DEVELOPMENT OF APPLICATION PROVIDING PUBLIC'S PERSPECTIVES ON OFFICIAL STATISTICAL INDICATORS

Novi Kanadia[1], Siti Mariyah[2]

[1] Statistics of Kampar Regency, Riau, Indonesia
[2] STIS Polytechnic of Statistics, Jakarta, Indonesia
e-mail: [1]novi.kanadia@bps.go.id, [2]sitimariyah@stis.ac.id

Abstract

BPS-Statistics Indonesia, as an official data producer, puts data quality as a top priority. Public acceptance and trust in data reflect data reliability which is one of the data quality indicators. The existing survey that collects user acceptance, trust, and perspective to data produced by BPS can only reach data users in a limited number. Perspectives from a large number of data users cannot be captured. This research aims to build an application that collects and provides users' perspectives and sentiment to official statistics sourced from online news. One feature in this application is Named Entity Recognition, which extracts public perspectives in entities such as names, organizations, statistical indicators, quotes or opinions, etc. This application objectively measures the sentiment of news discussing or citing statistical indicators. This application also facilitates BPS to do social network analysis to understand the relationships between fellow data users for each statistical indicator. The implicit goal is to effectively provide insights into how frequently society uses and refers to statistical indicators produced by BPS in any domain and their perspective on data. All models and features provided in this application have been evaluated based on standard performance metrics.

Abstrak

Badan Pusat Statistik (BPS) sebagai produsen data resmi, menempatkan kualitas data sebagai prioritas utama. Penerimaan dan kepercayaan publik terhadap data mencerminkan keandalan data yang merupakan salah satu indikator kualitas data. Survei yang telah ada untuk mengumpulkan penerimaan, kepercayaan, dan perspektif pengguna terhadap data yang dihasilkan oleh BPS hanya dapat menjangkau pengguna data dalam jumlah terbatas. Perspektif dari sejumlah besar pengguna data tidak dapat ditangkap. Penelitian ini bertujuan untuk membangun sebuah aplikasi yang mengumpulkan dan menyediakan perspektif dan sentimen pengguna terhadap statistik resmi yang bersumber dari berita online. Salah satu fitur dalam aplikasi ini adalah Named Entity Recognition (NER), yang mengekstrak perspektif publik dalam bentuk entitas seperti nama, organisasi, indikator statistik, kutipan atau opini, dll. Aplikasi ini secara objektif mengukur sentimen berita yang membahas atau mengutip indikator statistik. Aplikasi ini juga memudahkan BPS untuk melakukan analisis jejaring sosial (Social Network Analysis) untuk memahami hubungan antar sesama pengguna data untuk setiap indikator statistik. Tujuan implisitnya adalah untuk memberikan wawasan tentang seberapa sering masyarakat menggunakan dan mengacu pada indikator statistik yang dihasilkan oleh BPS dalam domain apa pun dan perspektif mereka terhadap data secara efektif. Semua model dan fitur yang disediakan dalam aplikasi ini telah dievaluasi berdasarkan metrik kinerja standar.

INTRODUCTION

BPS-Statistics Indonesia is the official data producer in Indonesia. BPS' vision is to be the provider of qualified statistical data for advanced Indonesia. To realize this vision, BPS has been conducting monitoring to perceive users' perspectives against data reliability. One of the monitoring is to identify data users' needs and satisfaction to improve the quality of data and public services. BPS data users are anyone, it can be the government or the public who use BPS data. Data users usually use BPS data for research and policy studies. Monitoring BPS data for research and policy studies by individuals, groups, or organizations can be seen from the amount of BPS data downloaded from the official BPS website and data requests directly to BPS. However, this monitoring is still quantitative and requires qualitative monitoring that can capture the data users' perspective.

The existing method to collect data users' perspectives is The Data Needs Survey (SKD). However, the survey is limited to data users who request data directly from BPS or fill out the SKD questionnaire. In addition, the cost of conducting a survey is expensive and requires considerable time to obtain data. On the other hand, there are opportunities from online news which the number of news and sources is increasing day by day. Online news usually contains news related to government policies, which states BPS data on news text.

Based on the analysis result of the SKD, it is stated that the primary data users of BPS are dominated by consumers who come from ministries and government institutions, students and educational institutions, the private sector, BUMN, and BUMD. Public figures in ministries and government institutions who are often reported on online news are also the primary data users. Therefore, the users' perspective can be seen from online news, including opinions or quotes from public figures regarding BPS data.

In practice, the Public Relations Bureau of BPS has been monitoring or evaluating news about official statistics by classifying the sentiment of the news to be positive or negative. However, the news sentiment is still classified manually by reading the news one by one. As a result, the sentiments produced have high subjectivity and are less effective in time and resources. The evaluation is still limited to looking at news sentiment and has not captured the data users' perspective of official statistics in depth. The development of web-based applications to capture data users' perspective of official statistics on online news is a solution to overcome the existing constraints. BPS data users, namely government public figures, usually give their perspectives or responses to online news. The response can be seen from the direct and indirect quotations in the news text. So that the application built is useful for evaluating the response of public figures objectively and effectively by determining who the public figures are responding to, whether the responses given are positive or negative, and how the relationship between public figures against BPS data.

