Jurnal Teknik Informatika (JUTIF)
Vol. 4, No. 1, Februari 2023, hlm. 11-13

DOI: https://doi.org/10.20884/1.jutif.2023.4.1.614
p-ISSN: 2723-3863
e-ISSN: 2723-3871

PREDICTION OF STUNTING PREVALENCE IN EAST JAVA PROVINCE WITH RANDOM FOREST ALGORITHM

M. Syauqi Haris[*1], Mochammad Anshori[2], Ahsanun Naseh Khudori[3]

[1,2,3]Informatika, Institut Teknologi, Sains, dan Kesehatan RS dr. Soepraoen Kesdam V/Brw, Indonesia
Email: [1]haris@itsk-soepraoen.ac.id, [2]moanshori@itsk-soepraoen.ac.id, [3]ahsanunnaseh@itsk-soepraoen.ac.id

(Naskah masuk: 19 September 2022, Revisi : 23 September 2022, diterbitkan: 10 Februari 2023)

Abstract

Stunting or cases of failure to thrive in toddlers is one of the most serious health problems faced by the people of Indonesia. Based on data from the Ministry of Health and the Central Statistics Agency, East Java Province has a stunting prevalence value of 26.8% which is categorized as a high prevalence value according to the standards of the World Health Organization (WHO). Random forest is one of the machine learning algorithms in the field of artificial intelligence that can learn patterns from labeled data so that it can be used as a method for predicting or forecasting data. This approach is considered very suitable to be used in predicting the value of stunting prevalence because stunting prevalence data is usually accompanied by other data in the health sector according to survey results. Previous studies on the prediction of stunting prevalence used secondary data sourced from one survey only. Therefore, this study is one of the efforts to contribute in providing solutions for the stunting problem in East Java Province by combining several data from different surveys in the same year. The results of this study show that from 20 factor candidates for predicting stunting prevalence value, only 12 factors are suspected to be causative factors based on their correlation value. However, the prediction results obtained using the random forest algorithm in this study, with data consisting of 12 features and a dataset consisting of only 38 data, have results with error values of 1.02 in MAE and 1.64 in MSE that are not better than multi-linear regression which can produce smaller error values of 0.93 in MAE and 1.34 in MSE.

Keywords: *Machine Learning, Prediction, Python, Random Forest, Stunting.*

PREDIKSI NILAI PREVALENSI STUNTING DI PROVINSI JAWA TIMUR MENGGUNAKAN ALGORITMA RANDOM FOREST

Abstrak

Stunting atau kasus kegagalan tumbuh pada balita adalah salah satu masalah kesehatan yang cukup serius dihadapi oleh masyarakat Indonesia. Berdasarkan data dari Kementerian Kesehatan dan Badan Pusat Statistik, Provinsi Jawa Timur memiliki nilai prevalensi stunting sebesar 26,8% yang dikategorikan sebagai nilai prevalensi tinggi sesuai standar dari World Health Organization (WHO). *Random forest* merupakan salah satu algoritma *machine learning* dalam bidang kecerdasan artifisial yang bisa mempelajari pola dari data-data berlabel sehingga bisa digunakan sebagai metode untuk melakukan prediksi atau peramalan data. Pendekatan ini dirasa sangat sesuai digunakan dalam prediksi nilai prevalensi stunting pada suatu daerah karena data prevalensi stunting biasanya diiringi dengan data-data lainnya di bidang kesehatan sesuai hasil survei. Penelitian-penelitian tentang prediksi prevalensi stunting sebelumnya rata-rata hanya menggunakan data sekunder yang bersumber dari salah satu sumber survei saja. Oleh karena itu, penelitian ini merupakan salah satu upaya untuk memberikan kontribusi dalam memberikan solusi untuk permasalahan stunting di Provinsi Jawa Timur dengan menggabungkan beberapa data hasil survei yang berbeda-beda pada tahun yang sama. Hasil penelitian ini menunjukkan bahwa dari 20 kandidat faktor untuk memprediksi nilai prevalensi stunting, hanya 12 faktor yang diduga sebagai faktor penyebab berdasarkan nilai korelasinya. Namun hasil prediksi yang diperoleh dengan menggunakan algoritma *random forest* pada penelitian ini, dengan data yang terdiri dari 12 fitur dan dataset yang hanya terdiri dari 38 data, memiliki hasil dengan nilai error 1,02 pada MAE dan 1,64 pada MSE yang tidak lebih baik dari regresi multi linier yang mampu menghasilkan nilai error yang lebih kecil yaitu 0,93 pada MAE dan 1,34 pada MSE.

