BANK INDONESIA

Sentiment Analysis of User's Reviews on Non-Bank Payment Service Apps[1]

N. Heru Praptono[2], Arinda D. Okfantia, M.Khoyrul Hidayat, M. Hafiruddin

Statistics Department

Basel, Switzerland
August 2022

[1] The expressed views belong to the authors and **do not** necessarily reflect the institution.

[2] Corresponding email: nursidik_hp@bi.go.id/pra.heru@gmail.com

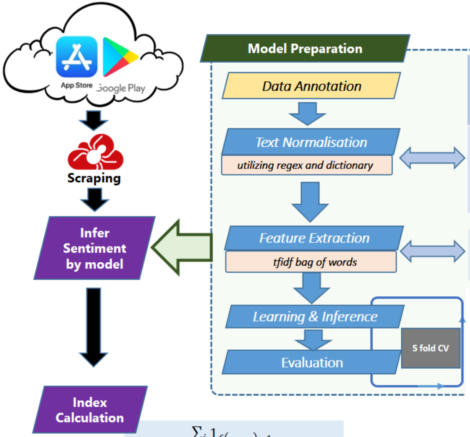
BANK INDONESIA

Outline

Background

- More banking and payment activities are currently being conducted into mobile application platform than the traditional ones, legally provided by the authorised entities either bank or non-bank institutions.
- In Indonesia for example,
  - By the end of 2021, there have been at least 41 non-bank payment service providers (PSPs) who take a part in such banking and payment services.
  - The growing amount of non-bank PSP customers has found to be increasing recently, moreover since the pandemic of COVID-19.
  - Have infiltrated on nearly the whole of about 270 million people of the Indonesian population $\rightarrow$ play significant role in promoting and accelerating the economic growth.
  - Need to be supervised and monitored by the policy maker (i.e. Bank Indonesia) to anticipate any systemic risk.
- Measuring and monitoring the quality of non-bank payment service apps is difficult – but possible by considering user's review. Survey is costly, solution: apps review (i.e. in Google Play Store, Apps Store) as a proxy.
- Related works on mobile apps user's review analysis, e.g. Vu et al 2015 (mobile apps), Leem & Eum 2021 (m-banking).
- Some remaining issues (including but not limited to): non-formal text, limited training data, imbalanced training data, further utilisation for monitoring.

Overall Methodology



Unnormalised Text (raw text)	Normalised Text (cleaned text)
Aplikasi yg Bagus.. Tampilannya menarik..	aplikasi yang bagus tampilannya menarik *(in English: nice application, yet interesting user interface)*

$$tfidf_{w,d} = tf_{w,d} \ x \ \log \frac{N}{df_w}$$

1. **Security**: Represents the security of mobile apps. some issue related to the problem on security aspects including log in/log out, registration, security of e-money.
2. **Feature**: Represents the variability and functionality of the feature offered by non-bank PSP's
3. **Design**: Represents how well the design related to the intuitiveness and user experiences.
4. **Cust. Services**: Represents how responsible, reliable, and effectiveness of the customer services.
5. **Technology**: Represents how advanced the technology the non-bank PSP's is, and represent how well the application is from the technological issues.

e.g.: Bach et al, 2020, Leem & Eum 2021

$$\text{idx}_{o,t} = \frac{\sum_i 1_{f(x_{o,i,t})=1}}{N_{o,t}} x \log N_{o,t}$$

Annotated Data and Model Evaluation

Annotated data are relatively imbalanced → utilise SMOTE (Chawla et al, 2002) during learning process.
Some experimented models → rule based model, SVM (with RBF Kernel), Decision Tree (Gini Splitting Criteria), and Logistic Regression.

Annotated Data

Annotated Dataset — Table 1

	Negative	Positive or Not Related	Total
Security Aspect	449	821	1270
Feature Aspect	420	850	1270
Design Aspect	306	964	1270
Customer Service Aspect	390	880	1270
Technology Aspect	427	843	1270

[1] Annotation is performed by 3 annotators.

