Big Data for Small Area Estimation: Happiness Index with Twitter Data

S D Aziz[1], A Ubaidillah[2]

[1]Statistics Indonesia, Jakarta, Indonesia
[2]Polytechnic of Statistics STIS, Jakarta, Indonesia

*Corresponding author's e-mail: sheerin.dahwan@bps.go.id

Abstract. Data availability for small area level is one of the keys to the success of regional development. However, direct estimation of small areas can produce high error due to inadequate sample sizes so the estimation is not reliable. One of alternative solution to this problem is to use the Small Area Estimation (SAE) method which can improve precision by "borrows strength" of the corresponding region information or auxiliary variable information that is strongly related to the response variable. This study uses two SAE models, namely SAE EBLUP Fay-Herriot model with auxiliary variables Podes data and SAE with Error Measurement with auxiliary variable Twitter data. Estimation results using the SAE method are better than direct estimates. This is shown by the RSE value which produced from SAE method, both the EBLUP model and Measurement Error, is smaller than the direct estimate. Therefore, big data can be used as an alternative variable in the SAE model because the data is available in real-time, covers up to the smallest area, and relatively low cost.

1. Introduction

Today, the need for data at the micro level is increasing. Moreover, Indonesia implements a decentralized system in running its government so that each region requires data to plan its development. The role of regional development in achieving the SDGs in Indonesia is certainly very important. However, currently some of the SDGs indicators for this level have not been fulfilled.

The happiness index is one of the indicators that BPS has not fulfilled from the 17 SDGs goals, namely strengthening implementation facilities and revitalizing global partnerships for sustainable development [1]. This index is an attempt by BPS to follow the international focus on measuring subjective well-being. This index is composed of three dimensions, namely the dimension of life satisfaction, the dimension of feeling, and the dimension of the meaning of life [2]. To collect the data needed to make up the happiness index, BPS conducted a Survei Pengukuran Tingkat Kebahagiaan (SPTK).

The data collection methods commonly used by BPS so far are through censuses and surveys. The census covers the entire population and the results can be presented to the smallest area. However, the census has limited time, effort, cost, and the variables collected are limited. On the other hand, surveys can cover many variables and are more time, effort and cost efficient than censuses. However, estimation with survey data is often faced with the problem of sample adequacy. Direct estimation with insufficient samples can result in large standard errors [3]. To overcome the problem of sample adequacy, estimation can be done indirectly using the Small Area Estimation (SAE) method.

Small Area Estimation (SAE) is an indirect estimation method that "borrows the power" of the appropriate area information or information on auxiliary variables that have a strong relationship with the observed variables [3]. This method tends to produce a Mean Squared Error (MSE) which is smaller than the direct estimate. To produce a good estimate, the SAE method requires available auxiliary variables up to the level to be estimated. So far, the SAE auxiliary variable commonly used by BPS comes from Podes data. However, Podes data is only collected 3 times in 10 years, so sometimes the use of Podes data as an auxiliary variable for the SAE model becomes less relevant.

On the other hand, big data is an innovative data source that is available in real-time and covers up to the smallest level. These innovative data sources can usually be obtained quickly [4], are relatively inexpensive, and can also produce information about aspects of life that traditional data sources may not capture.

One source of big data that is easily available is Twitter data. Twitter gives developers official access to its data. Based on statistics obtained from the Statista website, in 2017 Twitter had 18.9 million users from Indonesia or around 25.47% of the total social media users. Social media users tend to express the things they experience on their respective accounts, Twitter users are no exception. So, it can be said that almost every tweet that is thrown describes his feelings. Thus, it is not impossible to use Twitter data as an auxiliary variable of the Happiness Index SAE model.

However, the auxiliary variables obtained from the processed Twitter data have errors, while the SAE method assumes that the auxiliary variables are measured without errors. When the auxiliary variables used in the model have errors, estimators who ignore these errors can produce estimates that are worse than direct estimates [5]. Therefore, the model in this study will be formed with a measurement error. In addition to Twitter data, the research will use Podes data as a comparison.

Based on the identification of the problem, the purpose of this research is as follows.

- Estimating the Happiness Index 2017 in Java at the district/city level using the SAE method.
- Comparing the results of the Happiness Index 2017 estimation in Java at the district/city level between direct and indirect estimates using the SAE method.
- Assessing the application of the SAE model by utilizing big data as an auxiliary variable.