Kata kunci: *Machine Learning, Prediksi, Python, Random Forest, Stunting.*

1. INTRODUCTION

Stunting is one of the most important health and welfare crises in the world today and is a concern for global public health. With a frequency of 27.7%, Indonesia has a relatively high stunting prevalence. In other words, in 2019, stunting affected 28 out of every 100 children under five years old. Although the fact that the prevalence has reduced over the years, the stunting prevalence remains high in comparison to other middle-income countries [1]. The stunting prevalence can be categorized as very high, high, medium, low, and very low. A very high category is for stunting prevalence beyond 30%, high for prevalence value between 20 and 30%, while a medium is between 10 and 20%, between 2.5 and 10% is categorized as low, and less than 2.5% is categorized as very low.

Children's stunting is linked to social outcomes and both short as well as long-term health. Mortality and morbidity are two of the most common short-term consequences of stunting [2]. Childhood stunting has long-term consequences such as poor maternal health outcomes, low academic success, developmental motor delays, and poor cognitive development [3].

The identification of characteristics that are independently linked with stunting in children under five has been done throughout time using traditional statistical models [4]. These techniques, however, are often not reliable when there is multi-correlation among the variables or when the number of covariates exceeds the number of observations. Furthermore, the current challenge is predicting the prevalence value of stunting so that stunting problems can be avoided in the future. Machine learning techniques, namely random forest, can be used for prediction.

2. RELATED STUDY

A similar study regarding machine learning in Zambia in 2022 on stunting has been done by Chilyabanyama et al. using data from the ZDHS 2018, they determined that in Zambia, the random forest model provided the greatest predicted accuracy for stunting among children under the age of five [5]. Titaley et al. in 2019 also conducted a study regarding stunting utilizing data from the 2013 Indonesia Basic Health Survey towards children under two years old in Indonesia. They indicated that there are several requirements to decrease stunting prevalence in Indonesia such as multi-sectoral coordinated treatments during the prenatal and postnatal periods [6].

Another research regarding stunting in children was also conducted by Muche et al. in 2021, they used the data from the Ethiopian Demographic and Health Survey (EDHS) conducted from January 18 to June 27, 2016, and implied that children stunting in Ethiopia were caused by community and individual level factors after being analyzed using Bayesian multilevel logistic regression [7]. Bitew et al. in 2022 also studied regarding stunting in Ethiopia using data from the Ethiopian Demographic and Health Survey of 2016. They compared two algorithms to predict the stunting prevalence [8]. In addition, Rahman et al. in 2021 also conducted a study regarding identifying and predicting the main risk factors for underweight children, wasting, and stunting in Bangladesh utilizing a machine learning algorithm [9]. Furthermore, this study focuses on random forest algorithms in machine learning to predict stunting prevalence in East Java, Indonesia. It is because when creating predictions, random forest is highly reliable as it has been widely used in the field of medicine. This study aims at predicting the prevalence of stunting in East Java, Indonesia. A descriptive quantitative method using a random forest algorithm was used in this study.

3. LITERATURE REVIEW

3.1. Stunting

Stunting term is often known as being short or stunted, is a condition where children under the age of five (toddlers) fail to grow due to severe malnutrition or repeated illnesses that happen within the first 1,000 days of life (HPK). Combining short as well as extremely short child nutrition difficulties is referred to as stunting. When compared to the World Health Organization (WHO) standard, short toddlers are those whose nutritional status is based on height or length according to age. The z-score value must be less than minus two (-2) standard deviations to be considered short, and it must be less than minus three (-3) standard deviations to be considered very short [10].

Impaired growth or stunting has adverse functional consequences on the child, especially on their development. The consequences include poor cognition and academic performance, lost productivity, and low adult income. Furthermore, when the disease is worsened by excessive weight gain later in childhood, the risk of nutrition-related chronic disorders in adulthood increases [2]. Children's stunting is frequently caused by poor hygiene and inadequate food security in the poorest households [11].