Model Evaluation

Performance matrix of models from data (F1 in %) — Table 4

	Security Aspect	Feature Aspect	Design Aspect	Customer Service Aspect	Technology Aspect
Rule Based[1]	56,49	56,14	55,43	53,78	55,85
SVM	76,14	74,67	66,67	66,55	*65,77*
Decision Tree	72,00	61,90	63,37	59,36	63,69
Logistic Regression	*78,17*	*74,80*	*69,32*	*79,07*	65,61

[1]Rule based: $r(S) = \exists(\mathrm{kw}_{\mathrm{aspect}})$ in S $\wedge \exists(\mathrm{kw}_{\mathrm{negative\_sentiment}})$ in S

Incorporating Prior Knowledge

We adopt the methodology incorporating prior knowledge described by Schapire et al, 2002.

- **Prior Knowledge:** $r(S_x)$ (essentially the result of rule based model)
- **Learning**: Given the training dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$, and $y \in \{0, 1\}$. The objective function is to minimise negative log likelihood, controlled by the prior information.

$$J = \sum_i [\ln(1 + e^{(-(2y_i - 1)f(x_i))}) + \underbrace{\eta D_{\mathrm{KL}}(\pi(x_i)||\sigma(f(x_i)))}_{\text{control on prior information}}]$$

where f is linear function, σ is the logistic function. Here π is our "prior information" quantification, defined by:

$$\pi(x) = p(y = 1|x) = \begin{cases} 0.9; & \text{if } r(S_x) = 1 \text{ (true)} \\ 0.1; & \text{otherwise} \end{cases}$$

- **Inference**:

$$p(y = 1|x) = \sigma(f^*(x)); f^* = f + h_0$$

In this case, h_0 is the "prior term" defined as the inverse of logistic function of $\pi(x)$, that is $h_0(x) = \sigma^{-1}(\pi(x)) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right)$

Incorporating Prior Knowledge (Cont'd)

Experimental Result

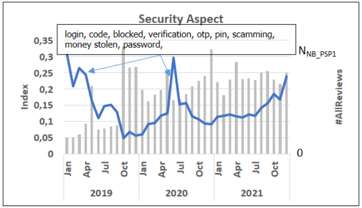
- Initially, when the number of training data is small, the Logistic Regression gives result the lowest F1 score.
- Adding prior knowledge at this stage helps increase the inference process, although the performance is still below the rule based model.
- As the size of training data grows, the performance of both two models (model Logistic Regression and Logistic Regression with Prior Knowledge) increases.
- The Logistic Regression with Prior Knowledge however gives the better performance at any training size over the Logistic Regression model.
- The rule based model shows constant score as this is a deterministic function.
- Once the training data is sufficient, the Logistic Regression with Prior Knowledge achieves highest score.
- We then leverage the trained Logistic Regression with prior knowledge to be applied into full data so that the index can further be constructed based on the inferred results.

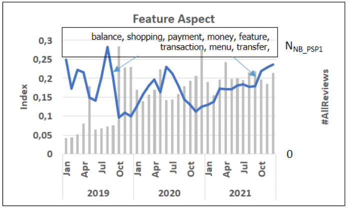
Performance on Logistic Regression with Prior Knowledge (F1 in %) Table 5

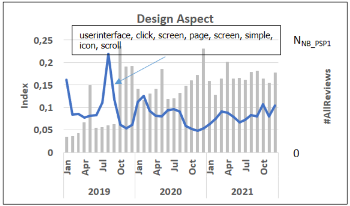
	# of Training Data	Logistic Regression	Prior Knowledge (Rule Based Model)	Logistic Regression with Prior Knowledge
Security Aspect	100	54,11	56,49	55,00
	200	66,67	56,49	67,86
	500	72,64	56,49	75,20
	800	75,76	56,49	78,82
	1016	78,17	56,49	**81,36**
Feature Aspect	100	50,39	56,14	57,65
	200	51,81	56,14	74,90
	500	62,50	56,14	77,67
	800	64,15	56,14	77,87
	1016	74,80	56,14	**78,39**
Design Aspect	100	45,24	55,43	52,38
	200	60,32	55,43	64,43
	500	63,25	55,43	67,95
	800	63,45	55,43	69,39
	1016	69,32	55,43	**73,52**
Customer Service Aspect	100	45,27	53,78	54,86
	200	54,58	53,78	74,58
	500	68,00	53,78	74,58
	800	74,16	53,78	77,46
	1016	79,07	53,78	**82,78**
Technology Aspect	100	40,00	55,85	47,50
	200	55,81	55,85	65,21
	500	63,64	55,85	70,61
	800	65,16	55,85	75,30
	1016	65,61	55,85	**77,90**