This study uses two SAE models, namely the SAE with Measurement Error (SAE ME) model with Twitter data as an auxiliary variable containing errors and SAE EBLUP Fay-Herriot with Podes data as a comparison variable. The research systematics used are as follows:

- Processing Twitter data with text mining to get a happiness score;
- Estimating the Happiness Index using the direct estimation method and the indirect estimation method using the SAE with Measurement Error model and SAE EBLUP Fay-Herriot;
- Examine the use of big data by comparing the results of direct and indirect estimates with the SAE model.

2. Data Preprocessing

Secondary data was also collected from Twitter data for the period January to April 2017 by ignoring languages other than Indonesian and English. Twitter data collection is done using the ‘twint’ library in the Python programming language [6]. The attributes of the data used in this study consist of 'id', 'tweets', and 'place'. 'id' is a unique number that represents each tweet, 'tweets' is the content of the tweets in the form of text, and 'place' is where the tweets came from. The Twitter data that has been collected is filtered so that only tweets are found on the island of Java. Then, duplicate tweets that don't have the 'place' attribute are removed. Next, the 'place' attribute is classified into districts/cities. Then, several preprocessing stages were carried out. The preprocessing stages of the text are as follows.

1. Replacement of characters with capital letters to lowercase/regular characters;
2. Deletion of urls, symbols, usernames, hashtags, punctuation marks, numbers, and double words because only text in the form of words is analyzed;
3. Separation of words into arrays for cleaning each word on each tweet;
4. Standardization of words to match the available dictionaries;
5. Removal of affixes and stopwords that are not needed in the analysis so as to produce basic words;

6. Substitution of negative words (words that contain the word ‘not’) with the antonym of the word after it (e.g. not happy to be sad) to reduce errors in scoring because scoring is done lexically.

id	place	tweet	new_place	kabkot	new_tweet
8.53E+17	Baros, Indonesia	alhamdulillah,, msh dapat kado setelah 2 bulan lebih nikah nuhun https://www.instagram.com/p/BS4-aM2Fk-Qat8e5E0fRaEXzlem7VQYi6efZFw0/Â â€¦	baros	kota sukabumi	alhamdulillah masih dapat kado telah bulan lebih nikah nuhun
9.48E+17	Kota Sukabumi, Jawa Barat	Di malam spesial aku sendiri pic.twitter.com/fvjNBm5q4s	kota sukabumi	kota sukabumi	di malam spesial aku sendiri
8.80E+17	Pilangkenceng, Indonesia	Halal bi halal dlu y dek di halalinya nnti dlu• @missfitria_22 @ Desa Kuwu Kec. Balerejo https://www.instagram.com/p/BV28ydXlJdk/Â	pilangkenceng	madiun	halal bi halal dulu ya adik di halal nanti dulu desa kuwu camat balerejo
9.43E+17	Baros, Indonesia	Karena hdup slalu mengajarkan... Jika kta benar2 tulus https://www.instagram.com/p/Bc0bkOdBZOFM0Otzxmbun5levk934HYGxHZkN00/Â	baros	kota sukabumi	karena hidup selalu ajar jika kita tulus
8.97E+17	Baros, Indonesia	Seenggak-nya hidup aku gk penuh drama . . #latepost #me #happy #yellow #maroonâ€¦ https://www.instagram.com/p/BXvHbPIDtjq/Â	baros	kota sukabumi	enggak hidup aku drama latepost me bahagia yellow maroon longgar

Figure 1. Preprocessing

Then, the clean data is scored using the LabMT library [7]. The LabMT library was chosen because it can produce relevant results [8]. After scoring is successful, the outlier data is removed. Data that does not contain outliers are 455,132 tweets. Then, the data is aggregated into district/city level with an average data concentration measure. The result is used as an auxiliary variable (X) in the SAE method.

new_tweet	score
selamat siang para sahabat kuliner selamat nikmat weekend akhir di tahun selamat	7.206667
norak full haha jalan labuh	5.52
silakan lanjut apa perlu aku tuntun agar kamu tidak nikung dasartukangnikung	3.96
waktu tuju senja selalu saja mampu buat aku untuk jatuh cinta ya aku suka senja	6.2
perihal cinta biar jadi urus dengan agar kelak rangkai betul	7.3
ajar hadap kecewa supaya tidak gunung	4.266667
dulu sempat kecewa pergi ke pantai ini karena mungkin salah jalan	4.333333
gadis satu ini tiap libur pasti main ke trans studio bandung	6.076667
terus senyum buah hati ku santa sea waterpark	6.626667
yeaayy happy anniversary vovoevitasarii	7.18