Today's toddlers still face several dietary issues, including the prevalence of stunting. Around 150.8 million children under five years old, or 22.2%, were stunted in 2017. Nevertheless, this number is lower than the 32.6% stunting rate in 2000. More than half (55%) of stunted children under five in 2017 were from Asia, while more than a third (39%) were from Africa. The majority (58.7%) of Asia's 83.6 million young children under five were from South Asia. In the Southeast Asia/South-East Asia Regional (SEAR) area, Indonesia ranks third among the countries with the greatest prevalence of stunting among children under the age of five, according to WHO data [10].

In Indonesia, the prevalence of short toddlers is generally stable. According to the 2007 Indonesian

Basic Health Survey, 36.8% of children in Indonesia had stunting. The percentage dropped slightly to 35.6% in 2010. However, the proportion of stunted toddlers rose once more to 37.2% in 2013 [12]. Although it is considered to be generally stable, the prevalence of children who had stunting in Indonesia is still high.

3.2. Machine Learning

Machine learning (ML) is an effective method for identifying unknown patterns or relationships by combining statistical learning with artificial intelligence. Without direct human rule determination, machine learning may create intelligent computer systems. Rather, the system is designed to detect patterns in training data sets to develop a model that can be used to estimate a value (regression) or collection of values (classification) [13]. Machine learning is possible if data is provided as input for the analysis of massive data sets to find specific patterns.

There are two types of data in machine learning, namely training and test data. Training data is used to train algorithms, and test data is used to assess the performance of trained machine learning, such as when finding new data that has never been provided in training data [14]. In general, machine learning algorithms create a model based on sample data, often known as training data, to make predictions or decisions without being explicitly taught.

Medical research has used a variety of ML methods. For example, machine learning (ML) methods like support vector machines, artificial neural networks, and Random Forest (RF), have been utilized to forecast the prognosis of illnesses including diabetes and acute appendicitis [8]. In general, to generate predictions or choices without being explicitly taught, machine learning algorithms develop a model based on sample data, often known as training data.

Several researchers have used machine learning algorithms to make predictions and categorize issues related to stunting and toddler nutrition, including the use of the K-Nearest Neighbor (KNN) algorithm or KNN with backward elimination feature, the Nave Bayes algorithm, and the Decision Tree. While just for one sub-district, the Random Forest algorithm has also been used to categorize children who are stunted [15][16][17]. When compared to traditional statistical models, modern ML algorithms have demonstrated improved predictive performance when dealing with categorization difficulties (See Figure 1).

Random Forest (RF) is a supervised learning technique that can be used to solve classification and regression issues. This learning approach combines numerous decision trees into a single model. Random Forest has the advantage of being suitable for vast amounts of data, overcoming noise and missing information. The disadvantages include difficult interpretation and accurate adjustment of the model [18]. Random Forest is a versatile and user-friendly algorithm. In certain circumstances, this method produces good results. Random Forests can be produced in the following steps, determine the number of trees to be formed, then the random samples are taken as many as N from the dataset. After that, a random subset of predictors is taken in each tree and repeated in the second and third process until as many trees [19].

