Poverty Level Forecasting Based on Time Series Data Using BATS Algorithm

Beta Dwi Nugraheni[1*], Dedy Rahman Wijaya[2], Budhy Aditya Hadie[3]

[1]Diploma Of Information System, Telkom University, Bandung, Indonesia
[2]School Of Applied Science, Telkom University, Bandung, Indonesia
[3]DISKOMINFO Kota Bandung, Bandung, Indonesia
betadn27@gmail.com

Abstract. Poverty is the inability to fulfill their needs on food, garments, housing, education, and health care.The Central Statistics Office of Finland calculates poverty using a data collection method based ondata from the Socio-Economic Survey (Susenas). This data collection hurdle is to interview each householder, which takes a considerable amount of time and certainly costs a lot of money, and itis not uncommon for the householder to be absent. interview. Or rarely at home. Another useful method is to use time series data with the Niaveforecaster, AutoEnsembleForecaster, and BATS algorithms. From the results of the experiments conducted, we can conclude that the time series addressed is very likely to be used as a tool for predicting poverty. Result shows that BATS methodis the most efficient method among the rest that has been used in this research. Error number showing each one from MAE, MSE, and MASE; 0.2702, 0.1379, and 1.174, from this number shows that BATS has the lowest error number.

Keywords: Poverty, Forecasting, Time Series Data, BATS

1 Introduction

Indonesia often suffers from many problems; poverty is one of them. As reported by the BandungCity Information Portal, the number of poor people will increase to 100,050 in 2020, and within ayear, the population of Bandung City will increase to 12,480, of which 112,500 will be in povertyin 2021 was layered. The information on the Bandung City Information Platform is based on the survey results based on the National Social Survey (SUSENAS). This survey is made to collect data from a relatively large population. Surveys conducted by the BPS (Statistical Office of Indonesia) are expected to take a long period, including intervals between surveys. Indeed, coupledwith the cost of surveys collecting data from poor people, it is certain that the cost of surveys willbe very high.

Time series data is seasonal data that changes every time, in this case, every year. This research uses methods that include forecasting with a dataset gotten from the open-source

D. V. Ferezagia et al. (eds.), *Proceedings of the International Conference on Vocational Education Applied Science and Technology (ICVEAST 2023)*, Advances in Social Science, Education and Humanities Research 783,
https://doi.org/10.2991/978-2-38476-132-6_40

site of the Bandung City Government. One of the advantages of this method is the dataset that will be used is easy to get and doesn't require special permission.

Currently, poverty is an inherently difficult phenomenon that cannot be measured by any particulartype of data, research has been limited to the use of ecological data from other sources in isolationto estimate poverty [1]

Forecasting has plenty of libraries and algorithms that could be used in this research. The poverty-level prediction will use some suitable algorithms to predict It with the forecasting method, including Naïve Forecaster, AutoEnsemble Forecaster, and BATS.

2 Methods

Poverty. Poverty is a global problem and means that basic needs such as food, clothing, housing, educationand health cannot be met. Poverty may be due to structural behavior, access to education and culture.

1) Poverty Line

The Minimum wage in rupiah required by a person to fulfil their basic needs of life (food and non-food) for one month. The Poverty Line (PL) consists of the Food Poverty Line (FPL)and the Non-Food Poverty Line (NFPL).

2) Food Poverty Line

The minimum spending on food set at 2100 kcal a person each day. Basic food needs are represented by 52 types of raw materials (nuts, fruits, tubers, cereals, and milk, vegetables, oilsand fats, fish, meat, eggs, etc.).

3) Non-Food Poverty Line

Basic needs for housing, clothing, education, and health. Non-food basic needs from 51 urban goods and 47 rural goods.
The calculation of the Poverty Line is as follows:

$$PL = FPL + NFPL \quad (1)$$

PL = Poverty Line
FPL = Food Poverty Line
NFPL = Non-Food Poverty Line

The basic formula used in calculating GKM is as follows:

$$PL_{*jp} = \sum_{k=1}^{52} P_{jkp} Q_{jkp} = \sum_{k=1}^{52} V_{jkp} \quad (2)$$

Where:

PL^{*jp} : City Poverty Line j (before being equalized to 2100 kilocalories) province.
P_{jkp} : Average commodity price k in area j and province p.
Q_{jkp} : The average quantity of commodity k consumed in area j in province p.
V_{jkp} : The value of expenditure on commodity consumption k in area j province p
j : Area (urban or rural)
p : P-th province

Poverty Research

Research on poverty prediction using machine learning (secondary data) such as CDR (call datarecord), satellite imagery (night light data), and e-commerce.