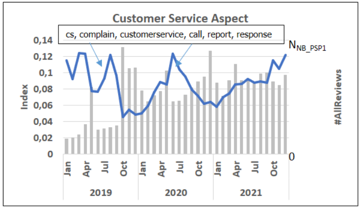
Result: Index (NB_PSP1)

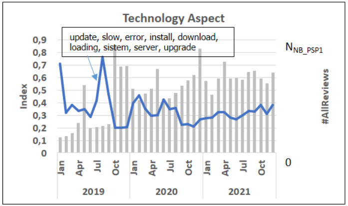
In March and April 2019, the NB_PSP1 showed the increasing index on Security aspect, Feature aspect, as well as Customer Service aspect. In the period of July 2020 when the pandemic of COVID-19 had started to hit the country, the complaint and negative sentiments were about the Security, Feature and Customer Service aspect.The Technology and Design aspects, however, showed only once dramatic increasing curve in July 2019 and declined afterwards → it was identified that some improvement happened.







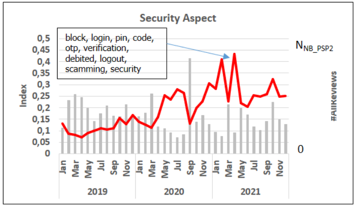


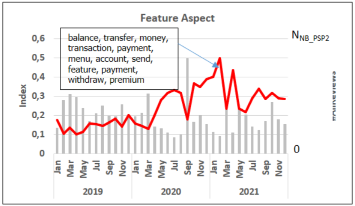


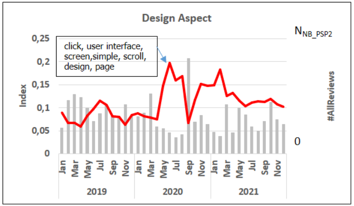
$$\text{idx}_{o,t} = \frac{\sum_i 1_{f(x_{o,i,t})=1}}{N_{o,t}} \, x \log N_{o,t}$$

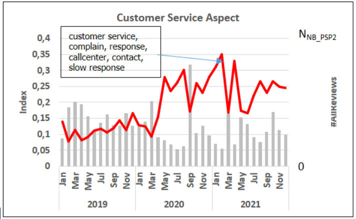
Result: Index (NB_PSP2)

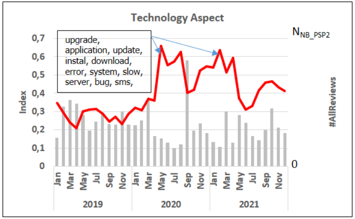
During the period of 2019 until 2021 the number of overall reviews showed relatively no significant increase, except in September 2020. The index, however, started to increase at the time the pandemic of COVID-19 hit the country. It is identified that the app's demand was highly increasing, and some possible troubles or unsatisfactory things were reported by user more than before. The most complained aspect was Technology aspect, 2020-Q2, 2021-Q1.











$$\text{idx}_{o,t} = \frac{\sum_i 1_{f(x_{o,i,t})=1}}{N_{o,t}} \, x \log N_{0,t}$$

BANK INDONESIA

Conclusion and Future Directions

Conclusion

1. We conducted some experiments on the inference models: deterministic model (rule based), model-from-data (machine learning), and model-from-data incorporating prior knowledge. Prior knowledge can help to improve the inference process when the number of training data is limited.
2. Having demonstrating it onto 5 aspects, we suggest that this approach can be used to infer the user's sentiment of mobile non-bank payment apps by some aspects in a big data quickly.
3. We also propose an index that is based on the inferred sentiment and the number of reviews. The series constructed by the index calculation represents the recent condition because the reviews can be scrapped at any time from the application's store.
4. By automatically analyse the sentiment and monitoring the series periodically we can suggest that this approach can be used in order to timely monitor the nonbank-PSP performance, e.g. as a leading indicator so that any further systemic risk can be anticipated in advance.

Future Directions

1. Advancement on Text Normalisation Methods.
2. Enrich more various user's expression.
3. Add more Human Languages.