Figure 2. Scoring

3. Estimated Happiness Index in Java Island in 2017

The Happiness Index is a subjective measure that describes the level of well-being of the objective conditions of various domains of human life taking into account the feelings and meaning of one's life. The Happiness Index is a composite index consisting of three dimensions, namely life satisfaction which is divided into sub-dimensions of personal life satisfaction and social life satisfaction, meaning of life (eudaimonia), and feelings (affect). The formula used to calculate the Happiness Index is as follows. [9]

$$I_{Kebahagiaan} = \frac{w_1 * I_{Kepuasaan\ Hidup} + w_2 * I_{Perasaan} + w_3 * I_{Makna\ Hidup}}{w_1 + w_2 + w_3} \quad (1)$$

The estimation of the Happiness Index in 2017 was carried out by applying the Small Area Estimation (SAE) method EBLUP Fay-Herriot and SAE with Measurement Error (SAE ME) developed by Ybarra and Lohr. The EBLUP Fay-Herriot model uses Podes 2018 data as an auxiliary variable, while the SAE ME model uses data processed by Twitter called Twitter scoring as an auxiliary variable. The two models will estimate the Happiness Index 2017 in Java at the district/city level.

3.1 Model SAE EBLUP Fay-Herriot

After direct estimation is done, the resulting Happiness Index is tested for normality using the Shapiro Wilk test. Based on the test, the resulting p-value is 0.34, it can be said that the assumption of normality is met. Then, the selection of auxiliary variables sourced from the Podes 2018 data was carried out using the stepwise elimination method. The selection of auxiliary variables is carried out with the aim of obtaining auxiliary variables that can increase accuracy in estimation. Then, a multicollinearity test was conducted to determine whether there was a correlation between the auxiliary variables used. The multicollinearity problem does not occur in a good regression model. Therefore, it is necessary to check the multicollinearity in the model. To detect the presence of multicollinearity in the model, can be seen from the value of the Variance Inflation Factor (VIF). The auxiliary variable with VIF > 10 indicates multicollinearity. The auxiliary variables which indicated there was multicollinearity were omitted from the model. Furthermore, the Fay-Herriot SAE EBLUP modeling was carried out using the 'sae' package in the R application [10]. The modelling results were reviewed for the significance of the auxiliary variables included in the model. With a significance level of 5%, non-significant auxiliary variables were excluded from the model.

The selection of these variables produces 14 auxiliary variables that will be used in the SAE EBLUP Fay-Herriot model. The auxiliary variables include the number of brawls between community groups between villages (X1), the number of inter-ethnic brawls (X2), the occurrence of criminal acts of theft (X3), the occurrence of criminal acts of fraud/embezzlement (X4), the incidence of rape / decency crime (X5), criminal acts of drug abuse/trafficking (X6), the number of flood incidents in 2017 (X7), the number of hurricanes/hurricanes/typhoons in 2017 (X8), the number of the incidence of volcanic eruptions in 2017 (X9), the number of forest and land fires in 2017 (X10), the number of high school education levels (X11), the number of Polindes (Pondok Bersalin Desa) (X12), the number of diphtheria sufferers (X13), and the number of village community institution: ‘Karang Taruna’ (X14). With these 14 auxiliary variables, the SAE EBLUP Fay-Herriot modeling was carried out again. The equation of the model is as follows: [3]

$$y_i = X_i^T \beta + v_i + e_i, , \quad i = 1, \dots, m \quad (2)$$

Where $y_i = Fay - Herriot\ estimator$
$X_i = auxiliry\ variable$
$\beta = regression\ coefficient$
$v_i = random\ effect\ area$
$e_i = sampling\ error$

3.2 SAE Model with Measurement Error

Before it can be used as an auxiliary variable, Twitter data that has been collected using the twint library needs to be processed first. Twitter data that has the 'place' attribute outside Java or does not have the 'place' attribute is omitted. Then, the duplicate/duplicate tweet was deleted. Furthermore, each tweet is grouped into districts/cities based on the 'place' attribute listed. After that, several preprocessing stages were carried out as described in the previous chapter.

The data that has been cleaned and ready to be processed are 533,474 tweets. Then, each tweet is scored lexically using the LabMT library. The mechanism for assigning numbers to tweets with this library is to assign a number to each word contained in the tweets based on the LabMT dictionary, then the average number of each tweet is calculated. Data from scoring results which are outliers are removed from the data set. Data that does not contain outliers are 455,132 tweets. Then, the data is aggregated into district/city level with mean data concentration measure. Furthermore, the aggregated data is called Twitter scoring in this study. This Twitter scoring will be used as the auxiliary variable (X) in the SAE with Measurement Error model.