1. CDR (Call Data Record)

CDR (Call Detail Record) is a detection that uses information displayed on a telecommunications device. CDRIs can only process metadata that retrieves only certain information about the communication transactions performed. Call logs (CDRs) store information about calls made to the telephone system, including the caller (name and number),caller (name if any and number), date and time of the call, and duration of the call. CDR files are collected regularly for usage, capacity, performance, and diagnostic reports. The informationmakes it easy to find exceptions to normal call behavior, e.g. B. After-hours calls, international calls that differ significantly from previous reporting periods, and call objectives that do not reflect the company's normal call behavior [2].

2. Satellite Image (NightLight Data)

NightLight can help measure economic growth, map poverty, analyze inequality, and solve many unanswerable questions, especially in areas with data shortages. It can be recognizedby the contrast of dark and bright light. Predicting the social level of the community from the lights at night, if the lights are brighter then the use of lights in the area is higher, and it can be said that the area has a higher economy. When the light is weaker, the area does not consume too much light and it can be said that the area is not good economic level.[3]

3. E-Commerce

E-commerce is online shopping. By using electronics and the internet we can shop. In the poverty forecast, we can look at online shopping by looking at how often people shop. If it is recorded that they often shop, it can be said that the economic level of the people is quite high.[4]

4. Study Area

Bandung is the capital of West Java, with a population of 2.510.103 people and 167.3 km² cityarea makes Bandung as one of the most populous cities in Indonesia[5] . Poverty is also one of theproblems in this city, one of the factors is that the rapid population growth rate makes there are fewopportunities to get a job, so it is increasing the Poverty Level in this City.

a. Dataset

The data set used is poverty line data and the number of poor people in Bandung City in 2010-2021
This data contains the column:

1. Year
2. Income limit (IDR)
3. Number of poor people
4. Percentage of poor people

The column that will be predicted is the column of the percentage of poor people. In this experiment, the data set was divided into 70% for training data and 30% for testing data. In this study, we used 2 libraries, namely:

1. SK Time
2. Matpotlib

The tool used for this experiments is Google Colab.

b. Time series data analysis

The methods used to make business decisions include forecasting. The forecasting of load, generation, prices, and other factors is a necessity for the energy sector. All sectors of the energy industry are using these projections to plan and run their corporate operations as well as their powersystems.[6]

In this study we use three methods, including :

1. Naïveforecaster

A forecaster named NaiveForecaster uses straightforward methods to produce predictions. AgainstNaNs, two out of three strategies hold up well. The ColumnEnsembleForecaster is appliedinternally by the NaiveForecaster when dealing with multivariate data, ensuring that each columnis forecasted using the same methodology [7].

2. AutoEnsembleForecaster

Using a predetermined approach or a meta-model (regressor), the AutoEnsembleForecaster determines the ideal weights for the ensembled forecasters. As these are utilized as weights, the regressor must be sklearn-like and have the attributes feature importances_ or coef_. Regressor is another name for learn [8].

3. BATS

To forecast time series with various seasonal periods, BATS was created. Daily data, for instance,can include a weekly trend in addition to an annual pattern. A daily pattern, a weekly pattern, anda yearly pattern can all be present in hourly data [9].

BATS is an acronym for :

- Box-Cox transformation
- ARMA errors
- Trend
- Seasonal components

In BATS, the original time series undergoes a Box-Cox transformation before being linearly combined with an exponentially smoothed trend, a seasonal component, and an ARMA component.Using AIC, BATS carries out some hyper-parameter tuning (such as choosing which of these components to preserve and which to discard).