To determine the response of public figures, we need a method that can identify entities in the news text. Named-Entity Recognition (NER) is a method that can identify certain entities from text files, recognize and classify them into certain entities such as names, locations, organizations, numbers, and other categories [1]. NER is used to produce a comprehensive analysis by identifying names, positions, organizations, locations, indicators, and quotes (responses) of public figures in the news text. Name, positions, and organizations entities are useful to see who the public figures are responding to; quote entities are useful to see what the response of public figures to BPS data and indicator entities are useful to see what public figures respond to indicators of BPS data. In addition, news text analysis to capture data users' perspective of official statistics can be done using sentiment analysis and Social Network Analysis (SNA). Sentiment analysis aims to see the

response of individuals to a particular event [2]. This analysis can predict news sentiments in the form of negative, neutral, or positive sentiments. Meanwhile, SNA is a methodology that can study connections or relationships between individuals in social groups [3]. The results of SNA analysis are usually visualized in the form of graphs consisting of node (as actors) and edge (as relations) that can see the tendency to use official statistics. This study aims to build web-based applications that objectively and effectively evaluate public figures' responses in online news. The development of the NER model is carried out to produce a comprehensive analysis, which can identify entities in the news text in the form of names, positions, organizations, locations, indicators, and quotes of public figures. In addition, news text analysis was carried out using sentiment analysis and SNA to find out the perspective of public figures and the relationship between public figures on official statistics.

LITERATURE REVIEW

Monitoring is needed to determine how far the official statistics produced by statistical institutions are useful for data users in each country. The National Bureau of Statistics (NBS) in the Maldives has conducted a review of the National Statistical System (NSS) to produce a comprehensive analysis that can provide recommendations for the development of official statistics in the Maldives [4]. The review is carried out using a questionnaire and distributed to stakeholders that are relevant to official statistics. One of the review results was regarding data sources and data quality in the Maldives.

In addition, NBS in China has used big data for official statistics [5]. They collect data from the internet to produce official statistics. They succeeded in producing CPI (Consumer Price Index) statistics by utilizing online shopping data since 2014, producing domestic retail sales sold on the internet, and predicting agricultural area and output using satellite image data. For future work, NBS China plans to independently produce public opinion surveys based on data from the internet. The utilization of online sources is an opportunity in the current era of big data. [6] explains three big data opportunities: big data as a control, a substitution, and a new product. The opportunities big data as a new product means that online sources can provide statistical data that has never been produced before. For example, BPS Indonesia utilizes social media (Twitter) or online news to analyze users' sentiments. This sentiment is useful for knowing data users' perceptions about the quality of statistical data and BPS services to the public.

RESEARCH METHOD

This study aims to build an application for providing data users' perspectives of official statistics, especially from public figures. There are four methods carried out in this study, starting from the data collection methods, data processing method, analysis method, to evaluation method. We collected news using the Kofax Kapow application. The cleaning stage reduced the number of news and produced 9,873 clean ones. The information collected from the news is sources, dates, headlines, and contents. We limited the news sources and only collected from credible sources such as Detik.com, Kompas.com, Antaranews.com, Bisnis.com, and Okezone.com. The keywords used in the news search engines are "BPS", "Badan Pusat Statistik", and 113 social and economic indicators included in the Encyclopedia of BPS Indicators 2011.

In the data processing method, we develop NER to identify entities such as people, places, organizations, etc. that are related to a particular word from a document and can also classify words in different categories [1]. The development of the NER model in this study uses the SpaCy library [7]. SpaCy is open-source software that provides an advanced library for NLP using the Python and Cython programming languages. Some studies show that SpaCy has the fastest syntactic parser. In [8], showed that the accuracy of SpaCy was

only 1% different from the best parser ever. Other more accurate parsers were 20 times slower or more than SpaCy's speed in parsing sentences or tokens per second. In [9] developed a non-monotonic transition system for dependency parsers of SpaCy produced 91.85% accuracy.

SpaCy has a NER feature and provides methods to train and build their own NER models that can be saved and load later. With many features and conveniences provided by SpaCy, we use SpaCy to build NER models to identify entities in the news text with Bahasa Indonesia. The entity is named, position, organization, location, indicator, and quote to evaluate the response of public figures. To train NER models, the collected news data is labeled manually. Labeling is divided into training sets as 450 news and testing sets as many as 50 news (90% training set; 10% testing set). The training set is used to build models, while the testing set is used to evaluate models that have been built. Labeling news entities is done using the GATE Developer v8.5.1 application [10]. The label of the entity consists of "person", "position", "organization", "location", "indicator" and "quote". Training set and testing set labeling are adjusted to the data structure at SpaCy (Figure 1). The NER training data structure consists of news text, start index, end index, and entity label. The entity label consists of person, position, organization, location, indicator, and quote. To improve accuracy, the dataset is separated into six training sets and six testing sets. The NER model that has been built is then loaded to predict the rest of the news that has been collected.

