JOIN (Jurnal Online Informatika)
p-ISSN: 2528-1682, e-ISSN: 2527-9165
Volume 9 Number 1 | June 2024: 100-110
DOI: 10.15575/join.v9i1.1275

SAER : Comparison of Rule Prediction Algorithms on Constructing a Corpus for Taxation Related Tweet Aspect-Based Sentiment Analysis

Annisa Mufidah Sopian[1], Ridwan Ilyas[2], Fatan Kasyidi[3], Asep Id Hadiana[4]
[1,2,3]Department of Informatics, Faculty of Science and Informatics, Universitas Jenderal Achmad Yani, Indonesia
[4]Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia

Article Info

Article history:

Received January 10, 2024
Revised March 27, 2024
Accepted April 02, 2024
Available Online May 08, 2024

Keywords:

Aspect-Opinion Extraction
Rule Prediction
SAER
Syntactic feature
Taxation

ABSTRACT

Twitter is a popular social media in Indonesia, and sentiment analysis on Twitter has an important role in measuring public trust, especially in taxation issues. Aspect extraction is an important task in sentiment analysis. In this research, we propose SAER, a Syntactic Aspect-opinion Extraction and Rule prediction, that used language rule-based approach using syntactic features for aspect and opinion extraction, and we compare several algorithm for rule prediction such as Random Forest Regression, Decision Tree Regression, K-Nearest Neighbor Regression (KNN), Linear Regression, Support Vector Regression (SVR), and Extreme Gradient Boosting Regression (XGBoost) that can generate rules with a tree-based approach. By employing syntactic features and rule prediction, it has been able to explore important features in a sentence. In rule prediction, comparison results show that Support Vector Regression (SVR) was identified as the most effective model for aspects rule prediction, providing the best results with a Mean Squared Error (MSE) of 0.022, Root Mean Squared Error (RMSE) of 0.150, and Mean Absolute Error (MAE) of 0.123. While XGBoost was identified as the most effective model for opinions rule prediction, with MSE of 0.013, RMSE of 0.117, and MAE of 0.075. Since we used syntactic feature-based approaches and rule prediction in this work, it is expected to be implemented for other cases, with other domain datasets.

Corresponding Author:

Ridwan Ilyas,
Department of Informatics, Faculty of Science and Informatics, Universitas Jenderal Achmad Yani
Jl. Terusan Jend. Sudirman, Cimahi, Jawa Barat, Kota Cimahi, Jawa Barat 40525, Indonesia
Email: ilyas@lecture.unjani.ac.id

1. INTRODUCTION

Tax is one of the key elements in the economics and governance of a country, including Indonesia. The performance of tax officials has a significant impact on public trust and the effectiveness of national revenue [1]. Social media is a networking platform that incorporates the interaction and multiplicity of public communication [2]. Twitter is one of the most popular social media platforms in Indonesia, where people can easily find out information about current trending topic [3]. Public sentiment analysis is very important for evaluating the tax policy effectiveness and finding opportunities for improvement [4].

The research about sentiment analysis, especially in a more detailed sentiment analysis, which is aspect based sentiment analysis (ABSA), has become highly required and has gained great attention from many researchers in recent years [5]. In general, there are two primary processes in ABSA, which are aspect extraction and sentiment classification. One important task in ABSA is aspect extraction, that extracted aspect-opinion pair within a sentence [6].

There are several studies that have been proposed in aspect extraction for ABSA. In general, it can be divided into various categories, including rule-based approach[7][8], sequence model approach[9][10], topic modelling approach[11][12], and machine learning approach[13]. In the language rules approach, rules are derived from contextual patterns, that capture the diverse properties of words and their relationships in the text. In the sequence model approach, essentially the task of aspect extraction is considered similar to the task of sequence labeling, since aspects, entities, and opinion expressions are frequently interrelated and arise in the formation of certain sequential patterns in sentences. One of the most highly used sequencing models is Conditional Random Fields (CRF)[9][10]. In the topic modelling approach, the concept is that the document collections are usually a mixture of many topics, and each of these topics contains a probability distribution of words, so a topic model can be used by assuming each aspect is a unigram language model, which means, a multinomial distribution of words. This is used to categorize words into groups of similar aspects. Machine learning approaches perceives aspect extraction as a classification task and applies supervised or semi supervised learning methods to extract the aspects contained in an entity[13].

Implementation of Aspect-Based Sentiment Analysis (ABSA) becomes very important in the context of sentiment analysis related to taxation. This is due to its ability to provide a deeper understanding of public sentiment towards specific aspects. We propose an architecture that uses a language rule-based approach and a rule prediction in aspect and opinion extraction. The language rule-based approach using syntactic features for aspect and opinion extraction can build a more intuitive understanding for Indonesian language cases. The advantage of this language rule-based approach is that it a reasons of used rule are traceable, and when errors occur, it can be solved immediately [5].