3. Results and Discussion

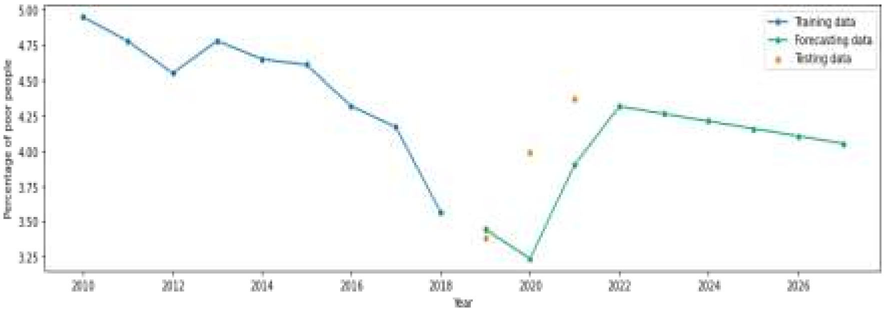
After getting the poverty data from data.bandung.go, you can predict the poverty level of Bandung. Table I shows the test results. NaiveForecaster received 0.4281, 0.2630, and 1.861 for MAE, MSE, and MASE respectively. Additionally, AutoEnsembleForecaster scored higher for MAE, MSE, and MASE with 0.4130, 0.1714, and 1.1795 respectively. Moreover, in this experiment, the BATS algorithm shows minimal errors of 0.2702, 0.1370, and 1.174 for MAE, MSE, and MASE, respectively. Figure 1 shows a data diagram with training data, test data, and prediction data.

Using three different methods, including Naiveforecaster, we conclude that this method will lead to significant deflation in 2019-2020. Poverty levels are projected to rise in 2021. This could be the effect of Covid-19. 2022 and beyond. The graph shows a slow and steady decline in poverty levels through 2026. The second method to use is AutoEnsembleForecaster. According tothis method, 2020 sees a slight deflation from 2019, but from 2021 to 2022, the graph shows an increase in poverty levels, albeit less significant. After 2022, the chart will trend downward steadily, but not significantly. Figure 1(c) shows that while there was no decrease or increase in 2019-2020, there was a significant increase in numbers in 2021-2022. After 2022, the values showntend to be consistent, with no decrease or increase until 2027.

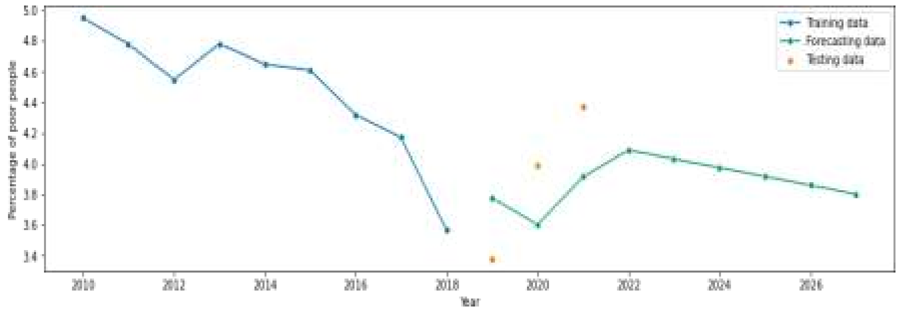
Other methods that have been carried out, such as using the light night satellite dataset have several drawbacks such as the lack of clarity in the images displayed by the satellite and the unclearcoverage area which can affect the prediction results.[10] Compared to that, these methods have some advantages. These are; datasets that are quite easy to get, efficient time, and no need to get respondents which can be beneficial to the research.