The Twitter scoring data generated from a collection of tweets certainly contains errors so it cannot use the EBLUP model to estimate the Happiness Index. The SAE model that can be used with auxiliary variables containing errors is SAE with Measurement Error. The error resulting from Twitter scoring is calculated by the mean variance (with assumption $bias^2 = 0$). After that, the SAE Measurement Error modeling was carried out using the 'saeME' package on R [11].

In Small Area Estimation, it is the auxiliary variable with the actual value for the area . If all components X_i are known, model (2) is used for estimation with $v_i \sim (0, \sigma_v^2)$ is the error of the model and $e_i \sim (0, \psi_i)$ is the error of the design survey for y_i. Since X_i may contain errors, the estimators $\hat{X}_i$ are substituted for X_i as follows. [12]

$$y_i = \hat{X}_i^T \beta + r_i(\hat{X}_i + X_i) + e_i, \tag{3}$$

Where $r_i(\hat{X}_i + X_i) = v_i + (X_i - \hat{X}_i)^T \beta$

4. Comparison of the Estimated Results of the Happiness Index in 2017

After calculating the estimated Happiness Index in 119 districts/cities in Java Island in 2017 with direct and indirect estimation methods, the estimation results will be compared to find out the best results. The value that will be compared from the three methods is the estimated value of the Happiness Index 2017 and the error rate for the district/city level.

4.1 Comparison of the Estimated Values of the Happiness Index in 2017

Based on the graph in Figure 4, it can be seen that the estimated value of the Happiness Index in 2017 in 119 districts/cities in Java Island is not much different between direct estimates, the SAE EBLUP Fay-Herriot method, and SAE with Measurement Error. This is also shown by the Pearson correlation between the three estimation methods with a number of more than 0.95 which is significant with a p-value that is smaller than 5 percent. The results of the Pearson correlation coefficient show that districts / cities that have a high Happiness Index in the direct estimation results also have a high Happiness Index in the Fay-Herriot SAE EBLUP and SAE with Measurement Error results.

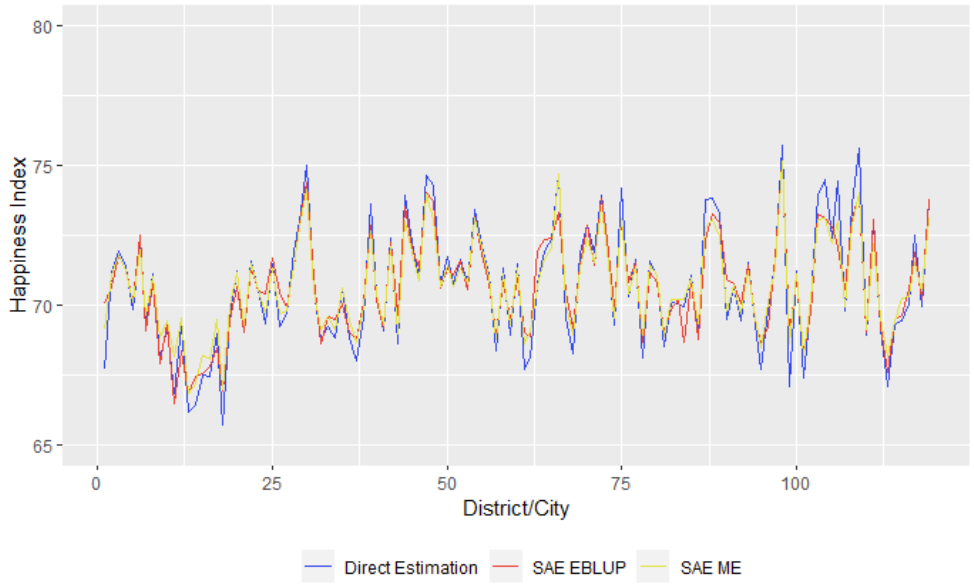

Figure 4. Statistical comparison graph of direct estimation results, SAE EBLUP, and SAE ME
Source: BPS and Twitter data, processed