The analytical method we used in this study is sentiment analysis and social network analysis. Sentiment analysis processes and extracts textual data automatically to obtain information, namely the response to a particular event in the form of a positive or negative response [2]. Indonesian sentiment analysis research has been conducted by [11] who built the system and adapted the SentiStrength classifier by [12]. SentiStrength is an algorithm that can determine the sentiment of a text with a lexicon-based approach. SentiStrength Lexicon with default English translated into Indonesian according to [12] permission. To compare the results of system sentiments, labeling was carried out by empowering several human coders. The results of the comparison between the systems that were built, and the labeling of

```
TRAIN_DATA = [('''Jakarta - Badan Pusat Statistik (BPS) mencatat impor periode November
2018 sebesar US$ 16,88 miliar atau Rp 244,76 triliun (kurs Rp 14.500). Jumlah ini tumbuh
11,68% dibandingkan periode yang sama tahun 2017. Dari data BPS, impor non migas turut
mendorong peningkatan impor. Misalnya impor sayuran periode November 2018 tercatat naik US$
57 juta atau 140%. Paling banyak impor berasal dari China sebanyak 94 ton atau US$ 81 juta
pada November. Namun BPS tidak merinci jenis sayuran yang paling banyak diimpor dari negara
Tirai Bambu tersebut.  Kemudian Indonesia juga masih mengimpor sayuran dari Ethiopia
sebanyak 3 ton atau sebesar US$ 3,04 juta. Kemudian dari Australia sebanyak 1,4 ton atau
senilai US$ 1,4 juta. BPS menyebutkan jenis sayuran impor yang masuk ke Indonesia adalah
bawang putih senilai US$ 78 juta naik 140,49% dibandingkan periode Oktober 2018 US$ 22,7
juta. Selanjutnya untuk kacang tercatat US$ 5,1 juta tumbuh 39,73% dibandingkan periode
Oktober. Lalu ada impor kentang sebesar US$ 1,9 juta tumbuh 98,27% dari periode Oktober US$
1,1 juta. Lalu bawang bombai US$ 5,9 juta tumbuh 4,75% dibanding Oktober 2018 US$ 5,7
juta.Lalu impor minuman juga mengalami kenaikan mencapai SU$ 75,2 juta atau naik 470% dari
periode Oktober 2018. Selanjutnya disusul oleh impor nikel yang meningkat 404%, naik US$
45,8 juta dibandingkan bulan sebelumnya. Selanjutnya, impor lokomotif dan peralatan kereta
api juga mengalami kenaikan US$ 21,5 juta atau naik 58% di November 2018.Namun demikian,
ada juga sejumlah barang yang mengalami penurunan impor, seperti kulit berbulu yang turun
90,98% atau senilai US$ 1,24 juta di November 2018 dibandingkan bulan sebelumnya.Disusul
oleh bahan bakar mineral yang turun 5% atau senilai US$ 1,41 miliar di November 2018,
begitu juga dengan impor sutra yang turun 51% atau senilai US$ 542 juta.''',
{'entities': [(47, 52, 'indicator'),(224, 229, 'indicator'),(268, 273, 'indicator'),(284,
289, 'indicator'),(371, 376, 'indicator'),(747, 752, 'indicator'),(979, 984, 'indicator'),(
1277, 1282, 'indicator'),(1373, 1378, 'indicator'),(1549, 1554, 'indicator'),(1783, 1788,
'indicator'),(1149, 1154, 'indicator')]})]
```

Figure 1. The NER training data structure of indicator entity

the human coder showed an accuracy of 57.33%.

Two approach methods can be used in sentiment analysis, namely supervised and lexicon-based. In this study, we used lexicon-based sentiment analysis because it's easier to use and develop. Lexicon-based is an approach that utilizes the sentiment of the lexicon. The Lexicon is considered an important indicator for sentiment, which is called an opinion word. SentiStrength is one of the lexicon-based algorithms that estimate positive and negative strength in a short text, even for informal languages, which are determined based on the theory of human-level accuracy [12].

Meanwhile, social network analysis is a methodology for studying the connections and behavior of individuals in social groups [3]. Individuals or actors in SNA are usually represented by nodes or dots. While the relationship between actors is represented by edge or line. Edge can describe the relationship between two individuals, such as acquaintances, likes, beliefs, or dislikes. The pattern produced between nodes and edges is usually visualized in the form of a graph. For SNA, we visualized it using D3.js in the form of a force-directed graph. D3.js needs a JSON dictionary to make a force-directed graph. Its JSON consists of nodes and links. Nodes contain "id" represents actor names and "group" represents in what quarter did actor give the response. While links contain "source", "target" and "value". "source" and "target" represent the relation between the actor, and "value" represents the number of the relation between the actors. NER evaluation is done by using hold-out validation. Hold-out validation is the simplest model evaluation technique. In this technique, the dataset is labeled and divided into two parts, that is the training set and the testing set. Training sets are used to train the models and testing sets are used to evaluate the model [13].

According to [13], there are four stages in hold-out validation to produce the final model. First, label the dataset and divide it into training sets and testing sets. The distribution of training and testing sets, in general, is 60:40, 70:30, 80:20, or even 90:10. Second, determining learning algorithms and training the training set to produce models. Third, predicting using the model and evaluating it using testing sets to see the performance of the model. Model performance can be seen from the value of precision, recall, and F1 score. Last, training uses a learning algorithm in the entire dataset so that the final model is produced. Precision, recall, and F1 score are measuring instruments used to evaluate the performance of a model. Precision, recall, and F1 scores are obtained using a confusion matrix. Precision is a comparison between labels that are correctly predicted positively on all positive predictive results. The recall is a comparison between labels that are correctly predicted positively on all positive actual data. While the F1 score is the harmonic average that is adjusted from precision and recall.

RESULT AND DISCUSSION

The result of this study is a web-based application that can capture data users' perspectives of official statistics. In this study, we not only build applications but also analyze the result of the dataset we run in the application. In addition, the NER model that we built is also evaluated and discussed here.

We have built an application using PHP and Python as the back end, and CSS and JavaScript as the front end. The framework we used in this application is CodeIgniter and Flask. In this study, we design the application to accommodate the need to capture data users' perspectives of official statistics on online news into menus in the application. There are five menus in the application, they are "Dashboard", "News", "NER", "Sentiment", and "Social Network".