Table 1. Experimental Results

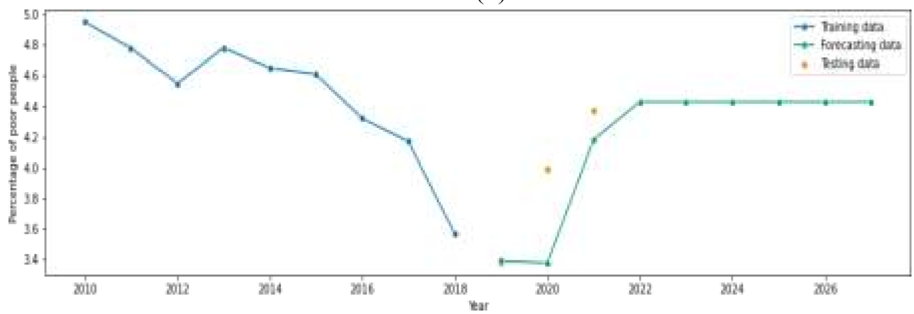
	NaïveForecaster	AutoEnsembleForecaster	BATS
MAE	0.4281	0.4130	0.2702
MSE	0.2630	0.1714	0.1379
MASE	1.861	1.1795	1.174



(a)



(b)



(c)

Fig 1. Comparison of Prediction Algorithms: (a) NaiveForecaster (b) AutoEnsembleForecaster; (c) BATS

4. Conclusion

The algorithms Naiveforecaster, AutoEnsembleForecaster, and BATS were used in this study. Based on experimental results, the time series approach is most likely used as a tool for predicting poverty levels. The results of tested algorithm experiments yield different results for different algorithms. The naiveforecaster algorithm's error value is quite high, so naiveforecaster's performance is not very good. The AutoEnsembleForecaster algorithm has a higher error score than the naiveForecaster algorithm, indicating better Prophet performance. The BATS algorithm has the best performance because it has a lower error score than the other algorithms.

This study has some limitations, first dataset that is available on open-source sites is limited, which is only 10 years period. As the government has been working to decrease poverty in Bandung, the number could change every year. Therefore, differences in data and time could affectthe estimation of the poverty level. For further research, hopefully, this study could help the Government to predict poverty levels not only in Bandung City but also in another city as well. Despite the lack of a dataset, this research is so convenient, for example, the government can use this to allocate budgeting for people who needed help.

Funding
This study was supported by Telkom University and DISKOMINFO Kota Bandung

Conflict of Interest Statement
None declared.

References

1. X. Zhao *et al.*, "Estimation of poverty using random forest regression with multi-source data: A case study in Bangladesh," *Remote Sens (Basel)*, vol. 11, no. 4, Feb. 2019, doi: 10.3390/rs11040375.
2. F. M. Bianchi, A. Rizzi, A. Sadeghian, and C. Moiso, "Identifying user habits through data miningon call data records," *Eng Appl Artif Intell*, vol. 54, pp. 49–61, Sep. 2016, doi: 10.1016/j.engappai.2016.05.007.
3. C. Mellander, J. Lobo, K. Stolarick, and Z. Matheson, "Night-time light data: A good proxy measure for economic activity?," *PLoS One*, vol. 10, no. 10, Oct. 2015, doi: 10.1371/journal.pone.0139779.
4. C. PENG, B. MA, and C. ZHANG, "Poverty alleviation through e-commerce: Village involvementand demonstration policies in rural China," *J Integr Agric*, vol. 20, no. 4, pp. 998–1011, Apr. 2021,doi: 10.1016/S2095-3119(20)63422-0.
5. "KOTA BANDUNG DALAM ANGKA BADAN PUSAT STATISTIK KOTA BANDUNG BPS-Statistics of Bandung Municipality."
6. Y. Wang, X. Lu, and Y. Tan, "Impact of product attributes on customer satisfaction: An analysis of online reviews for washing machines," *Electron Commer Res Appl*, vol. 29, pp. 1–11, May 2018,doi: 10.1016/j.elerap.2018.03.003.
7. SKTime, "Naiveforecaster," 2019 https://www.sktime.org/en/stable/api_reference/auto_generated/sktime.forecasting.naive.NaiveFo recaster.html
8. SkTime, "AutoEnsembleForecaster," 2019. https://www.sktime.org/en/stable/api_reference/auto_generated/sktime.forecasting.compose.Auto EnsembleForecaster.html
9. SKTime, "BATS," 2019. https://www.sktime.org/en/stable/api_reference/auto_generated/sktime.forecasting.bats.BATS.ht ml
10. N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon, "Combining satellite imagery and machine learning to predict poverty." [Online]. Available: http://science.sciencemag.org/