P2P Lending Sentiment Analysis in Indonesian Online News

Ryan Randy SURYONO[1,2] and Indra BUDI[1*]

[1]*Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia*
[2]*Faculty of Engineering and Computer Science, Universitas Teknokrat Indonesia, Indonesia*
Corresponding author: indra@cs.ui.ac.id

ABSTRACT

Fintech has improved from a few years ago and has put regulators under pressure to find a legal framework that allows fintech to operate in the formal financial sector and provide appropriate protection for customers. At present, many online news in Indonesia contain articles about Fintech, especially P2P (Peer to peer) Lending. The positive and negative sides of the development of P2P Lending are interesting for further investigation. This study aims to determine the best text classification techniques from P2P Lending sentiment analysis on Indonesian Online News. This research compared four algorithms which are Multinomial Naïve Bayes (NB), Logistic Regression (LR), Support Vector Machine (SVM) and Random Forest (RF). The experiment was carried out using features combination and the model was measured using 10-fold cross validation. The result is the SVM classification model achieves the highest accuracy score of 63.61% on the TFIDF Unigram-Trigram feature.

Keywords: *fintech, P2P Lending, sentiment analysis, classification*

INTRODUCTION

With the rapid expansion of the internet market and the development of small and medium enterprises (SMEs), it encourages Financial Technology (Fintech) to innovate in funding and other financial activities. The term Fintech which consists of Financial and Technology is basically understood as a technology used to change and improve financial business models [6]. Like several other countries, Fintech especially P2P Lending is one of the fastest growing sectors in Indonesia. The Fintech has increased from last year and has made regulators under pressure to find a legal framework that allows fintech to operate in the formal financial sector and provide appropriate protection for customers [4]. By utilizing modern communication technology, online lending can reduce certain transaction costs, making it economically feasible to raise funds through small contributions from a large number of investors [19]. As an attractive alternative to traditional bank financing, online lending have experienced exponential growth worldwide, since zopa.com, the world's first P2P platform, emerged in the UK in 2005. In a short period of about 10 years, global online lending have reached more than USD 100 billion at the end of 2015, and has been projected to exceed USD one trillion by the end of 2025 [6].

Compared to traditional financial intermediaries such as banks, in a P2P online lending platform, borrowers can apply for lending directly to lenders and use textual narratives to describe their lending [19]. The borrower will assess the feasibility of the lending with a variety of credit scoring methods [13]. Previous research explains that trust in borrowers has a stronger effect on lending intentions than trust in platforms. When considering a P2P online lending platform, it is more important to provide accurate and updated information to gain trust from the lender [19]. The development of funding models without the mediation of financial institutions makes peer to peer lending get public attention. Using data from the Lending Club, which is one of the popular online P2P lending houses, exploration of the characteristics of P2P lending has been carried out, evaluating credit risk and measuring lending performance. The results of the study found that credit ratings, debt to income ratios, FICO scores, and utilization of detour lines played an important role in lending defaults. The higher interest rates charged to high-risk borrowers are not enough to compensate for the probability of a higher lending default [11]. However, not only seen from the side of the borrower, the P2P Lending Platform in China has a moral hazard and is very easy for borrowers to falsify lending information [12]. The problem of moral danger and adverse selection is triggered by asymmetric information, so it also creates regulatory challenges [20].

The positive and negative sides of the development of P2P Lending are interesting for further investigation. In Indonesia, quite a lot of online news that contains information about P2P Lending. The news contains positive, negative, and neutral values for the P2P Lending topic. There is some news that the emergence of P2P is a solution to the limitations of banks in an effort to increase SMEs credit. P2P Lending acts as a financial service institution that provides access to SMEs who are actually eligible for credit (creditworthy), but have difficulty obtaining bank credit. Business processes that are concise, fast, and transparent make the lending process easier. The use of information technology will also make credit equalization throughout Indonesia. However, there is also negative news about P2P Lending related to improper billing, high interest, and even other moral hazard actions. News related to P2P Lending began to talk about the fact P2P Lending as a loan shark [17].