The use of syntactic features will generate rule sets. These rule sets then used as learning data by the prediction model, in order to predict the right rule on certain data. We use several prediction algorithms such as Random Forest Regression, Decision Tree Regression, K-Nearest Neighbor Regression (KNN) [14], Linear Regression, Support Vector Regression (SVR), and Extreme Gradient Boosting Regression (XGBoost) [15], with the consideration of using methods that can be used for regression tasks, and methods that can generate rules with a tree-based approach.

Since we use tweet sentence data, which contains aspect and opinion candidate phrases, we need to build an effective rule set to get them, and the problem of finding an effective rule set that able to extract aspect or opinion phrases is quite a complex problem. This combination of language rule-based approach and rule prediction are further expected to contribute in the field of sentiment analysis, especially in developing aspect and opinion extraction methods.

2. METHOD

We propose an architecture that uses a language rule-based approach and a rule prediction named SAER, a Syntactic Aspect-opinion Extraction and Rule prediction. It begins with data preparation which consist of collecting tweets data from Twitter, and data annotation. In the next step, we implement language rule-based approach using syntactic features for aspect and opinion extraction, that can build a more intuitive understanding for this case. Then we train a rule prediction model, in order to predict the right rule on certain data. The SAER architecture shown in Figure 1 below.

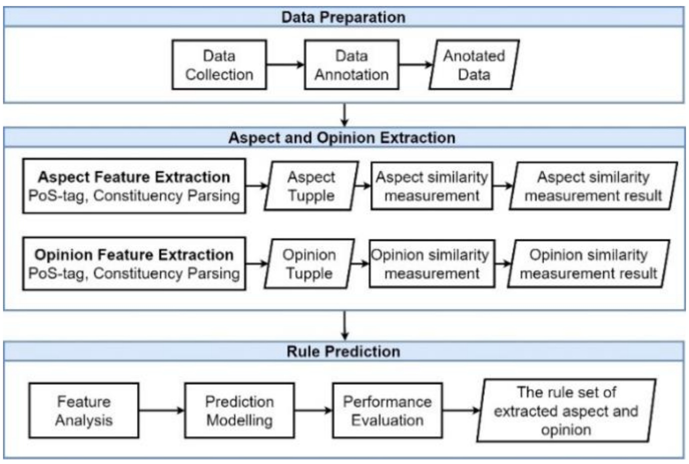

Figure 1. SAER Architecture

2.1 DATA COLLECTION

The data used in this research is tweet data from Twitter. The tweet collection was carried out in the range of March 23, 2023 to April 16, 2023. This timeframe was chosen because during this time, there was a case in Indonesia involving a tax official's son who did a criminal act, which led to many public reactions on social media[16][17]. Using this strategy, we were able to conduct a more detailed exploration of Twitter users' attitudes, viewpoints, opinions, and conversations on tax-related topics within the specified timeframe.

2.2 DATA ANNOTATION

In order to enrich the dataset, the data annotation process has a crucial place in finding the aspects, opinions, and sentiments contained in long tweet sentences. These data annotations represent phrases that are carefully identified to capture the aspects, opinions, and sentiments expressed in the wider content of each tweet. This annotation process becomes a valuable process in highlighting the main sentiments, attitudes or emotional tones present in the text. At this stage, the process of aspect and opinion labeling, and sentiment labeling contained in the tweet data is carried out manually by annotator.

a) Aspect and Opinion Labeling
Aspects are features, attributes, components, or functions of an entity, and opinions are perspectives or judgments towards an aspect [5]. On the tweets data used in this research, aspects are usually words or phrases containing the word "pajak" or taxation, and opinion are perspectives or judgments towards that taxation issues.

b) Sentiment Labeling
Sentiment labeling is performed to categorize the expressions contained in each aspect-opinion pair into three main classes: neutral, positive, and negative [3]. By categorizing the aspect-opinion pairs into these three sentiment classes, it can help to understand the overall sentiment landscape surrounding tax-related discussions on Twitter during the specified time span.

2.3 ASPECT AND OPINION EXTRACTION

The extraction of aspects and opinions can be carried out using syntactic features [5]. Syntactic features are features related to syntax, which are the rules or structures of words, phrases, or sentences. Syntactic features that are important in aspect-based sentiment analysis including Part-of-speech tagging [18], phrases structure, and dependency relationship[19]. The combination of syntactic features used in this research are part-of-speech tagging and phrases structure that were obtained through constituency parsing. Constituency parsing requires POS tag information for each word or token, and then each phrase in the sentence is labelled with a constituency label. Constituency tree forms a collection of nodes consisting of internal nodes and leaf nodes. Each internal node represents the constituency label of a phrase and each leaf node represents a word in the sentence that has been assigned the POS tag of the word [20]. POS tagging structure and constituency parsing structure can be generated by Stanford POS tagger and constituent parser [5]. POS tagging structure and constituency parsing structure shown in Figure 2 and Figure 3.

