us conducting background screenings is to detect potential fraud cases. So based on my personal experience, I will rank financial factors first.
Then maybe industry, reputation, longevity, size, location and other factors.

Q3: Do you have in-house tools(e.g. formula, website) to calculate clients' credit scores?
A3: Yes. But it is industry and company specific. I am not convenient to share the details here. You may read some papers/research online to enhance your understanding, Some of them provide very insightful information.

Q4: Do you engage external services to assess clients' credibility?
A4: Yes. We cross-check for accuracy to lower our operational risks. But as shared earlier, it is not cheap to have these checks.

So for SMEs (which is your intended users), as far as I know, most of them do not do background screenings unless really required. Because returns cannot outweigh inputs… I guess your idea of providing low-cost screenings will be popular among them (provided your screening results are meaningful…).

Q5: Where do you retrieve clients' information (e.g revenue, board of directors, etc) from?
A5: We usually ask clients to provide the information themselves. Then we validate. There are quite a number of platforms in Singapore for you to do validations. The common ones will be ACRA, MAS,Bizfile and credit bureau.

Q6: Kindly help to look through the workflow attached & share how can we improve the workflow to increase the accuracy of our predictions?
A6: Step 1, you may use UEN (Unique Entity Number) for verification. UEN is a standard identification number for entities to interact with government agencies.

7.3 Post Interview Follow-Up

There are some clarifications from the expert on Machine Reasoning and rules used and they have been addressed below.

Clarification 1: What does CF mean?
Response 1: CF is certainty factor. We use it to calculate the level of confidence we have about our results.

Clarification 2: What do you want to check with CEO IC?
Response 2: CEO name and IC can be used to check if the management is involved in any criminal cases or bankruptcy cases.