It's important to know the sentiments in P2P Lending. This can provide input for the government in making policies and is beneficial for the community in choosing Fintech P2P Lending as an alternative funding. And to do that, we can use classification techniques/ approaches. This study aims to determine the best text classification techniques from P2P Lending sentiments on Indonesian Online News.

Because there are no methods with stable accuracy in a number of studies, this study compares four algorithms which are Multinomial Naïve Bayes (NB), Logistic Regression (LR), Support Vector Machine (SVM) and Random Forest (RF) [9,7]. The uniqueness of classifying online news is a challenge for this research. It's different from doing sentiment analysis on Twitter data which has a limited number of characters. Research with a long amount of text must consider the model and features used. For this reason, this research contributes to find out which models and features have the best accuracy for P2P Lending news. Finally, this paper is organized as follows: In Section 2, we conducted a literature study on P2P Lending, Indonesian Online News, and Sentiment Analysis. Then in Part 3 it explains the research steps applied in this experiment. Section 4 explains the results of the experiment. And the results of the study are explained in Section 5.

LITERATURE REVIEW

This section discusses the study of P2P lending, Indonesian Online News, and Sentiment Analysis.

P2P Lending

The P2P lending model has attracted great attention from industrial and academic fields. In the financial industry, the P2P model provides a new pattern on the design and management of financial products [18]. Peer to Peer Lending (P2P Lending) is the process or method of lending money to individuals or businesses [17]. According to data from the Financial Services Authority (OJK) up to August 7, 2019, the total number of registered and licensed fintech operators in Indonesia was 127 companies. The types of fintech that developed in Indonesia were mostly P2P Lending platforms. From the previous month, there were an additional fifteen registered fintech organizers namely, qazwa.id, bsalam, onehope, LadangModal, Dhanapala, Restock, Solution, lending here, AdaPundi, Tree +, Assetkita, Edufund, Finanku, Tunasaku, and Uatas. This proves that the P2P Lending platform is highly developed in Indonesia. However, the OJK calls on the public to use the services of fintech peer to peer lending that have been registered/licensed from the OJK [16]. Prospective users who intend to borrow or lend must create an account, provide personal information, such as name, address, telephone number, and social security number. Some online P2P lending platforms (e.g. Prosper) also require users to provide bank account information. The information is then verified. Lenders make lending decisions according to the borrower's register information and personal information [3]. With the easy process of a traditional lending model with banks, the most important problem with online lending is the lack of legal basis in online P2P intermediary lending supervision and the lack of security guarantees for lenders [17].

Indonesian Online News

News is "the communication of information chosen about current events", in which the selection is by "valuable news" or "which attracts public interest". News is also a story, from which readers usually expect answers to: who, what, when, where and why, which are often added "how". News-style writing - as opposed to, for example, writing comments - generally strives for objectivity and / or neutrality (representation of different views at the event) [2]. In terms of linguistics, news has different characteristics from blogs and social media. News contains content created by people who have knowledge in the field of journalism and are deemed to have met the standards in conveying information. This can provide an opportunity to analyze about the content using analytic data so that people can find out information that is grouped based on events and time. In terms of the availability of news content, online news is one source that is easy to obtain using internet access [15]. Cyber media has a special character that requires guidelines so that its management can be carried out in a professional manner, fulfilling its functions, rights and obligations in accordance with the Law. For this reason, Indonesia formed a press council (dewanpers.or.id) to verify online news to safeguard factual information.