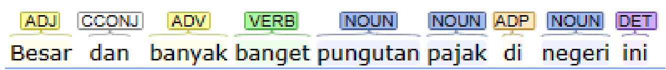

Figure 2. Part of Speech Tagging (POS tagging)

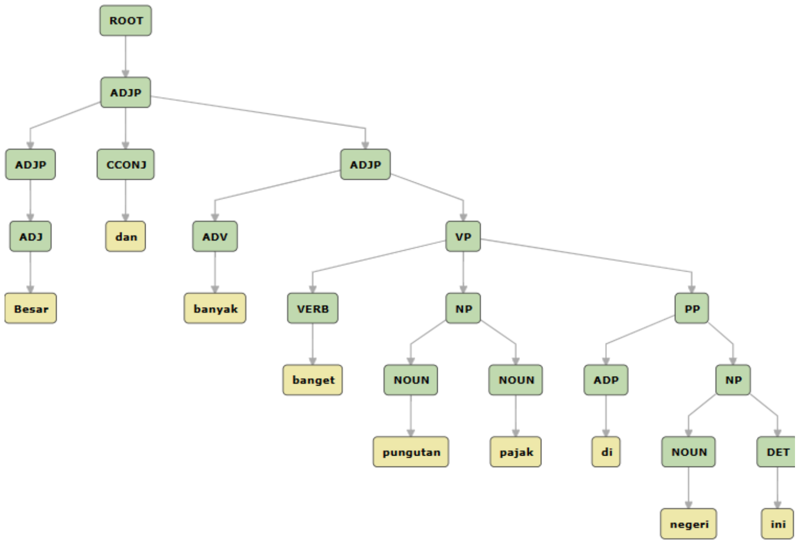

Figure 3. Constituency Parsing

The extracted aspects are chunks of the whole sentence. According to [21] nouns or noun phrases are usually represents as an aspect term. In the case of taxation, we also found that there were many aspects of the sentence that were verb phrases. In syntactic feature extraction, extraction of some additional factors is also carried out to determine and group important phrases. These extracted features are chunk position or constituency label position, constituency sub-tree depth, and leaf node length. These features then will be used as a rule set for the rule prediction process in the next step.

The steps in the opinion extraction process are almost similar to the steps carried out in the aspect extraction process, that we perform an exploration of constituency labels, labels position, sub-tree depth, and leaf node length, that allow us to extract all possible opinion phrases in the sentence.

After aspects and opinions have been extracted, then similarity values will be calculated between the aspects and opinions obtained from the labeling by annotator in the previous stage, with the aspects and opinions obtained from the extraction process with syntactic features at this stage. Similarity calculation is carried out using the cosine similarity function.

2.4 RULE PREDICTION

In this step, we predict the right rule on certain data based on an existing features and rules. This prediction process compare several algorithms which is possible to predict the rule , such as Random Forest Regression, Decision Tree Regression, K-Nearest Neighbor Regression (KNN) [14], Linear Regression, Support Vector Regression (SVR), and Extreme Gradient Boosting Regression (XGBoost) [15]. In this process, the rule set used as a feature. This rule set are extracted based on the results of constituent tree mapping which includes constituency labels, label positions, constituency sub-tree depth (height), and leaf node lengths, which represent the length of extracted phrases. This prediction model uses Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) as an evaluation metrics to evaluate model performance[14]. The metric are presented in equation[22].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y_i})^2 \tag{1}$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y_i})^2} \tag{2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \tilde{y_i}| \tag{3}$$

3. RESULT AND DISCUSSION

3.1 Data Collection

The data used is Indonesian-language tweet data that contains the keyword "pajak". Twitter data collection is performed through crawling techniques using the Twitter Application Programming Interface (API) [3]. The crawling techniques successfully collected 100,001 tweets, in the range of March 23, 2023 to April 16, 2023. This timeframe was chosen due to the incident in Indonesia that involved a tax official's son who did a criminal act, which led to many public reactions on social media [16][17].

3.2 Data Annotation

The collected tweet, then will be used in the data annotation which consists of aspect and opinion labeling processes, and sentiment labeling. The table 1 below shows an example result of the aspects and opinions labeling.

Table 1. Aspect and Opinion Labeling Result

No	Original Tweet	Labeled by Annotator
1	Besar dan banyak banget pungutan pajak di negeri ini.	Opinion: Besar dan banyak banget. Aspect: Pungutan pajak.
2	Begini cara bayar pajak yang benar, sukarela	Aspect: Bayar pajak. Opinion: Benar, Sukarela

In the first sentence, the labeled aspect is "pungutan pajak" which refers to the payment of taxes, and the extracted opinion is "besar dan banyak banget" which states that the value of the tax payment is expensive with a large amount. In the second sentence, the labeled aspect is "bayar pajak" which refers to tax payment, and the labeled opinion, is “sukarela" which means paying tax with awareness.

In each aspect-opinion pair that has been extracted, it will be labeled based on the sentiment value contained in it. Table 2 below contains an example of the results of sentiment labeling.