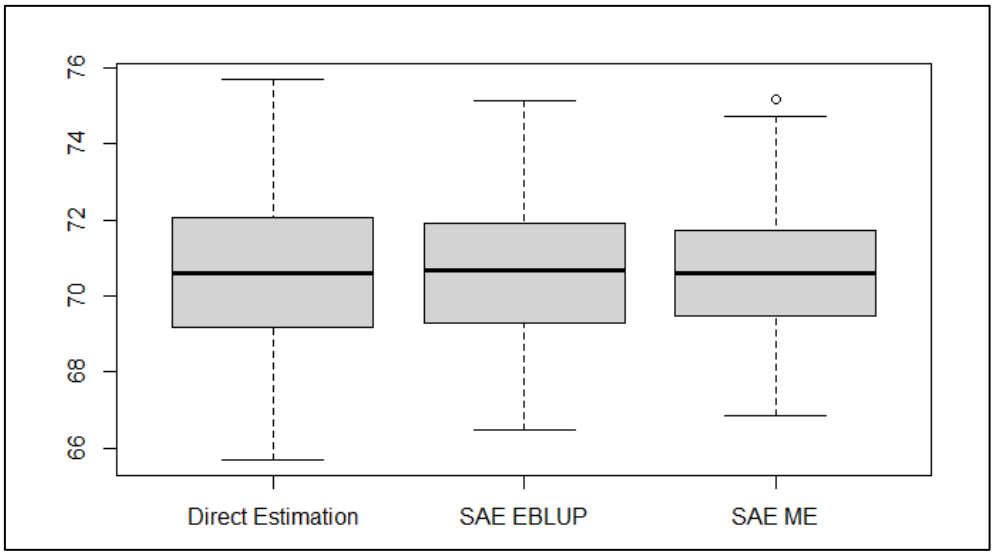

Figure 5. Boxplot of statistical comparison of direct estimation results, SAE EBLUP, and SAE ME
Source: BPS and Twitter data, processed

It can be seen in Figure 5 that the statistics produced by each method have a similar average. The boxplot width shows the variance of each estimation result. The variance resulting from the direct estimation is the largest among the three methods, which is 4.84, while the variance for the results of SAE EBLUP and SAE Measurement Error is 3.26 and 2.81, respectively. This shows that the direct estimation results tend to be more heterogeneous than the estimation using the SAE method. The estimation results from SAE with Measurement Error model are more homogeneous than the direct estimation and indirect estimation using the SAE EBLUP method.

4.2 MSE Comparison of Happiness Index 2017 Estimates

A review of whether or not the estimation results obtained through direct estimation, SAE EBLUP Fay-Herriot, and SAE with Measurement Error is carried out by taking into account the MSE value generated from each model. Figure 6 is a comparison graph of the MSE directly estimated, SAE EBLUP Fay-Herriot, and SAE with Measurement Error.

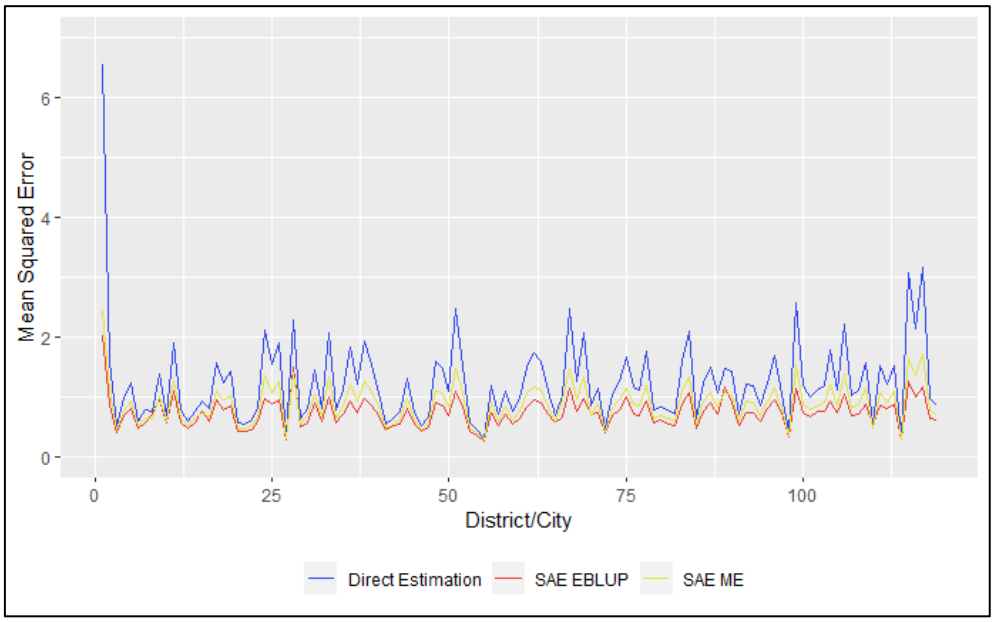

Figure 6. Comparison graph of MSE direct estimation results, SAE EBLUP, and SAE ME
Source: BPS and Twitter data, processed