The dashboard menu (Figure 2) contains overall descriptive statistics, like total news, predicted news which means predictive data using NER and sentiment analysis, news sources, negative sentiments, neutral sentiments, and positive sentiments. On the Dashboard menu, there

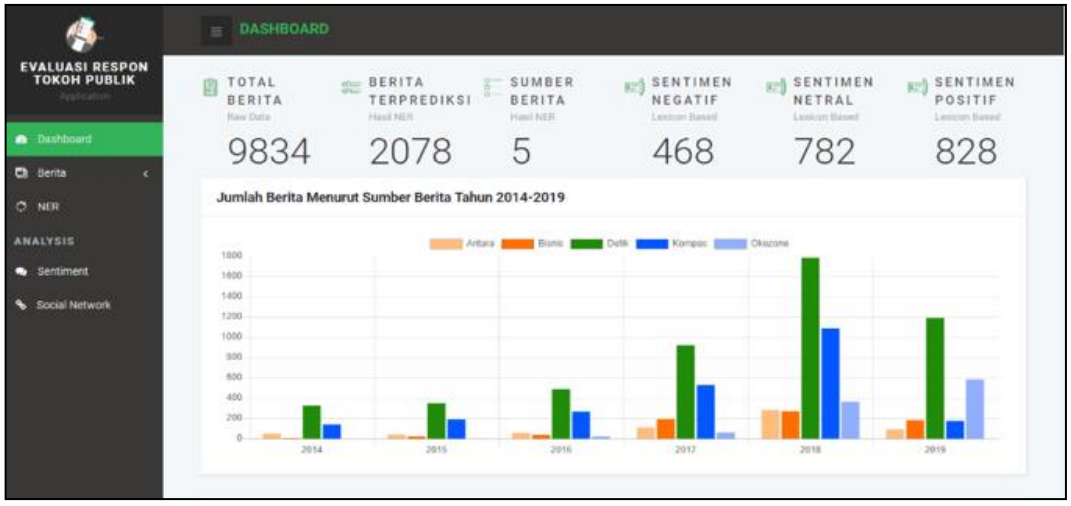

Figure 2. The implementation of the "Dashboard" menu

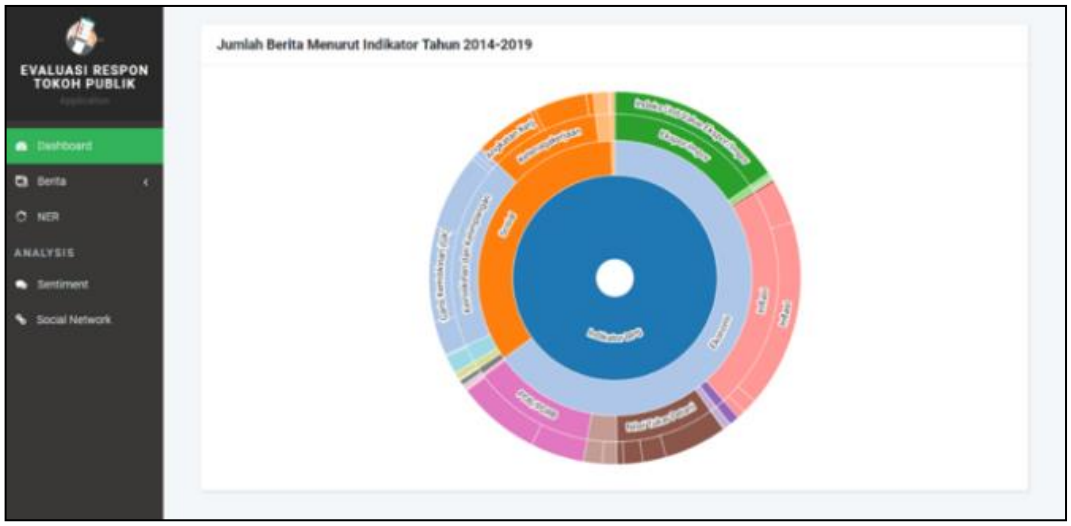

Figure 3. The implementation of the "Dashboard" menu

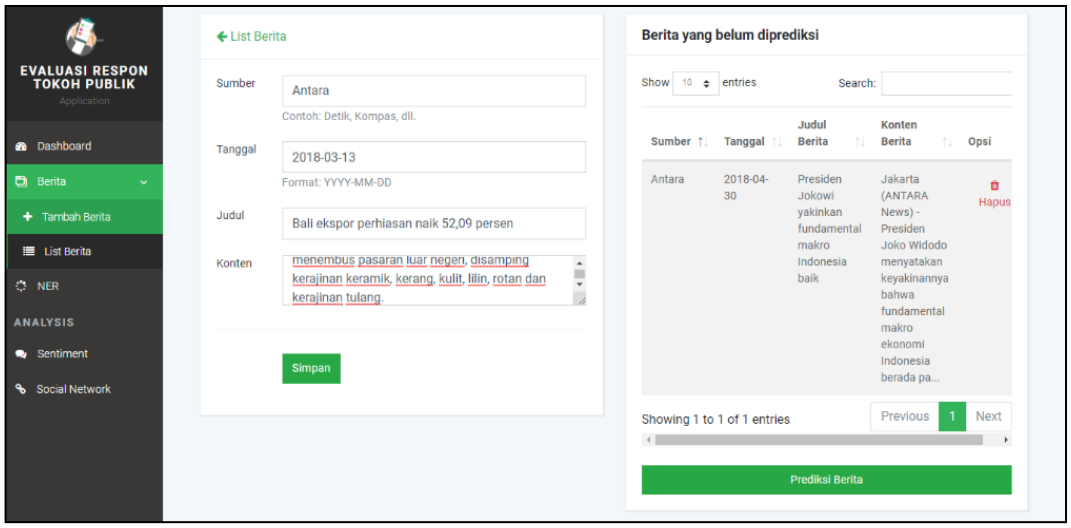

Figure 4. The implementation of the "Add News" menu

are two graphs, that is the bar chart and sunburst zoomable. The bar chart shows the amount of news according to news sources in 2014 - 2019. Whereas the sunburst zoomable (Figure 3) shows the number of news according to the indicators 2014 - 2019. The news menu contains the function of adding news and predicting news (Figure

4). Add news function allows users to add news manually to the database. While predict news function allows users to predict the news.