Table 2. Sentiment Labeling Result

No	Aspect-Opinion Phrase	Class
1	Besar dan banyak banget pungutan pajak di negeri ini.	Negative
2	Begini cara bayar pajak yang benar, sukarela.	Positive

The first examples categorized into negative sentiment class, because the opinions contained in the sentence have emotional expressions that reflect disappointment or dissatisfaction as indicated by the opinion "besar dan banyak banget". The second example categorized into positive sentiment class, because the opinion contained in the sentences have emotional expressions that reflect satisfaction as indicated by the opinion “benar, sukarela”.

The tweets that processed at this step are 1500 labeled tweets. The aspect labeling results produced 1500 tweets containing aspects, and the opinion labeling results produced 554 tweets containing opinions. The sentiment labeling process produced 68 tweets with positive sentiment, 440 tweets with negative sentiment, and 992 tweets with neutral sentiment.

Table 3 Data Annotation Results

Data	Count
Labeled Tweets	1500
Tweets contains aspect	1500
Tweets contains opinion	554
Positive tweets	68
Negative tweets	440
Neutral tweets	992

3.3 Aspect Extraction

Aspects are features, attributes, components, or functions of an entity [5]. According to the Indonesian Dictionary (KBBI) aspect is the representation or interpretation of ideas, problems, situations, and so on as considerations seen from a certain point of view [23]. Based on the definition, we will analyze features in the form of aspect word types, where the word types are various, such as nouns, verbs, etc. The table 4 below shows an example result of the aspects extraction with syntactic feature.

Table 4 Aspects Extraction with Syntactic Feature

No	Original Tweet	Labeled by Annotator	Extracted with Syntactic Feature
1	Besar dan banyak banget pungutan pajak di negeri ini.	Aspect: Pungutan pajak.	['pungutan pajak']
2	Begini cara bayar pajak yang benar, sukarela	Aspect: Bayar pajak.	['bayar pajak']

According to [21] nouns or noun phrases are usually represents as an aspect term. In the case of taxation, we also found that there were many aspects of the sentence that were verb phrases. Based on the results of aspect extraction with syntactic features, there are 273 aspects with Verb Phrase (VP) labels and 1236 aspects with Noun Phrase (NP) labels. From this process, we also obtained the distribution of chunk position values or constituency label positions, the distribution of sub-tree depth values (height), and the distribution of extracted phrase length values (length). The distribution of aspect constituency labels shown in Figure 4.

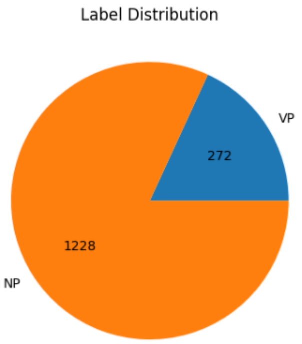

Figure 4. Aspect Constituency Label Distribution

The distribution of the label positions in this aspect extraction are shown in Figure 5 and Table 5.

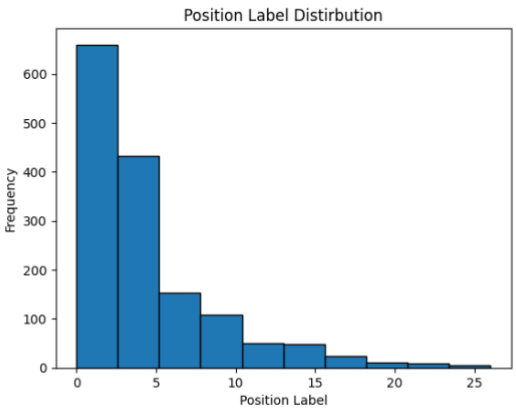

Figure 5. Position Aspect Label Distribution

Table 5. Position Aspect Label Frequency

Position	Frequency	Position	Frequency
0	318	14	11
1	142	15	13
2	200	16	10
3	170	17	7
4	163	18	6
5	99	19	1
6	79	20	10
7	74	21	5
8	50	22	2
9	32	23	1
10	27	24	2
11	29	25	2
12	22	26	1
13	24		

The distribution of sub-tree depth on each phrase in this aspect extraction are shown in Figure 6 and Table 6.

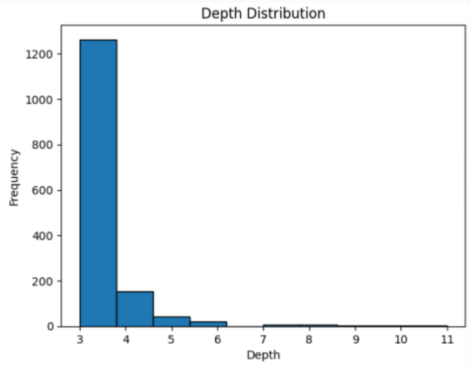

Figure 6. Aspect Tree Depth Distribution

Table 6. Aspect Tree Depth Frequency

Depth	Frequency
3	1263
4	153
5	43
6	22
7	6
8	7
9	2
10	3
11	1

The distribution of leaf node length that represent the number of words (tokens) within an extracted phrase in this aspect extraction are shown in Figure 7 and Table 7.