In the figure, it can be seen that both the SAE EBLUP Fay-Herriot method and the SAE with Measurement Error method have a smaller MSE than the direct estimate. This shows that the estimation results obtained through the SAE method, both the EBLUP Fay-Herriot model and the Measurement Error model, are more reliable than direct estimates. This is in line with the theory put forward by Rao & Molina [3] and Ybarra & Lohr [12]. The MSE estimation results from the SAE EBLUP Fay-Herriot method is the most reliable estimate which is shown by the line graph which is the lowest compared to the other two methods. The average MSE generated by the direct estimation method, SAE EBLUP Fay-Herriot, and SAE with Measurement Error were 1.25, 0.75, and 0.89, respectively.

5. Mapping Happiness Index 2017

Figure 7 shows the mapping of the Happiness Index in Java Island in 2017. The white color represents the average Happiness Index in Java Island. The red color represents the Happiness Index with a number below the average, while the blue color is the opposite. The darker color shows the farther Happiness Index in the district/city is from the average.

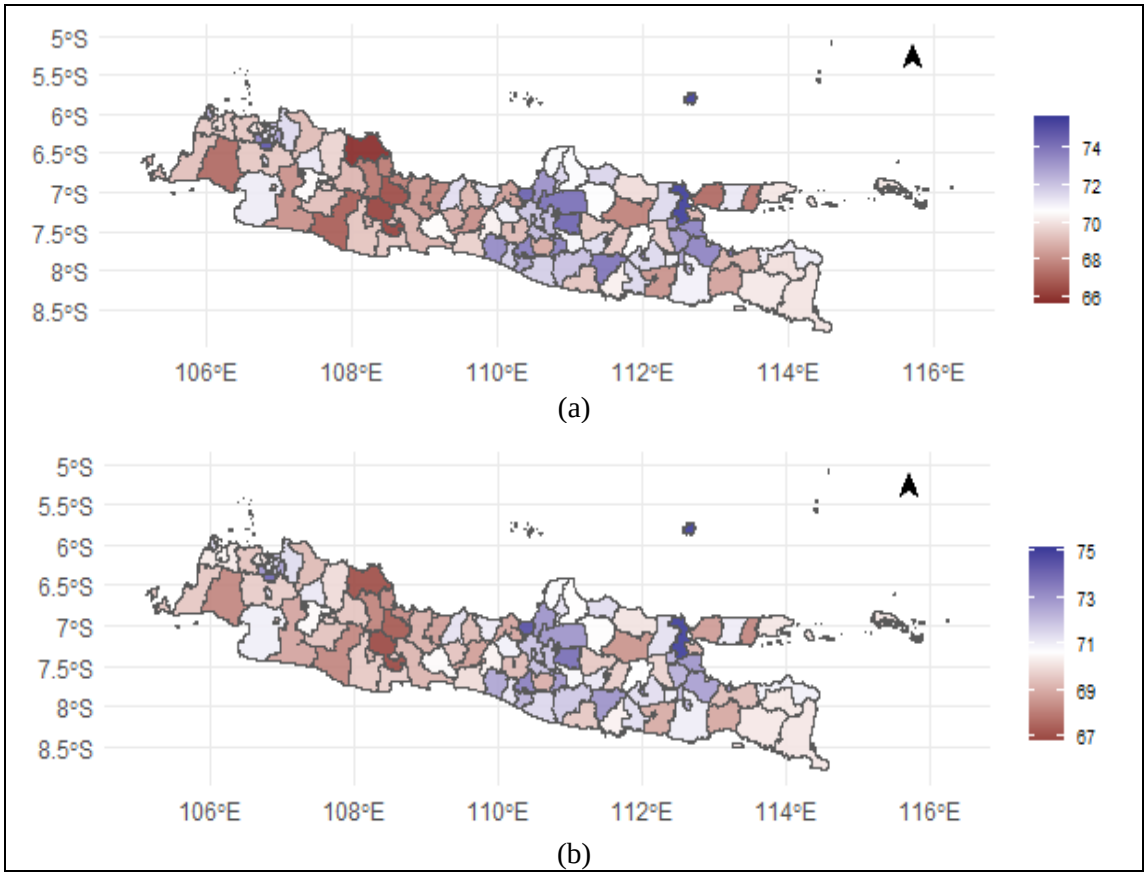

Figure 7. Estimated Happiness Index in Java Island in 2017 using a) SAE EBLUP Fay-Herriot method; b) SAE with Measurement Error
Source: BPS and Twitter data, processed