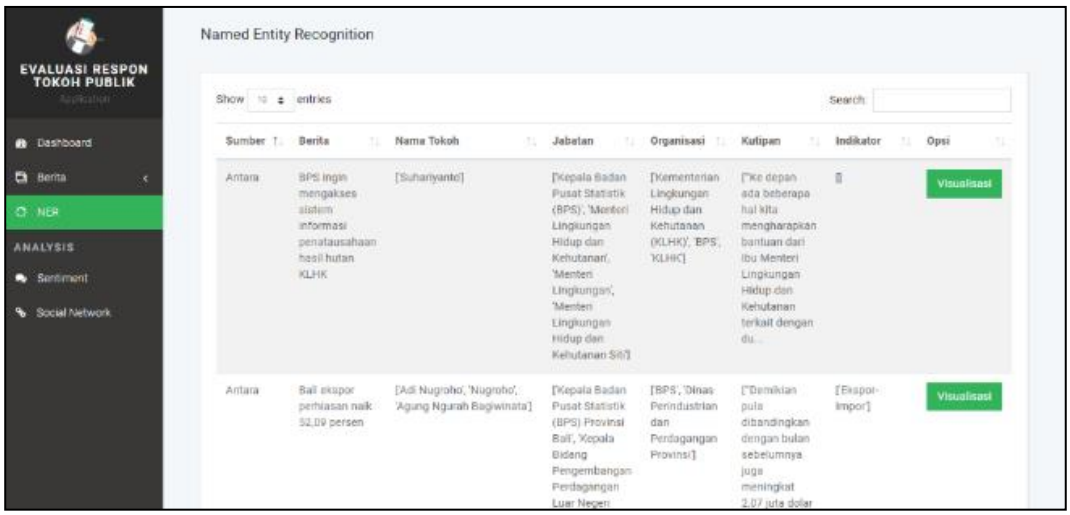

Figure 5. The implementation of the "Dashboard" menu

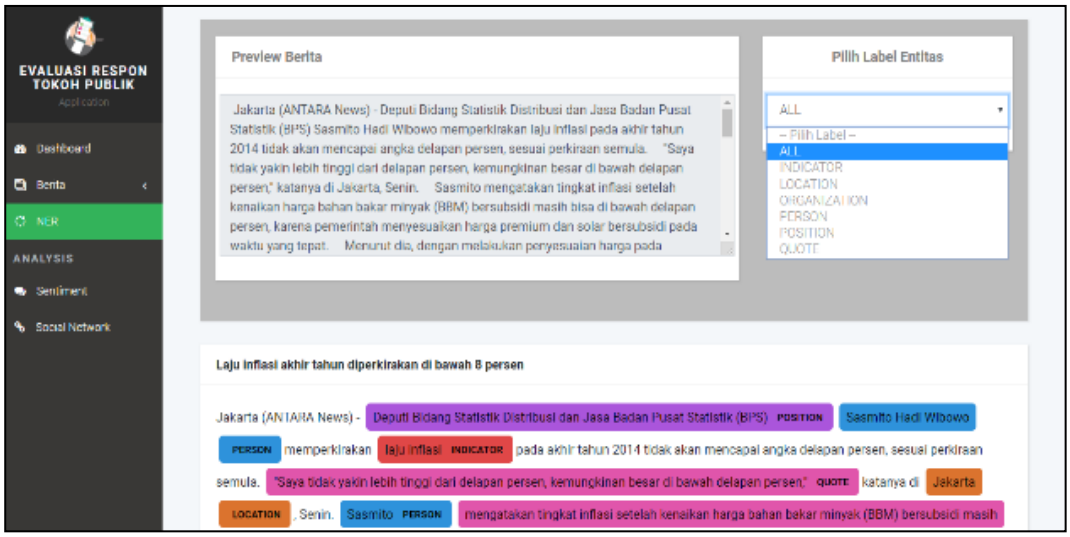

Figure 6. The implementation of the "Dashboard" menu

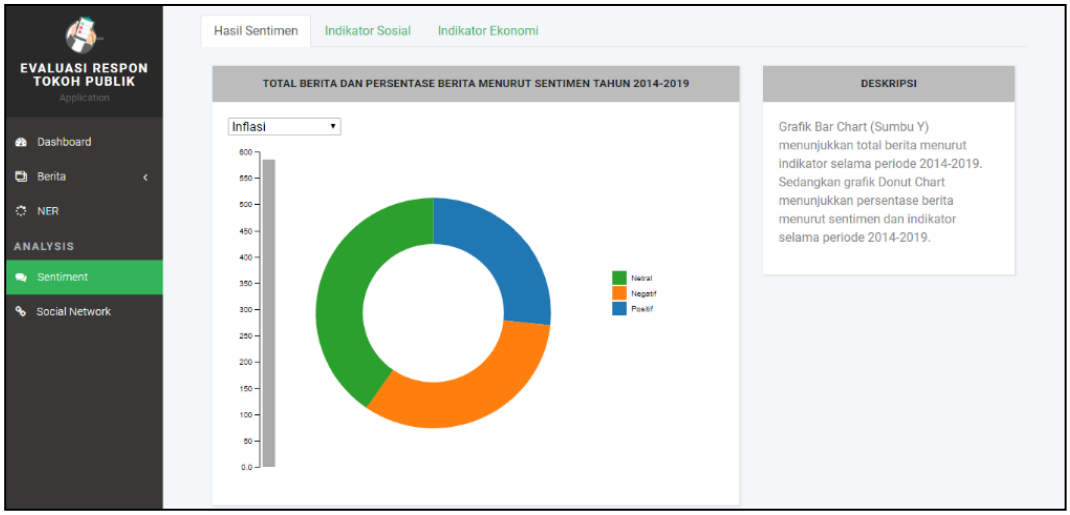

Figure 7. The implementation of the "Add News" menu

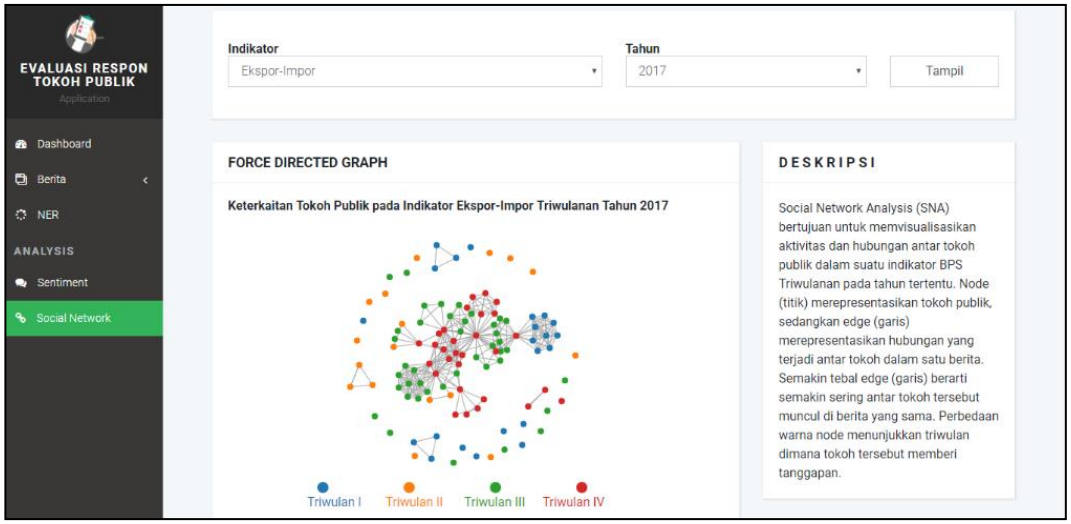

Figure 8. The implementation of the "Social Network" menu

The NER menu (Figure 5) contains a table of news predictions using NER models that have been built. The table consists of sources, news, public figures' names, positions, organizations, location, indicators, and quotes. In the options column, there is a "Visualization" button that contains the NER visualization page shown in Figure 6. NER visualization in this study was conducted using displaCy visualizer, which is one of the libraries of SpaCy. On the NER visualization page, there is a news preview and the option label of the entity want to visualize.

The sentiment menu (Figure 7) consists of three tabs, namely the sentiments result, social indicators, and economic indicators. The sentiments result tab contains a dynamic visualization from the percentage of news sentiments according to the selected indicators using the donut chart.