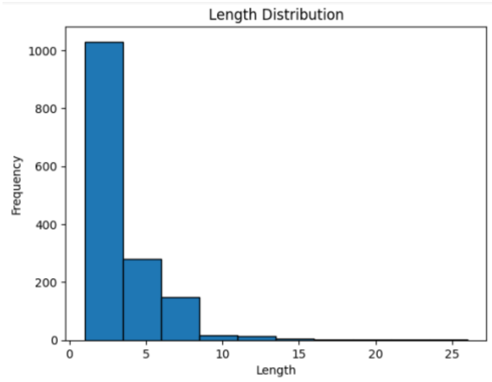

Figure 7. Aspect Phrase Length Distribution

Table 7. Aspect Phrase Length Frequency

Length	Frequency	Length	Frequency
1	128	11	8
2	596	12	3
3	306	13	2
4	173	14	2
5	108	15	3
6	76	18	1
7	37	20	1
8	36	21	1
9	12	26	1
10	6	36	1

The aspect extraction also produces a similarity value between the extracted aspects based on labeling results and extracted aspects based on the extraction results with syntactic features. The distribution of similarity results of this aspect extraction process is shown in Figure 8.

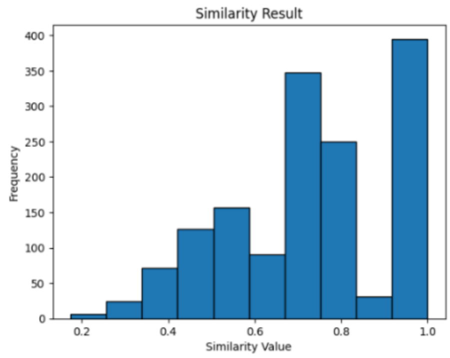

Figure 8. Aspect Similarity Result

Each of these extracted features brings its unique strengths and characteristics to the task of aspect extraction. By using this aspect extraction method, we aimed to capture a wide spectrum of linguistic nuances and variations that are inherent to Indonesian language, specifically in the context of taxation discussions on Twitter.

3.4 Opinion Extraction

According to [5] opinions are perspectives or judgments towards an aspect. If we refer to the Indonesian Dictionary (KBBI), opinion has the definition of opinion, thought, or stance [24]. From the existing definition, we will analyze the opinion feature of a sentence from its word type, where the word types are various, such as nouns, verbs, adjectives, etc. The table 8 below shows an example result of the opinion extraction with syntactic feature.

Table 8 Opinion Extraction with Syntactic Feature

No	Original Tweet	Labeled by Annotator	Extracted with Syntactic Feature
1	Besar dan banyak banget pungutan pajak di negeri ini.	Opinion: Besar dan banyak banget	['besar dan banyak banget']
2	Begini cara bayar pajak yang benar, sukarela	Opinion: benar, sukarela	['sukarela']

Figure 9 shows the distribution of extracted opinion labels. Based on the results of this opinion extraction, noun, verb, and adjective took the highest positions. Based on the constituency tree structure, we can obtain the information about the constituency label, label position, constituency sub-tree depth, and the leaf node length that represents extracted phrase length. The distribution of label positions in this opinion extraction shown in Figure 10.

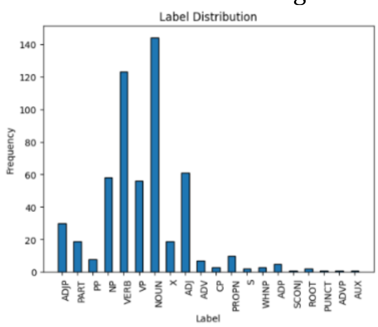

Figure 9. Opinion Label Distribution

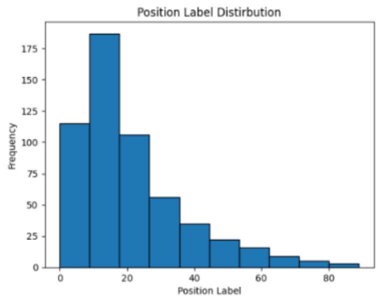

Figure 10. Position Opinion Label Distribution

The distribution of constituency sub-tree depth are shown in Figure 11 and Table 9.

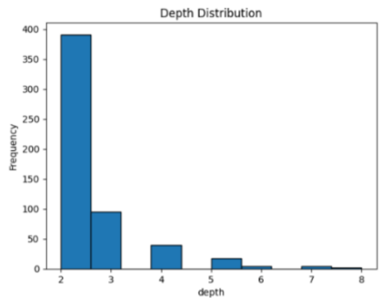

Figure 11. Opinion Tree Depth Distribution

Table 9. Opinion Tree Depth Frequency

Depth	Frequency
2	391
3	95
4	40
5	18
6	4
7	4
8	2

The distribution of phrase length or the number of words (tokens) within an extracted phrase in this opinion extraction process are shown in Figure 12 and Table 10.