In the figure 7, the SAE EBLUP Fay-Herriot and SAE with Measurement Error models produce a happiness index at the district/city level above 65. The Happiness Index itself is a scale range from 0-100 with 50 as the middle number. The Happiness Index, which is above 50 and close to 100, illustrates the living conditions of the population who are increasingly happier. That way, it can be said that the overall population on the island of Java has a level of life that tends to be happy. Overall, when viewed from the color scale, the number of districts/cities that are colored red is higher than that of regencies/cities that are colored blue. This means that the number of regencies/cities that have a Happiness Index below the average on Java Island is more than that of regencies/cities that have a Happiness Index above the average. At the district level, the district/city with the lowest Happiness Index 2017 in Java Island using the SAE EBLUP Fay-Herriot method and SAE with Measurement Error, respectively, is occupied by Garut Regency (West Java Province) with a figure of 66.49 and Ciamis Regency (West Java Province) with a figure of 66.85. On the other hand, Gresik Regency is the district/city with the highest Happiness Index using either the SAE EBLUP Fay-Herriot model and the SAE Measurement Error with figures of 75.16 and 70.70, respectively. The district/city with the lowest Happiness Index in 2017 on Java Island using the SAE EBLUP Fay-Herriot method and SAE with Measurement Error, respectively, is occupied by Garut Regency (West Java Province) with a figure of 66.49 and Ciamis Regency (West Java Province) with a figure of 66.85. On the other hand, Gresik Regency is the district/city with the highest Happiness Index using either the SAE EBLUP Fay-Herriot model and the SAE Measurement Error with figures of 75.16 and 70.70, respectively. The district/city with the lowest Happiness Index in 2017 on Java Island using the SAE EBLUP Fay-Herriot method and SAE with Measurement Error, respectively, is occupied by Garut Regency (West Java Province) with a figure of 66.49 and Ciamis Regency (West Java Province) with a figure of 66.85. On the other hand,

Gresik Regency is the district/city with the highest Happiness Index using either the SAE EBLUP Fay-Herriot model and the SAE Measurement Error with figures of 75.16 and 70.70, respectively.

6. Evaluation SAE Model implementation by Utilizing Big Data

Estimation directly at the small area level often has a sample adequacy problem so that the resulting estimator is less precise. To meet the assumption of sample adequacy, it is necessary to add more samples until the number of samples is met. However, of course this will cost a lot of money so that it can be said that the survey is not cost efficient. On the other hand, the precision of the estimator is very important so that the data describes the actual conditions and can be utilized by various parties. Of course, in conducting the survey, it is expected that the resulting data will have a very small error rate because the smaller error, the more reliable the data will be.

Small Area Estimation is an alternative to this problem because this method can increase precision without increasing the number of samples so that it is more cost efficient. In the previous subsection, the comparison between direct and indirect estimates of the Happiness Index 2017 has been explained. Of course, by fulfilling all assumptions, the estimation results obtained by the Small Area Estimation method are more reliable than direct estimates.

There are two Small Area Estimation models used in this study, namely the SAE EBLUP Fay-Herriot model and the SAE model with Measurement Error. Based on the explanation in the previous subsection, the estimation results using the SAE EBLUP Fay-Herriot model are better than the estimation results using the SAE with Measurement Error model. However, the Fay-Herriot SAE EBLUP model requires auxiliary variables that do not contain errors or in other words come from census data. Census data covers all elements of the population. This of course requires a very large cost. Therefore, the Population Census, Agricultural Census, and Economic Census are conducted every 10 years. The Potensi Desa (Podes) data is also carried out thoroughly so that it does not have a sampling error. Podes only done 3 times in 10 years, i.e. every 2 years before the census to support the smooth implementation of the census. Because Podes data is not available every time, sometimes the auxiliary variables sourced from Podes data become less relevant.

On the other hand, there is massive data available in real-time called big data. The availability of big data can certainly be useful if managed properly. In the previous subsection, the Small Area Estimation method has been combined with big data and produces a smaller error rate than direct estimation. Although the estimation results of the Small Area Estimation EBLUP Fay-Herriot model with the auxiliary variable Podes are more precise, big data (in this case Twitter data) can be an alternative when there is no relevant census/survey data to be used as auxiliary variables in the SAE method.