In the social network menu (Figure 8), users can choose the indicator and year to visualize. By pressing the "Show" button, SNA visualization will appear using a force-directed graph along with a description of the graph. The graph was built using D3.js. In force-directed graphs, the node represents a public figure, the edge represents the relation between the public figures and the different color shows in what quarter the public figure gives a response.

Sentiment analysis and SNA were used to analyze news in this study. We discuss indicators of poverty and inequality in social indicators, which have the highest reporting compared to other indicators we collected, as many as 384 news. According to sentiment analysis results shown in Figure 9, the indicators of poverty and inequality had 86 news (22 percent) of negative sentiment, 154 news (40 percent) of neutral sentiment, and 144 news (38 percent) of positive sentiment.

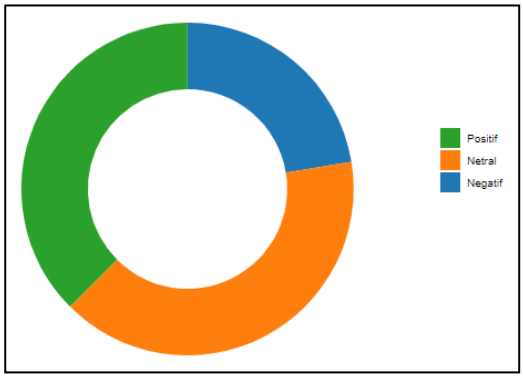

Figure 9. The percentage of news poverty and inequality indicator based on sentiment

According to NER prediction results, there were 609 public figures and 376 organizations included in the news of poverty and inequality indicators. In Figure 10, we can see the top 5 public figures and top 5 organizations on poverty and inequality indicators based on sentiments. In the negative, neutral and positive