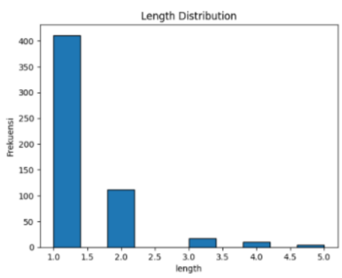

Figure 12. Opinion Phrase Length Distribution

Table 10. Opinion Phrase Length Frequency

Length	Frequency
1	411
2	112
3	17
4	10
5	4

This step also produces a similarity value between labeling results and the extracted opinion results with syntactic features. The distribution of similarity results of this opinion extraction process is shown in Figure 13.

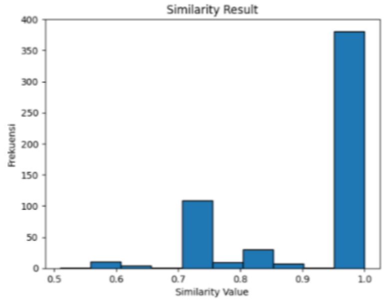

Figure 13. Opinion Similarity Result

Each of these extracted features brings its unique strengths and characteristics to the task of opinion extraction. By using this opinion extraction method, we aimed to capture a wide spectrum of linguistic nuances and variations that are inherent to Indonesian language, specifically in the context of taxation discussions on Twitter.

3.5 Rule Prediction Evaluation

As we mentioned above, rule prediction process compare several algorithms, such as Random Forest Regression, Decision Tree Regression, Extreme Gradient Boosting Regression (XGBoost), Linear Regression, Support Vector Regression (SVR), and K-Nearest Neighbors Regression (KNN) [14][15], with the consideration of using methods that can be used for regression tasks, and methods that can generate rules with a tree-based approach. Rule sets generated through constituent tree mapping are used as features in this process which include constituency labels, label positions, constituency sub-tree depth (height), and leaf node length. The table below shows the evaluation results of the rule prediction models.

Table 11. Evaluation Results of The Rule Prediction Models

Regression Model	Aspect			Opinion		
	MSE	RMSE	MAE	MSE	RMSE	MAE
Random Forest Regression	0,024	0.156	0,124	0,013	0,117	0,078
Decision Tree Regression	0,026	0.163	0,128	0,017	0,134	0,075
Extreme Gradient Boosting Regression	0,025	0.160	0,127	0,013	0,117	0,075
Linear Regression	0,026	0.161	0,133	0,017	0,130	0,104
Support Vector Regression	0,022	0.150	0,123	0,018	0,134	0,124
K-Nearest Neighbors Regression	0,029	0.171	0,134	0,019	0,139	0,100

The evaluation shows that Support Vector Regression (SVR) provides a promising results on aspect prediction, with MSE result of 0.022, RMSE result of 0.150, and MAE result of 0.123. Table 12 below shows the rule prediction result examples using SVR which is a great algorithm for the aspect.

Table 12 Aspect Prediction Result Using SVR

Rule Set				Tweet	Labeled by Annotator	Extracted aspect by rule set	Similarity Result
Constituency label	Label position	Constituency sub-tree depth	Leaf node length				
NP	5	3	3	#pajakdikebiripejabat saya berlindung dari godaan setan dan orang orang pajak yang terkutuk	Orang pajak	Orang orang pajak	0,95
NP	5	3	3	@abu_waras maaf nih bu, gak bayar pajak maksudnya kalo ke indomaret harganya turun 11% ?	Gak bayar pajak	pajak maksud nya	0,40

In the first data, by using the mentioned rule set, it has successfully extracted a quite appropriate aspect, with a similarity of 0.95. However, in another one, by using the same rule set, the extracted aspects were inappropriate. This can be caused by incorrect label position or incorrect sub-tree depth, so that the extracted sentence becomes less precise. Hence another prediction rule was needed to produce the appropriate aspects.

While in opinion prediction, Extreme Gradient Boosting Regression (XGBoost) produces a promising prediction results with MSE result of 0,013, RMSE result of 0,117, and MAE result of 0,075. The table 13 below shows the rule prediction result using XGBoost, which is a great algorithm for the opinion.

Table 13 Opinion Prediction Result Using XGBoost

Rule Set				Tweet	Labeled by Annotator	Extracted aspect by rule set	Similarity Result
Constituency label	Label position	Constituency sub-tree depth	Leaf node length				
VERB	19	2	1	@bebekchan26 @asumsico perkara dari anak belagu, bisa sampe merusak reputasi profesi seluruh karyawan pajak	merusak reputasi	merusak	0,70
VERB	19	2	1	@asumsico orang pajak masa pakai konsultan. bukan konsultasi itu mah kalo kabur bawa duid	kabur	kabur	1,00

For the first data, by using the mentioned rule set, it has extracted the right aspect with a similarity of 0.70, although there are still mistakes that may occur because of incorrect label position or wrong sub-tree depth. While in the second data, using the same rule set, it produces the precise aspect with a similarity of 1.00. This rule set prediction is one of the precise rule sets in opinion extraction. For other cases, it is still possible to explore rule set prediction, in order to enhance the aspect and opinion extraction results.