In addition to the availability of real-time data, the costs involved in collecting big data are relatively small. Big data can also be obtained to the smallest level of an area and even individuals. In addition, big data is also very diverse so that many fields can take advantage of this massive data.

In using Twitter data as an auxiliary variable, it is important to note that Twitter data has a self-selection bias. However, in this study there is limited information so it is assumed that this bias can be ignored as has been done in previous studies [4]. In addition, the Twitter scoring variable generated from processed Twitter data can be affected by measurement errors because sometimes there are "happy tweets" that do not match "happy people". Therefore, the SAE model that can be used with the auxiliary variables for Twitter data is the EBLUP Fay-Herriot model that has been developed by Ybarra and Lohr called SAE with Measurement Error [12]. This model takes into account random errors in the covariates when the auxiliary variables are derived from the survey.

Utilization of big data that is considered big for now can be considered normal in the future. What is needed is the development of skills to handle data and assess its use in statistical models, as has been done in the SAE model with Measurement Error in this study.

7. Conclusion

Based on the exposure in the previous chapters, the conclusions obtained from this study are as follows.

1. The districts/cities with the lowest Happiness Index 2017 in Java Island using the SAE EBLUP Fay-Herriot method and SAE with Measurement Error are respectively occupied by Garut Regency (West Java Province) with a figure of 66.49 and Ciamis Regency (West Java Province)

with a figure of 66.85. On the other hand, Gresik Regency is the district/city with the highest Happiness Index using both the SAE EBLUP Fay-Herriot model and the SAE with Measurement Error with figures of 75.16 and 70.70, respectively.

2. The MSE generated from the Small Area Estimation estimate is smaller than the direct estimate, meaning that it is proven that the estimator generated from the SAE method is more reliable.
3. The error rate of the SAE EBLUP Fay-Herriot model with the auxiliary variable Podes is lower than the SAE model with Measurement Error with the auxiliary variable Twitter scoring. Even so, big data can still be used as an alternative to the auxiliary variables for the SAE model because big data covers up to the smallest area and even individuals, is available in real-time, and the costs incurred are relatively low.

8. Suggestions

The suggestions that can be applied to further research are as follows.

1. Scoring tweets semantically.
2. Reviewing the combination of Small Area Estimation with other big data sources, such as Google Trends, Instagram data, Mobile Positioning Data, etc.
3. Create a web-based system or application that can process Twitter data as well as calculate the Happiness Index using the SAE with Measurement Error method.
4. Utilizing Twitter data as an auxiliary variable in the SAE model to estimate other indicators, such as households' share of food consumption expenditure in Indonesia.

References

[1] BPS. (2014). Kajian Indikator Sustainable Development Goals (SDGs). Jakarta: Badan Pusat Statistik.
[2] BPS. (2017). Indeks Kebahagiaan 2017. Jakarta: Badan Pusat Statistik.
[3] Rao, JNK, & Molina I. (2015). Small Area Estimation. New Jersey: John Wiley & Sons, Inc.
[4] Marchetti, S., et al. (2015). Small Area Model-Based Estimators Using Big Data Sources. Journal of Official Statistics, 31(2), 263–281.
[5] Hariyanto, S., et al. (2018). Measurement Error in Small Area Estimation: a Literature Review. IOP Conference Series: Earth and Environmental Science, 187(1).
[6] Github Twint Documentation. (2019). twintproject/twint. Accessed on October 10, 2019 via https://github.com/twintproject/twint
[7] Reagan, Andy. (2018). labMTsimple Documentation.
[8] Asri, AS, & Mariyah, S. (2018). Analisis Indeks Kebahagiaan Subjektif berdasarkan Twitter di Indonesia [Skripsi]. Jakarta: Politeknik Statistika STIS.
[9] BPS. (2018). Potensi Desa (Podes). Accessed on January 15, 2020 via https://www.bps.go.id/linkTableDinamis/view/id/960.
[10] Molina, Isabel & Marhuenda, Yolanda. (2015). sae: An R Package for Small Area Estimation. R Journal. 7.
[11] Mubarak, Muhammad & Ubaidillah, Azka. (2020). Package 'saeME' Title Small Area Estimation with Measurement Error.
[12] Ybarra, LMR, & Lohr, SL (2008). Small Area Estimation when Auxiliary Information is Measured with Error. Biometrics, 95(4), 919–931.