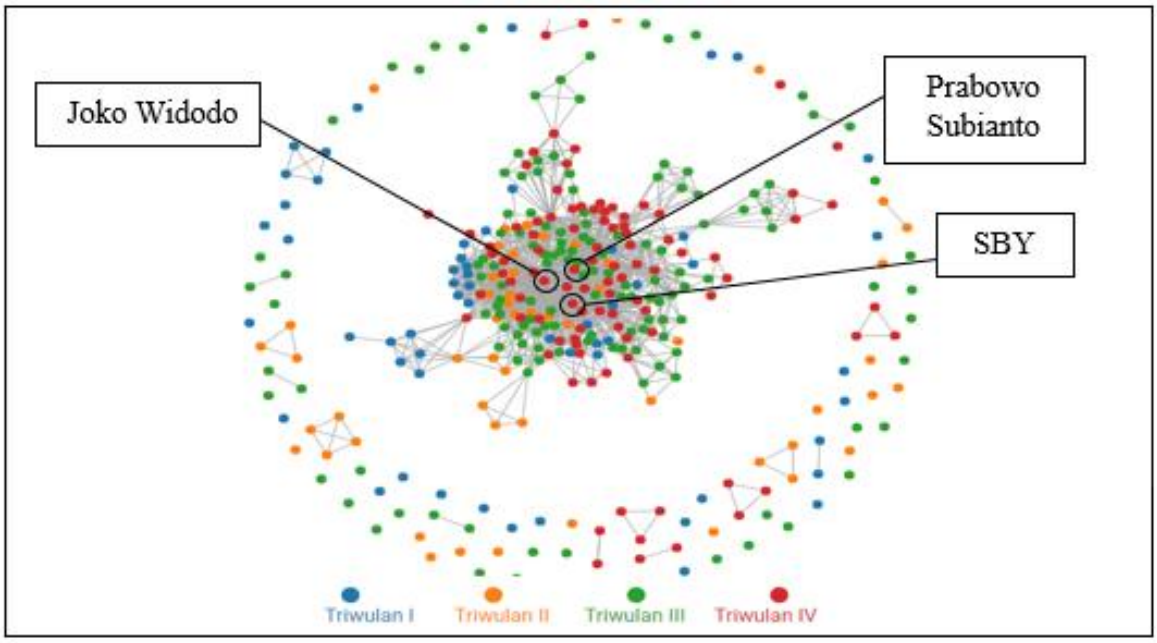

Figure 11. Relationship between public figures on poverty and inequality indicators quarterly in 2018

sentiment, Jokowi became the most emerging figure, while the World Bank became the most emerging organization.

Figure 11 shows the relationship between public figures on poverty and inequality indicators quarterly in 2018. Joko Widodo (President of Indonesia in 2014-2019) has the most related to other public figures, followed by Prabowo Subianto (Candidate for President in 2019-2024) and SBY (President of Indonesia in 2004-2014).

A large number of relations these three figures shows the big role each public figure of poverty and inequality. In 2018, it is known that the poverty rate in Indonesia reached single digits of 9.82 percent as of March 2018. This is certainly a big achievement for President Joko Widodo because the poverty rate is the lowest in history, followed by a decline in the Gini ratio. This achievement is associated with the reign of SBY, where even though the poverty rate has decreased, the Gini ratio has increased. In addition, Prabowo as a presidential candidate also gave many opinions on poverty and inequality. He said that only 1% of Indonesia's population lives well. While the rest or 99%, life is very mediocre. This opinion is much different from the BPS data, so it is widely reported.

The relationship between Joko Widodo and other figures are Prabowo Subianto, Jusuf Kalla (Vice President of Indonesia in 2014-2019), Darmin Nasution (Indonesian Economic Coordinating Minister for 2015-2019), and many other public figures. Based on the quarter, there were 84 figures responded in the first quarter, 76 figures in the second quarter, 342 figures in the third quarter, and 178 figures in the fourth quarter.

The relation between Joko Widodo and Prabowo Subianto occurs because of the presidential election that will be held in 2019, where both have different opinions regarding poverty data. The relation with Jusuf Kalla occurred because, in the reign of President Jokowi and Vice President Jusuf Kalla, both of them continued to strive to reduce poverty. Meanwhile, the relationship with Darmin Nasution occurred because the economic growth that occurred in Indonesia was not in line with the reduction of poverty and inequality at that time.

Based on news source Detik.com (Figure 12), there were 746 or about 36% of news obtained from Detik.com. Based on sentiment results, 162 news (22 percent) news had negative sentiment, 270 news (36 percent) had neutral sentiment and 314

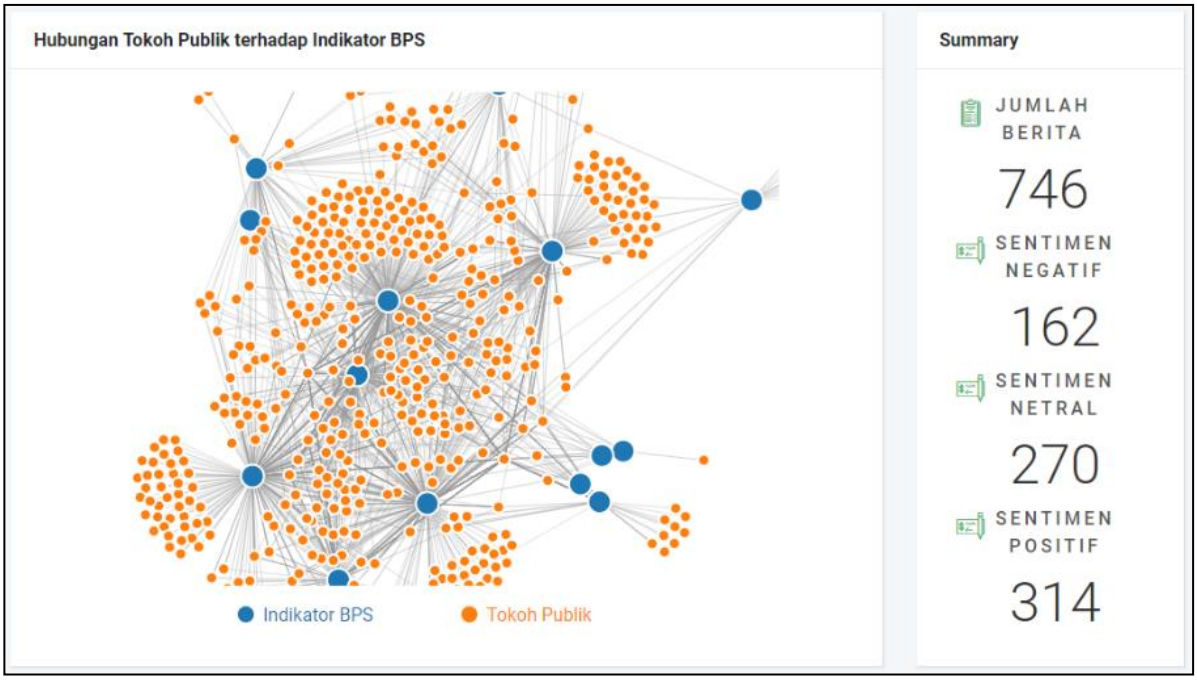

Figure 12. Summary of Detik.com

news (42 percent) had positive sentiment. Detik.com is a news source that has the highest amount of news compared to other news sources.