4. CONCLUSION

Based on the experimental results, we have been able to implement our proposed method, Syntactic Aspect-opinion Extraction and Rule Prediction (SAER). By utilizing syntactic features and rule sets in aspect and opinion extraction, this research is able to explore important features in a sentence such as the constituency label of each aspect and opinion phrase, label position, sub-tree depth, phrase length as seen from the extracted leaf nodes, extracted phrase, and similarity values that indicate the effectiveness of rules on certain data.

From the rule prediction results, it can be seen that Support Vector Regression (SVR) provides a promising performance in predicting aspects, with a Mean Squared Error (MSE) value of 0.022, Root Mean Squared Error (RMSE) of 0.150, and Mean Absolute Error (MAE) of 0.123. While for opinion prediction, the Extreme Gradient Boosting Regression (XGBoost) produced a promising performance with an MSE of 0,013, RMSE of 0,117, and MAE of 0,075. These findings show that rule prediction models, especially SVR and XGBoost, can be effectively used to analyze and predict ideal rules in the context of taxation tweet data.

The methodology used in this research can be implemented to build a corpus for other cases in future research, and this method can also be used to construct an aspect-based sentiment analysis system in further research. Using the baseline corpus that has been constructed in this work, a newer approach can be implemented, such as a generative approach using transformer.

ACKNOWLEDGEMENTS

This research is part of beginner lecturer research and received funding from the Directorate General of Higher Education (Dikti) Research Grant program for the Beginner Lecturer Research category on the 2023 period.

REFERENCES

[1] N. Sholihah, F. F. Abdulloh, M. Rahardi, A. Aminuddin, B. P. Asaddulloh, and A. Y. A. Nugraha, "Feature Selection Optimization for Sentiment Analysis of Tax Policy Using SMOTE and PSO," in *2023 3rd International Conference on Smart Cities, Automation & Intelligent Computing Systems (ICON-SONICS)*, IEEE, Dec. 2023, pp. 44–48. doi: 10.1109/ICON-SONICS59898.2023.10435269.

[2] S. A. Mirlohi Falavarjani, J. Jovanovic, H. Fani, A. A. Ghorbani, Z. Noorian, and E. Bagheri, "On the causal relation between real world activities and emotional expressions of social media users," *J. Assoc. Inf. Sci. Technol.*, vol. 72, no. 6, pp. 723–743, Jun. 2021, doi: 10.1002/asi.24440.

[3] M. A. Saputra and E. B. Setiawan, "Aspect Based Sentiment Analysis Using Recurrent Neural Networks (RNN) on Social Media Twitter," in *2023 International Conference on Data Science and Its Applications (ICoDSA)*, IEEE, Aug. 2023, pp. 265–270. doi: 10.1109/ICoDSA58501.2023.10276768.

[4] K. T. Shandana, A. Aminuddin, E. H. Saputra, F. F. Abdulloh, M. Rahardi, and B. P. Asaddulloh, "Sentiment Analysis of Google Classroom Application Using Machine Learning Techniques," in *2023 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA)*, IEEE, Nov. 2023, pp. 954–959. doi: 10.1109/ICAMIMIA60881.2023.10427706.

[5] F. Z. Ruskanda, D. H. Widyantoro, and A. Purwarianti, "Sequential Covering Rule Learning for Language Rule-based Aspect Extraction," in *2019 International Conference on Advanced Computer Science and information Systems (ICACSIS)*, IEEE, Oct. 2019, pp. 229–234. doi: 10.1109/ICACSIS47736.2019.8979743.

[6] F. Z. Ruskanda, D. H. Widyantoro, and A. Purwarianti, "Efficient Utilization of Dependency Pattern and Sequential Covering for Aspect Extraction Rule Learning," *J. ICT Res. Appl.*, vol. 14, no. 1, p. 51, Jul. 2020, doi: 10.5614/itbj.ict.res.appl.2020.14.1.4.

[7] P. Ray and A. Chakrabarti, "A Mixed approach of Deep Learning method and Rule-Based method to improve Aspect Level Sentiment Analysis," *Appl. Comput. Informatics*, vol. 18, no. 1/2, pp. 163–178, Mar. 2022, doi: 10.1016/j.aci.2019.02.002.

[8] M. K. Dadap *et al.*, "Aspect-Based Sentiment Analysis Applied in the News Domain Using Rule-Based Aspect Extraction and BiLSTM," in *2023 IEEE 6th International Conference on Computer and Communication Engineering Technology (CCET)*, IEEE, Aug. 2023, pp. 22–26. doi: 10.1109/CCET59170.2023.10335116.