Sentiment Analysis

Sentiment analysis on news articles has been carried out by social science researchers to investigate research questions related to public opinion and perception. The analysis was conducted around public image and community reputation, such as race, ethnicity, gender, religion, and social class [10]. Currently, there are studies using sentiment analysis for Abusive Language Detection in Indonesian Social Media. The study discusses coarse Indonesian on social media which generally comes from unpleasant conditions or something that is disgusting and prohibited by religion. The experimental results show that Naive Bayes (NB) is better than Support Vector Machine (SVM) and Random Forest Decision Tree (RFDT) for classifying abusive language using dataset in all scenarios [7]. In terms of data, most studies use Twitter datasets that are limited in number of characters. Existing studies consider the detection of hate speech as one of the tasks of text classification. Facebook, as the most widely used social media, has high potential to be used to spread hate speech. This experiment was conducted by Auliya et al, using the Support Vector Machine (SVM) as its classifier algorithm using TF-IDF, quad-gram char, word unigram, and lexicon features that yield f1-score of 85% [1]. Then what about the classification method that uses online news datasets. For automation of sentiment analysis, various approaches have been created to predict sentiments from words, expressions, and from documents. There are many natural language based processing and pattern based algorithms such as Naïve Bayes (NB), Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF) [5, 8, 14] The uniqueness of P2P Lending analysis sentiment on online news has a characteristic of its news content. Generally, news texts about P2P Lending contained in the words of regulators (for example "Financial Services Authority" and "KOMINFO") usually talk about P2P Lending violations. While the news text that contains the words "investment", "assets", "Funds" usually tells about the benefits of investment on the P2P Lending platform. This is interesting, because the combination of text in online news that is used is different from the text on Twitter data or social media data like Facebook.

METHODOLOGY

This section explains the method proposed in this study. This experiment consists of 6 steps shown in Fig. 1. The first is collecting data online news, the second step is data annotation, the third step is applying preprocessing techniques, the fourth is feature extraction, the fifth is an experiment with a classification model, and the last is a model evaluation.
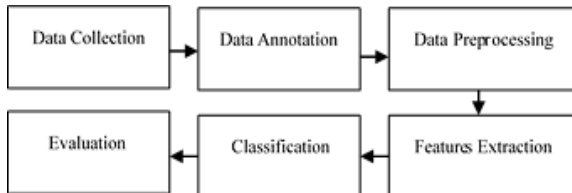

Figure 1. Stages of Experiment [7]

Data Collection

In this step, we indexed the top 10 news portal based on Alexa.com and this research collected 1015 articles on the Indonesian online news portal that was verified administration and factual by Dewanpers. Then obtained online news portals such as *Tribunnews.com, Detik.com, Okezone.com, Sindonews.com, Kompas.com, Liputan6.com, Merdeka.com, Cnnindonesia.com, Tempo.co,* and *Bisnis.com.* After that scrapped the online news site with query "P2P Lending". An example of the data shown in Fig. 2.
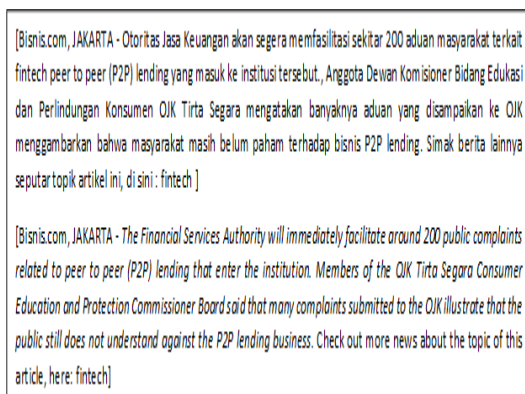

Figure 2. Collected Data from News Portal

Data Annotation

After retrieving data, this experiment provide annotations manually. In this annotation assignment, this study provide a data set from an online news database with a minimum length of 1000 characters. Majority voters are linguists and have graduated from masters. They are asked to perform two tasks, including determining whether the news provided contains Fintech P2P Lending and determining whether the news is pro, contra, or neutral against P2P Lending.