According to SNA force-directed graph visualization, the poverty and inequality indicators are indicators that have the most relationships with public figures such as Joko Widodo, Prabowo Subianto, Jusuf Kalla, Sri Mulyani (Indonesian Finance Minister), and other figures. Followed by inflation indicators that have relations with public figures such as Suhariyanto (Head of BPS-Statistics Indonesia), Abdullah Abu Bakar (Mayor of Kediri), and other figures.

Table 1. The results of NER model evaluation (%)

The Entity of The NER Model	Precision	Recall	F1 Score
Person	97.55	96.36	96.95
Position	89.55	90.91	90.23
Organization	87.62	72.54	79.37
Location	95.03	89.13	91.99
Indicator	76.44	80.90	78.61
Quote	92.52	88.89	90.67

We use hold-out validation to evaluate the performance of the model that has been built. Evaluation of model performance will be seen through a measuring instrument, namely precision, recall, and F1 score. After the NER model of each entity label is formed, an evaluation is carried out using the SpaCy command line, with results that can be seen in Table 1.

The table above shows that the NER model of person entity has the greatest value, precision 97.55%; recall 96.36%, and f1 score 96.95%. While the smallest value is the NER model of indicator entity with precision 76.44%; recall 80.90%, and f1 score 78.61%. From the NER model of person entity, the precision value means that 97.55% of the person entities in the predicted data are correctly predicted. The recall value means that 96.36% of the person entity in the actual data (test data) is correctly predicted. Whereas, the value of f1 score 96.95% indicates the results of the NER model of the person entity is already good. The more the f1 score is close to 100%, the better the resulting model will be.

Organization and indicator entities have a lower f1 score than other entities.

This can be caused by a training set that does not cover all existing organizations and BPS indicators. There are many organizations both domestic and overseas that are associated with BPS. However, not all of these organizations are included in the training set, so the model performance will be reduced when there are organizations that are not covered (not similar) to the models that have been built. However, the value of the f1 score tends to be close to 100%, which indicates that the model is good enough.

From the explanation above, it can be concluded that the evaluation of the NER model for all entities is quite high because it is above 50%. Thus, the NER model that has been built is good and feasible to predict the entity in the news.

CONCLUSION

Evaluation of BPS data (official statistics) is important to find out how much BPS data contributes to data users. The evaluation can be seen from the response of public figures to the quality of BPS data from online news. We have built an application that can capture the data users' perspective of official statistics. The development of the NER model to predict the entity name, position, organization, location, indicators, and quotes of public figures on the news text has also been successfully built. The evaluation using hold-out validation results in a high value of precision, recall, and f1 scores. This shows that the NER model built is good and feasible to predict the entity in the news text. In addition, analysis has been carried out using sentiment analysis to capture public perspective as news and quotation sentiment, and SNA visualization using force-directed graphs to see the relationship between public figures based on BPS indicators and the relationship between public figures and BPS indicators based on news sources.

The analysis results based on social indicators show that news sentiment tends to have neutral sentiments. While based on news sources, news sentiment tends to have positive sentiments. This shows that the public perspective of official statistics is good enough. In addition, it can be seen that Jokowi is the most influential figure in the BPS indicator as to the President of Indonesia because he has the most relation compared to other figures.

The future work in this study is to connect crawling robots to applications built and add more diverse news sources to evaluate the responses of public figures in real-time. Furthermore, the development NER model is expected to be able to overcome the limitations of research to identify full names and nicknames as the same person.

REFERENCES

[1] Vora, Komil, Avani Vasant, and Rachit Adhvaryu 2016 Named Entity Recognition and Classification for Gujarati Language In Proceedings of International Conference on Advance in Computing, Communications and Informatics pp 2269-2272.

[2] Nurdiansyah, Yanuar, Saiful Bukhori, and Rahmad Hidayat 2018 Sentiment Analysis System for Movie Review in Bahasa Indonesia Using Naive Bayes Classifier Method Journal of Physic: Conference Series

[3] Clifton, Allan and Gregory D. Webster 2017 An Introduction to Social Network Analysis for Personality and Social Psychologists Social Psychological and Personality Science pp 1-12

[4] Shazna, Ashiyanth and Aminath Musfiqa Ibrahim 2019 Review of the National Statistical System in the Maldives Asia-Pacific Economic Statistics Weeks

[5] Yajing, Cong 2017 Big Data for Official Statistics: Experience from China Asia-Pacific Economic Statistics Week.

[6] Mariyah, Siti 2014 Identification of Big Data Opportunities and Challenges in Statistics Indonesia In Proceedings of ICSS Conference

[7] Honnibal, Matthew SpaCy - Industrial-strength Natural Language

Processing in Python https://spacy.io/ (Accessed March 2019)
[8] Choi, Jinho D., Joel Tetreault, and Amanda Stent 2015 It Depends: Dependency Parser Comparison Using A Web-based Evaluation Tool In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing pp 387–396.
[9] Honnibal, Matthew and Mark Johnson 2015 An Improved Non-monotomic Transition System for Dependency Parsing In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing pp 1373–1378
[10] Cunningham, et al 2013 Getting More Out of Biomedical Documents with GATE's Full Lifecycle Open Source Text Analytics PLoS Computational Biology
[11] Sianipar, Raisa and Erwin Budi Setiawan 2015 Pendeteksi Kekuatan Sentimen pada Teks Tweet Berbahasan Indonesia Menggunakan SentiStrength In e-Proceeding of Engineering **Vol. 2, No.3**
[12] CyberEmotion SentiStrength http://sentistrength.wlv.ac.uk/ (Accessed March 2019)
[13] Raschka, Sebastian 2018 Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning