Text Mining Based on Tax Comments as Big Data Analysis Using SVM and Feature Selection

Abstract

- The tax gives an important role for the contributions of the economy and development of a country. The improvements to the taxation service system continuously done in order to increase the State Budget.
- One of consideration to know the performance of taxation particularly in Indonesia is to know the public opinion as for the object service.
- Text mining can be used to know public opinion about the tax system. The rapid growth of data in social media initiates this research to use the data source as big data analysis.
- The dataset used is derived from Facebook and Twitter as a source of data in processing tax comments. The results of opinions in the form of public sentiment in part of service, website system, and news can be used as consideration to improve the quality of tax services.
- In this research, text mining is done through the phases of text processing, feature selection and classification with Support Vector Machine (SVM).
- To reduce the problem of the number of attributes on the dataset in classifying text, Feature Selection used the Information Gain to select the relevant terms to the tax topic.
- Testing is used to measure the performance level of SVM with Feature Selection from two data sources. Performance measured using the parameters of precision, recall, and F-measure.

Existing

- The improvements to the taxation service system continuously done in order to increase the State Budget. One of consideration to know the performance of taxation particularly in Indonesia is to know the public opinion as for the object service.
- Text mining can be used to know public opinion about the tax system. The rapid growth of data in social media initiates this research to use the data source as big data analysis.
- The dataset used is derived from Facebook and Twitter as a source of data in processing tax comments. The results of opinions in the form of public sentiment in part of service, website system, and news can be used as consideration to improve the quality of tax services.

Proposed

- proposed a text mining processing through SVM method with classification optimization with Feature Selection.
- Feature Selection is used to select the relevant feature of the dataset in order to get a better performance of SVM as a classifier.
- Text mining aims to generate a classification on the sentiment about the problem of taxation based on data sources the public comments on Facebook and Twitter.
- In this study, the results of positive and negative sentiments are based on time period and the type of tax data namely service, website system, and tax news.
- For further research, information generated from this text mining can be used as considerable of taxation and support services for future Policies

Advantage

- Classification results based on positive and negative sentiments with three categories namely service, website, and news Facebook and Twitter showed the results of the performance algorithm that the average of the proposed method about 75%, 70% and 72% achieved

Conclusion

- Text mining in this research is obtained through the stages of text processing, feature selection using Information Gain and classification using SVM.
- The datasets obtained are derived from Facebook and Twitter comments on taxes.
- Classification results based on positive and negative sentiments with three categories namely service, website, and news.
- From the measurement of precision, recall and F-measure obtained on the dataset from Facebook and Twitter showed the results of the performance algorithm that the average of the proposed method about 75%, 70% and 72% achieved.
- This research can be used as a basis for big data analysis in tax case an evaluation of tax service based on public opinion

Reference

[1] Admin. (2017, October 15), State Revenue Realization – Ministry of Finance of the Republic of Indonesia. Available online:www.bps.go.id

[2] Leismester, C. 2015. Mastering Machine Learning with R. Published by Packt Publishing Ltd. Livery Place 35 Livery Street. Methods. USA: A Wiley-Interscience Publication. [3] Pathak, M, A. 2014. Beginning Data Science with R. Springer International Publishing Switzerland 2014.

[4] Jadon, E., Sharma R. Data Mining: Document Classification using Naive Bayes Classifier. International Journal of Computer Applications Volume 167 – No. 6, June 2017.

[5] Lu, H., Setiono, R. Liu, H. NeuroRule: A Connectionist Approach to Data Mining.2017. Available online : arXiv:1701.01358v1 [cs.LG].