Because it consists of three Majority Voters, this study uses majority voting to determine the label. From 1100 news articles, there are 85 no agreement articles, 156 news articles containing negative sentiment, 288 news articles containing neutral, and 571 news articles containing positive sentiment. The process of determining the majority voting is done by looking at the similarity of answers between rater. For example if rater1 fills the label with -1, rater2 fills the label with 0, and rater3 fills the label with -1 then the conclusion of the label contents is -1. Based on the annotation guide, sentiment in P2P Lending can be divided into three types, including:

- Positive Sentiment, which is a news article that contains positive values towards P2P Lending such as opinions about advantages of P2P Lending, benefits of P2P Lending, positive developments from P2P Lending and positive response from the government and society. For example, *"Jika dibandingkan dengan mengajukan pinjaman ke bank, koperasi, atau sektor swasta lainnya, proses mengajukan pinjaman di platform P2P sangatlah mudah."* ("When compared to applying for lending to banks, cooperatives, or other private sectors, *the process of applying for lending on the P2P platform is very easy.")*
- Negative Sentiments, namely news articles that contain negative values on P2P Lending such as opinions about P2P Lending weaknesses, Negative Issues of P2P Lending (Billing Case, Illegal Fintech) and Negative responses from the government and the community. For example, "Kekurangan pinjaman online langsung cair yang terakhir ini memang patut kamu perhatikan karena banyaknya penyedia pinjaman online langsung cair yang menjamur tapi tidak semuanya bisa terpercaya dan memiliki kredibilitas tinggi." ("The disadvantages of the last direct liquid online lending are worth noting because there are many liquid online direct lending providers that are mushrooming, but not all of them can be trusted and have high credibility.")
- Neutral Sentiment, which is a news article that contains neutral values towards P2P Lending (news that only displays information such as definitions and explanations about Fintech P2P Lending). For example, "Data yang dirilis Otoritas Jasa Keuangan Indonesia (OJK) menunjukkan, per Jumat (19/10), sudah ada 73 penyelenggara layanan pinjam meminjam berbasis teknologi atau fintech peer to peer (P2P) lending yang tercatat." ("Data released by the Indonesian Financial Services Authority (OJK) shows, as of Friday (10/19), there have been 73 providers of technology-based lending or peer-to-peer lending (P2P) lending services recorded.")

Data Preprocessing

Before starting the feature extraction step, this study do some basic text processing techniques to clear data. Some preprocessing procedures that have been carried out include lowercasing, removing punctuation, removing

extra space, removing numbers, removing stopwords, and stemming.

- Case Folding, aims to make all text lowercase such as 'ke UKM-UKM Indonesia' changed to 'ke ukm-ukm indonesia'.
- Removing punctuation aims to delete all non-alphabet characters. For instance, 'â??Kami optimistis,' (â??We are optimistic,) changed to 'Kami optimistis' (We are optimistic).
- Removing extra space we will remove or strip leading, trailing and duplicate spaces.
- Removing numbers is used to delete numbers that are always in front of and behind a word and also numbers that are not needed in research. For example, '(+62811-93004)'.
- Removing stopwords is processed in a sentence if it contains words that often come out and are considered unimportant such as time and conjunctions. We use Sastrawi library for removing stopwords such as 'bisa, maka, atau, dan' (can, then, or, and). For example 'kita bisa pakai transaksi non-tunai' changed to 'kita pakai transaksi non-tunai'.
- Stemming, the stage of stemming is the stage of finding the root word of each word filtering results. We used Snowball Stemmer from Sastrawi library for Indonesian. For example 'OJK siap memberikan akses' changed to 'OJK siap beri akses'.

Features Extraction

After the data preprocessing phase, this experiment extract features that are used in the model. Some of the features use are the term bigram and the vector is extracted by creating word representation vectors in online news. In the fields of linguistics and computational probability, the term n-gram is a sequence that is adjacent to n items from a text sample. Items can be syllables, phonemes, letters, words or language pairs. N-grams are usually collected from text or corpus of speech. N-gram size 1 is referred to as "unigram"; size 2 is "bigram"; size 3 is "trigram". This study used several features: Unigram TF-IDF; unigram Vector; Bigram TF-IDF; Bigram Vector; Trigram TF-IDF; Trigram

Vector; Unigram-Bigram; Unigram-Trigram; Bigram-Trigram. Then use a combination of stemming and stopwords steps to see whether the model's score can increase.

Classification and Evaluation

This study implemented some machine learning algorithms such as are Multinomial Naïve Bayes (NB), Logistic Regression (LR), Support Vector Machine (SVM) and Random Forest (RF) as the classifier. This algorithms were implemented using Scikit-learn library. For evaluation, we applied 10-fold cross validation. Then, F1-score was used as performance evaluation metrics in this study.

EXPERIMENT AND RESULTS

In this experiment, used four classifiers including Naïve Bayes (NB), Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF). This study conduct experiments by combining the features and classifying algorithms. For each classifier, apply feature combinations to classify data and measure its accuracy. Table 1 explain the result of features combination.

From the table, we can see that NB has the highest accuracy score for the Unigram-Trigram TF feature at 62.14%. Whereas LR has the highest accuracy score for the Unigram TFIDF feature of 62.62%. In addition, SVM has the highest accuracy score for the Unigram-Trigram TFIDF feature of 63.61%. Furthermore RF has the highest accuracy score for the Bigram-Trigram TFIDF feature of 61.34%.

In terms of features, for the Unigram TF feature, the best classifier is NB with an accuracy score of 61.35%. For TFIDF's Unigram feature, the best classifier is LR with an accuracy score of 62.62%. For the Bigram TF feature, the best classifier is LR with an accuracy score of 62.15%. For the Bigram TFIDF feature, the best classification is SVM with an accuracy score of 61.93%. For the Trigram TF feature, the best classifier is the LR with an accuracy score of 61.06%. For the TFIDF Trigram feature, the best classification is SVM with an accuracy score of 59.56%.

Table 1. The Result of Features Combination

No.	Features Combination	NB	LR	SVM	RF
1	Unigram TF	61,35%	60,56%	59,09%	60,24%
2	Unigram TFIDF	56,60%	62,62%	61,24%	57,60%
3	Bigram TF	59,29%	62,15%	61,37%	60,43%
4	Bigram TFIDF	57,00%	59,57%	61,93%	59,17%
5	Trigram TF	42,90%	61,06%	53,78%	59,17%
6	Trigram TFIDF	56,41%	57,10%	59,56%	58,27%
7	Unigram-Bigram TF	62,82%	60,47%	59,98%	59,96%
8	Unigram-Bigram TFIDF	56,31%	61,54%	62,53%	58,98%
9	Unigram-Trigram TF	62,14%	61,66%	61,27%	60,85%
10	Unigram-Trigram TFIDF	56,31%	61,05%	63,61%	60,83%
11	Bigram-Trigram TF	53,16%	61,85%	60,19%	59,86%
12	Bigram-Trigram TFIDF	56,51%	57,79%	60,84%	61,34%

Furthermore, for the Unigram-Bigram TF feature, the best classifier is NB with an accuracy score of 62.82%. For TFIDF's Unigram-Bigram feature, the best classifier is SVM with an accuracy score of 62.53%. For the Unigram-Trigram TF feature, the best classifier is NB with an accuracy score of 62.14%. For TFIDF's Unigram-Trigram feature, the best classifier is SVM with an accuracy score of 63.61%. In addition, for the Bigram-Trigram TF feature, the best classifier is the LR with an accuracy score of 61.85%. Finally, for the TFIDF Bigram-Trigram feature, the best classifier is RF with an accuracy score of 61.34%. Besides seeing the highest accuracy value, the NB model with the Trigram TF feature has the lowest accuracy value of 42.90%. Comparison of accuracy values can be seen in Fig. 3.
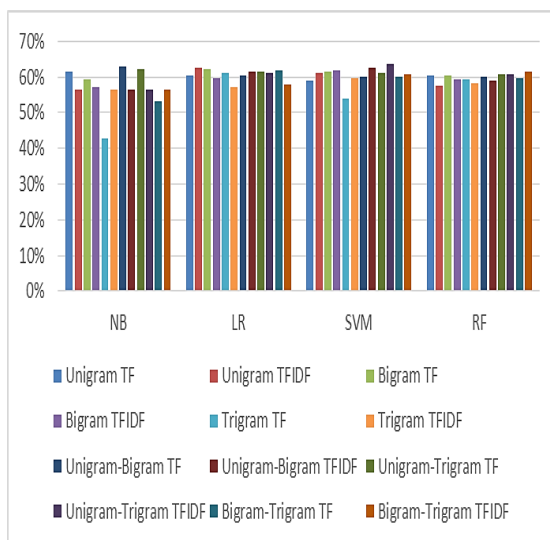

Figure 3. Comparison result of models

We can also see from the confusion matrix. This is one method that can be used to measure the performance of a classification method. Basically the confusion matrix contains information that compares the results of the classification carried out by the system with the results of the classification that should be. We take an example of a confusion matrix of 10 fold (See Fig. 4.) and compares the golden truth with the results of the model. The matrix sequence is negative, neutral, and positive (-1, 0, 1). There are 8 data that are assessed by majority voting and classifiers as negative sentiments. In addition, from 31 data, 4 data were classified as negative sentiment, 5 data as neutral, and 23 data as positive sentiment. Furthermore, 8 data were classified as negative sentiment, 11 data were considered neutral, and 43 data were rated by majority voting and classifiers were negative sentiment. Here we can see that news texts that are considered majority voting as neutral (5 data) are not supported classifier result. Most of the neutral sentiment classified as positive sentiment (23). Our analysis, this happen because the number of positive articles is higher than other sentiments. For example, "*The peer to peer lending platform guarantees the funds you invest based on funding risk. The annual interest received in P2P Lending investments usually has an amount twice the conventional savings account.*". From the article, it could have a positive meaning because

it explains about the benefits of P2P Lending but it can also be neutral if the statement is considered an information.
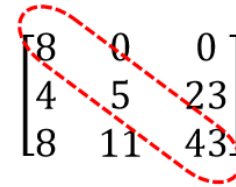

Figure 4. Confusion Matrix

CONCLUSION AND FUTURE WORKS

In this work, this study have examined the performance of four machine learning algorithms namely Naïve Bayes (NB), Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF) to classify P2P Lending sentiments on Indonesian online news. Evaluation is done by Preprocessing procedures that have been carried out including low-casings, removing punctuation, removing extra space, removing numbers, removing stopwords, and stemming. Furthermore, this experiment use features combination and the model is measured using 10-fold cross validation.

In conclusion, the results of the Support Vector Machine (SVM) classification model achieved the highest accuracy score of 63.61% in the TFIDF Unigram-Trigram feature compared to other models. Conducting sentiment analysis on long texts has a greater challenge especially since the text is online news. Online news has a combination of more formal words and it is difficult to distinguish neutral and positive meanings. This is indicated by the interpretation of the results of the confusion matrix. Our analysis, this is due to the unbalanced number of datasets for positive, neutral and negative sentiment labels.

In the future, we will try to use the scenario of under sampling and over sampling. We consider whether accuracy results will be higher if modifications are made to classification modeling. In addition, we will also add an online news dataset to produce better accuracy. Our next challenge is to analyze the content of the news and determine news trends based on timeline. Future studies can also do Text summarization to create a summary that preserves the most important points of the original document.

ACKNOWLEDGMENT

REFERENCES

[1]   N. Aulia, and I. Budi, Hate Speech Detection on Indonesian Long Text Documents Using Machine Learning Approach, (2019) 164–169.

[2]   B. Berendt, Text Mining for News and Blogs Analysis, Encycl. Mach. Learn. Data Min. (2019) 1247–1255.

[3]   D. Chen, F. Lai, and Z. Lin, A trust model for online peer-to-peer lending: a lender's perspective, Inf. Technol. Manag. 15 (2014) 239–254.

[4]   K. Davis, R. Maddock, and M. Foo, Catching up with indonesia's fintech industry, Law Financ. Mark. Rev. 11 (2017) 33–40.

[5]   K. S. Doddi, Y. V Haribhakta, and P. Kulkarni, Sentiment Classification of News Articles, Int. J. Comput. Sci. Inf. Technol. 5 (2014) 4621–4623.

[6]   R. H. Huang, Online P2P Lending and Regulatory Responses in China: Opportunities and Challenges, Eur. Bus. Organ. Law Rev. 19 (2018) 63–92.

[7]   M. O. Ibrohim, and I. Budi, A Dataset and Preliminaries Study for Abusive Language Detection in Indonesian Social Media, Procedia Comput. Sci. 135 (2018) 222–229.

[8]   M. Kaya, G. Fidan, and I. H. Toroslu, Sentiment analysis of Turkish political news, in *Proc. - 2012 IEEE/WIC/ACM Int. Conf. Web Intell. WI 2012*, 2012), pp. 174–180.

[9]   A. Khadjeh Nassirtoussi, S. Aghabozorgi, T. Ying Wah, and D. C. L. Ngo, Text mining of news-headlines for FOREX market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment, Expert Syst. Appl. 42 (2015) 306–324.

[10]   C. S. G. Khoo, A. Nourbakhsh, and J. C. Na, Sentiment analysis of online news text: A case study of appraisal theory, Online Inf. Rev. 36 (2012) 858–878.

[11]   H. Li, Y. Zhang, N. Zhang, and H. Jia, Detecting the Abnormal Lenders from P2P Lending Data, in *Procedia Comput. Sci.*, School of Information, Central University of Finance and Economics, Beijing, China, 2016), pp. 357–361.

[12]   J. Liu, The dramatic rise and fall of online P2P lending in China, Www.Techcrunch.Com (2018).

[13]   X. Ma, J. Sha, D. Wang, Y. Yu, Q. Yang, and X. Niu, Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning, Electron. Commer. Res. Appl. 31 (2018) 24–39.

[14]   T. B. H. . H. T. Magadza, A. Mukwazvure, and K. . P. Supreethi, Exploring Sentiment Classification Techniques in News Articles, Ijitkm 8 (2014) 55–58.

[15]   K. Nørvåg, and R. Øyri, News item extraction for text mining in Web newspapers, Proc. - Int. Work. Challenges Web Inf. Retr. Integr. WIRI'05 2005 (2005) 195–204.

[16]   OJK, *Penyelenggara Fintech Terdaftar dan Berizin di OJK per 7 Agustus 2019*, 2019).

[17]   R. R. Suryono, B. Purwandari, and I. Budi, Peer to Peer (P2P) Lending Problems and Potential Solutions : A Systematic Literature Review, Fifth Inf. Syst. Int. Conf. 2019 (2019).

[18]   H. Wang, K. Chen, W. Zhu, and Z. Song, A process model on P2P lending, Financ. Innov. 1 (2015).

[19]   Q. Yang, and Y.-C. Lee, Critical Factors of the Lending Intention of Online P2P: Moderating Role of Perceived Benefit, in *Proc. 18th Annu. Int. Conf. Electron. Commer. E-Commerce Smart Connect. World*, (ACM, New York, NY, USA, 2016), pp. 15:1--15:8.

[20]   Z. Zhu, Safety promise , moral hazard and financial supervision : Evidence from peer-to-peer lending, Financ. Res. Lett. 27 (2018) 1–5.