[9] D. Kun Indarta and A. Romadhony, "Aspect and Opinion Extraction of Indonesian Lipsticks Product Reviews using Conditional Random Field (CRF)," in *2021 13th International Conference on Knowledge and Smart Technology (KST)*, IEEE, Jan. 2021, pp. 113–117. doi: 10.1109/KST51265.2021.9415829.

[10] T. U. Tran, H. T. Thi Hoang, and H. X. Huynh, "Aspect Extraction with Bidirectional GRU and CRF," in *2019 IEEE-RIVF International Conference on Computing and Communication Technologies (RIVF)*, IEEE, Mar. 2019, pp. 1–5. doi: 10.1109/RIVF.2019.8713663.

[11] A. F. Pathan and C. Prakash, "Unsupervised Aspect Extraction Algorithm for opinion mining using topic modeling," *Glob. Transitions Proc.*, vol. 2, no. 2, pp. 492–499, Nov. 2021, doi: 10.1016/j.gltp.2021.08.005.

[12] S. J. Das and B. Chakraborty, "An Approach for Automatic Aspect Extraction by Latent Dirichlet Allocation," in *2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST)*, IEEE, Oct. 2019, pp. 1–6. doi: 10.1109/ICAwST.2019.8923417.

[13] S. Li, R. Wang, and G. Zhou, "Opinion Target Extraction Using a Shallow Semantic Parsing Framework," *Proc. AAAI Conf. Artif. Intell.*, vol. 26, no. 1, pp. 1671–1677, Sep. 2021, doi: 10.1609/aaai.v26i1.8346.

[14] D. E. Cahyani, F. F. Setyawan, A. D. Hariadi, L. Gumilar, and A. K. Junoh, "Comparison of Regression Methods for Estimation of State-of-Health in Lithium-Ion Batteries," in *2023 International Conference on Electrical and Information Technology (IEIT)*, IEEE, Sep. 2023, pp. 202–206. doi: 10.1109/IEIT59852.2023.10335524.

[15] E. Brilliandy, H. Lucky, A. Hartanto, D. Suhartono, and M. Nurzaki, "Using Regression to Predict Number of Tourism in Indonesia based of Global COVID-19 Cases," in *2022 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, IEEE, Sep. 2022, pp. 310–315. doi: 10.1109/AiDAS56890.2022.9918731.

[16] I. Sabilirrasyad, Z. Hasan, and M. Hermansyah, "Sentiment Analysis of Twitter Discussions on Rafael Alun: Multinomial Naïve Bayes and Decision Tree Approach," *PROCEEDING Int. Conf. Econ. Bus. Inf. Technol.*, vol. 4, pp. 803–809, Jul. 2023, doi: 10.31967/prmandala.v4i0.827.

[17] R. Gusdiana, I. Alfian, and C. Juliane, "IMPLEMENTATION OF TEXT PROCESSING FOR SENTIMENT ANALYSIS OF TAX PAYMENT INTEREST AFTER THE 'RUBICON' PHENOMENON," *J. Tek. Inform.*, vol. 4, no. 5, pp. 1157–1164, Oct. 2023, doi: 10.52436/1.jutif.2023.4.5.1014.

[18] D. Gunawan, H. P. Siregar, and O. Salim Sitompul, "Identifying Sentence Structure in Bahasa Indonesia by Using POS Tag and LALR Parser," in *2019 5th International Conference on Computing Engineering and Design (ICCED)*, IEEE, Apr. 2019, pp. 1–5. doi: 10.1109/ICCED46541.2019.9161125.

[19] F. Z. Ruskanda, A. Purwarianti, and D. H. Widyantoro, "Dependency-Based Feature and Pairwise Classifier for Cross-Domain Aspect Term Extraction," *SSRN Electron. J.*, 2022, doi: 10.2139/ssrn.4156881.

[20] H. Yufeng and Z. Fenglu, "A Rich-label Constituency Tree for Constituency Parsing," in *2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, IEEE, May 2019, pp. 186–190. doi: 10.1109/ITAIC.2019.8785583.

[21] A. S. Shafie, N. M. Sharef, M. A. Azmi Murad, and A. Azman, "Aspect Extraction Performance with POS Tag Pattern of Dependency Relation in Aspect-based Sentiment Analysis," in *2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP)*, IEEE, Mar. 2018, pp. 1–6. doi: 10.1109/INFRKM.2018.8464692.

[22] A. Khalid, A. Sundararajan, and A. Sarwat, "A Statistical out-of-Sample Forecast to Estimate Lithium-Ion Parameters That Determine State of Charge," *ECS Meet. Abstr.*, vol. MA2019-04, no. 4, pp. 208–208, Jun. 2019, doi: 10.1149/MA2019-04/4/208.

[23] "KBBI VI Daring." Accessed: Mar. 24, 2024. [Online]. Available: https://kbbi.kemdikbud.go.id/entri/aspek

[24] "KBBI VI Daring." Accessed: Mar. 24, 2024. [Online]. Available: https://kbbi.kemdikbud.go.id/entri/opini