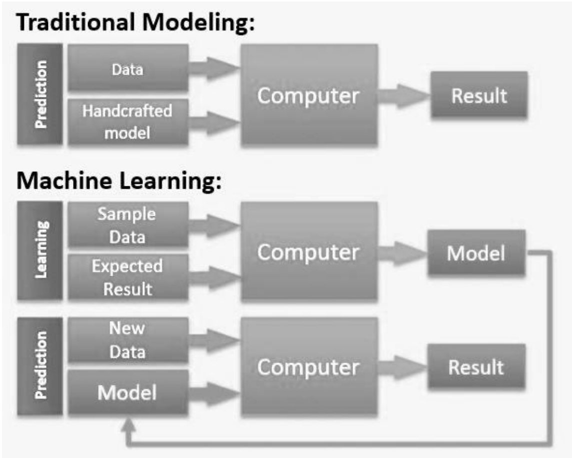

Figure 1. Traditional Modeling and Machine Learning Differences

The RF algorithm's classification accuracy is determined by the accuracy of each base classifier's classification and between the base classifiers' similarity [20]. It is positively related to the classification accuracy of each basis classifier and negatively related to the base classifiers' similarity. Random Forest uses an error-minimizing technique to select the variables to split into groups [21]. Random forest employs the method in such a manner that the correlation within subtrees is higher while the correlation between subtrees is lower. This algorithm is used frequently to generate a regression classification tree by determining the prediction for numerous data sets utilizing predictor variables random collection [8]. Following the formation of several of these trees, the prediction performance of each variable is tested, and the variable optimal collection is achieved. It is extremely adaptable as well as rapid, and it may be used for both classification and regression.

4. METHODOLOGY

The method used in this study is descriptive quantitative. There are several steps in conducting this study, namely literature review, collecting data, pre-processing data, feature selection, data splitting, data modeling, and model evaluation (See Figure 2).

Related studies are used to find the novelty of this research.

Furthermore, after determining the focus of this research, we collected data through five sources as secondary data from surveys in 2019, namely *Studi Status Gizi Balita Indonesia (SSGBI), Survei Sosial*

Ekonomi Nasional (SUSENAS), Laporan Direktorat Statistik Kesejahteraan Rakyat Badan Pusat Statistik, Statistik Perumahan dan Pemukiman, and Indeks Pembangunan Manusia.

From the results of the five surveys mentioned before, we obtained data for 38 regencies/cities in the province of East Java. From this data, we get 20 predictions of the factors causing stunting. These factors are shown in Table 1. There are 20 factors that are used to predict stunting in 38 districts and cities in the province of East Java.

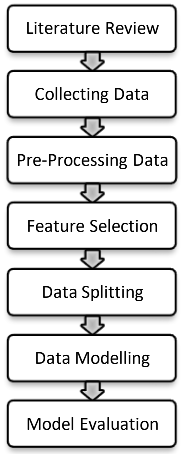

Figure 2. Research Process

Table 1. Stunting Factors in East Java, Indonesia

No	Factors
1	% Pregnant Women K4 Antenatal Care 2019 (*Ibu Hamil K4 2019*)
2	% Pregnant Women Iron Supplement (90 Tablets) (*Ibu Hamil Mendapat Tablet Tambah Darah (90 Tablet) 2019*)
3	% Parturition helped by Healthcare Workers 2019 (*Persalinan Ditolong Nakes 2019*)
4	% Women in post-partum took A (*Ibu Nifas Mendapat Vit A*)
5	% Low Birth Weight 2019 (*BBLR 2019*)
6	% New Born Baby who got Early Initiation of Breastfeeding 2019 *(Bayi Baru Lahir Mendapat IMD 2019*)
7	% Children Under Two Given Breast Milk (*Baduta diberi ASI 2019*)
8	Breastfeeding Average Period (month) (*Rata-rata Lama Pemberian ASI (bulan) 2019*)
9	Breastfeeding Average Period with Weaning Food 2019 (*Rata-rata Lama Pemberian ASI dengan Makanan Pendamping 2019*)
10	Toddler who gets Full Immunization (*% Balita Mendapat Imunisasi Lengkap 2019*)
11	6 to 11 Months Baby who get Vitamin A 2019 (*% Bayi 6-11 Bulan Mendapat Vit A 2019*)
12	12 to 59 Months Baby who get Vitamin A 2019 (*% Bayi 12-59 Bulan Mendapat Vit A 2019)*
13	6-50 Months Baby who get Vitamin A 2019 (*% Bayi 6-59 Bulan Mendapat Vit A 2019)*
14	Integrated Healthcare Scale for Toddler 2019 (*% Cakupan Pelayanan Kesehatan Balita 2019)*
15	Average Household 2019 (*Rata-rata Jiwa/Rumah Tangga 2019)*
16	Neighborhood Association that has proper sanitation 2019 (*% RT Memiliki Akses Sanitasi Layak 2019*)
17	Neighborhood Association that has Good Quality of Drinking Water Sources (*% RT Memiliki Akses Sumber Air Minum Layak 2019*)
18	Human Development Index (*IPM 2019*)
19	Per Capita Real Expenditure (*Pengeluaran Riil per Kapita 2019*)
20	Poor Community (*Penduduk Miskin 2019*)

5. RESULT AND DISCUSSION

In this research, we used the machine learning life-cycle for implementation under the *Jupyter Notebook* environment, which consists of seven stages, from data gathering, data preparation, data wrangling, data analysis, model training, until model testing, which is shown in Figure 3. While for the deployment stage we currently defer the process since we still focus on evaluating machine learning models for the accuracy of predicting the stunting prevalence value.

Data gathering is done by acquiring data from the survey results that are mostly published in PDF format. Online PDF to XLS (Excel Spreadsheet) converter is used in this step to produce table formatted data in CSV (Comma Separated Value) for further processing in the *Jupyter Notebook* environment. These semi-manual data preparation processes are done because processing PDF directly in *Jupyter Notebook* will be harder, since the files have very complex content, and separating the specified data needed from other information will need a more sophisticated method.

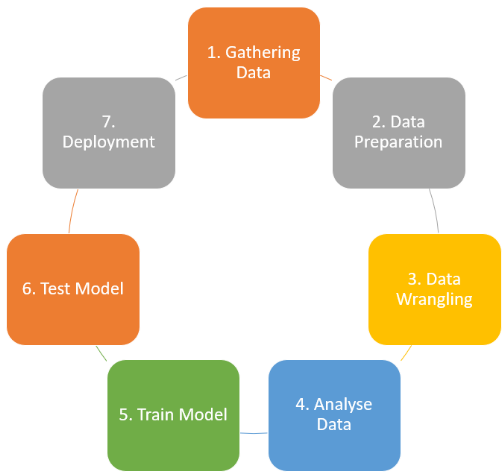

Figure 3. Machine Learning Life-cycle

Jurnal Teknik Informatika (JUTIF)
Vol. 4, No. 1, Februari 2023, hlm. 11-13

DOI: https://doi.org/10.20884/1.jutif.2023.4.1.614
p-ISSN: 2723-3863
e-ISSN: 2723-3871

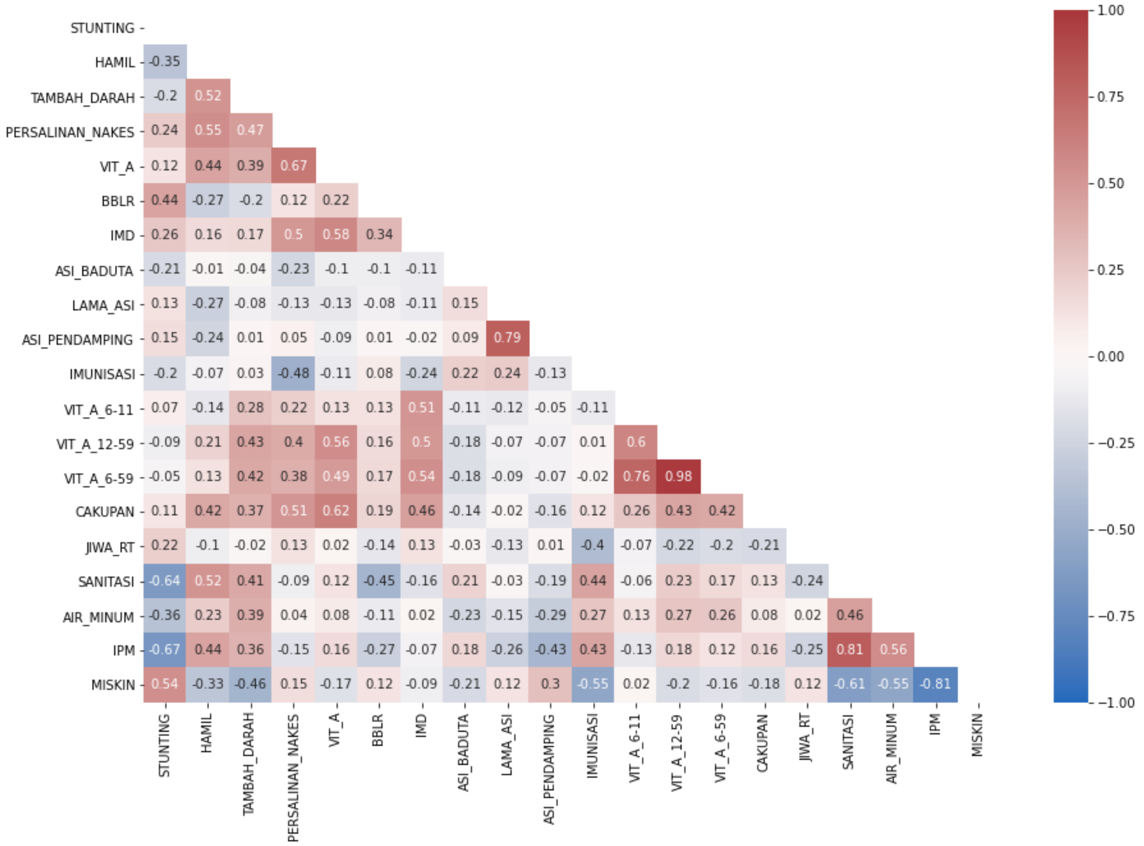

Figure 4. Feature Correlation Value

Data wrangling means cleaning data from missing data, noisy data, and inconsistent data. In this research, we benefited from the secondary data that are used. So, the data are already having no missing, noisy, and inconsistent data, since it must be well processed before. However, combining several survey results will need data scaling for normalizing or standardizing data to improve the performance of predictive modeling algorithms. We used the *StandardScaler* function from the python *sklearn* library for this data scaling process.

The data analyzing phase is conducted by comparing correlation values between 20 factors acquired to stunting prevalence value. The comparison is shown in Figure 4 which is generated with the *pyplot* library in the *jupyter notebook*. Red color means a strong positive correlation value while blue color means a strong negative correlation value. Since the factors will be used in predicting a value in machine learning, so we can call the factors are feature in machine learning.

In this phase, feature reduction can be done by analyzing the feature correlation values, we can differentiate which factors have medium correlation (between 0.2 and 0.4 for positive and between -0.4 and -0.2 for negative), strong correlation (greater than 0.4 or lesser than -0.4) and the factors which have weak correlation value (between -0.2 and 0.2). Factors with strong correlation will be maintained and the others will be eliminated. Factors that are maintained consist of 12 factors shown in Table 2.

Table 2. Selected Features Based on Correlation

No	Factors	Correlation
1	% Pregnant Women K4 Antenatal Care	- 0.35
2	% Pregnant Women Iron Supplement (90 Tablets)	- 0.2
3	% Parturition helped by Healthcare Workers	0.24
4	% Low Birth Weight	0.44
5	% New Born Baby who got Early Initiation of Breastfeeding 2019	0.26
6	% Children Under Two Given Breast Milk	- 0.21
7	Toddlers who get Full Immunization	- 0.2
8	Average Household	0.22
9	Neighborhood Association that has proper sanitation	- 0.64
10	Neighborhood Association that has Good Quality Drinking Water Sources	- 0.36
11	Human Development Index	- 0.67
12	Poor Community	0.54

RF model training and testing were using data from selected features by splitting the data into training data and test data with 80% to 20% in comparison. The model learning process is done by using the *RandomForestRegressor* function from the python *sklearn* library and subsequently, the testing process was done by comparing the model-predicted value to the real value from test data. The value deviation is represented by Mean Absolute Error

(MAE) and Mean Squared Error (MSE) as seen in Figure 5.

```
from sklearn.ensemble import RandomForestRegressor
Regressor3 = RandomForestRegressor()
Regressor3.fit(X_train,np.ravel(y_train))

RandomForestRegressor()

y_pred3 = Regressor3.predict(X_test)
print("Mean absolute error: %.2f" % np.mean(np.absolute(y_pred3 - y_test)))
print("Residual sum of squares (MSE): %.2f" % np.mean((y_pred3 - y_test) ** 2))

Mean absolute error: 1.02
Residual sum of squares (MSE): 1.64
```

Figure 5. RF Model Training and Testing Screenshot

To enrich the findings of this research, we also use Multi Linear Regression (MLR), one of the common statistical techniques for prediction, as a model comparison. MLR can be done by using the LinearRegression function from the sklearn python library, while MAE and MSE deviation values were also computed as a comparison as seen in Figure 6.

```
from sklearn import linear_model
regressor1 = linear_model.LinearRegression()
regressor1.fit (X_train, y_train)
print ('Coefficients: ', regressor1.coef_)
print ('Intercept: ',regressor1.intercept_)

Coefficients:  [[-0.27317458  0.11163811  0.44307547  0.17770062  0.0702026  -0.08148186
   0.46443851  0.12162765 -0.70659279  0.02192399  0.16996097  0.39807988]]
Intercept:  [0.12712418]

y_pred1 = regressor1.predict(X_test)
print("Mean absolute error: %.2f" % np.mean(np.absolute(y_pred1 - y_test)))
print("Residual sum of squares (MSE): %.2f" % np.mean((y_pred1 - y_test) ** 2))

Mean absolute error: 0.93
Residual sum of squares (MSE): 1.34
```

Figure 6. MLR Training and Testing Screenshot

Based on two regression models in this research, we can compare the prediction accuracy that is represented by the MAE and MSE value of error as seen in Table 3.

Table 3. Prediction Error Value Comparison

Regression Model	MAE value	MSE value
Random Forest Regressor	1.02	1.64
Multi Linear Regressor	0.93	1.34

This research uses Multi Linear Regression (MLR) prediction as a comparison to Random Forest (RF) prediction. As we know that a lower error value means better prediction, the result of this research shows that MLR with MAE value 0.93 and MSE value 1.34 outperform random forest regression with MAE value 1.02 and MSE value 1.64.

6. CONCLUSION

Machine learning modeling for predicting stunting prevalence value can be done by using several official survey result data that are publicly available. Combining several survey results in the same years will produce many features, or causative factor candidates in this research, that are needed to be analyzed and selected based on their correlation value to the prediction target value. This research shows that from 20 factor candidates for predicting stunting prevalence value, only 12 factors are suspected to be causative factors based on their correlation value.

The result of this research shows that Multi Linear Regression (MLR) prediction is still better than Random Forest (RF) prediction with MAE values 0.93 to 1.02 and MSE values 1.34 to 1.64. These results could be caused by the number of features, in this case, we used 12 features, and the number of datasets, in this case, we only have 38 data from each city and district in East Java. Most machine learning models will have better performance when it is trained using a large number of datasets. While the statistical method is better when the dataset is small and has minimum noisy data. East Java province which only has 38 districts and cities means the dataset will only consist of 38 data, so gathering more data from other provinces will probably increase the accuracy of random forest regression model for the next research.

7. ACKNOWLEDGEMENT

We would like to thank the Ministry of Education, Culture, Research, and Technology for funding this research as a part of novice lecturer research grants. We would also like to thank for all personnel involved for the accomplishment of this research.

REFERENCES

[1] Subdirektorat Statistik dan Perumahan, “Profil Statistik Kesehatan,” Jakarta, 2019.

[2] M. De Onis *et al.*, “Prevalence thresholds for wasting, overweight and stunting in children under 5 years,” *Public Health Nutr.*, vol. 22, no. 1, pp. 175–179, 2019, doi: 10.1017/S1368980018002434.

[3] K. G. Dewey and K. Begum, “Long-term consequences of stunting in early life,” *Matern. Child Nutr.*, vol. 7, no. SUPPL. 3, pp. 5–18, 2011, doi: 10.1111/j.1740-8709.2011.00349.x.

[4] B. Mzumara, P. Bwembya, H. Halwiindi, R. Mugode, and J. Banda, “Factors associated with stunting among children below five years of age in Zambia: Evidence from the 2014 Zambia demographic and health survey,” *BMC Nutr.*, vol. 4, no. 1, pp. 1–8, 2018, doi: 10.1186/s40795-018-0260-9.

[5] O. N. Chilyabanyama *et al.*, “Performance of Machine Learning Classifiers in Classifying Stunting among Under-Five Children in Zambia,” 2022.

[6] C. R. Titaley, I. Ariawan, D. Hapsari, and A. Muasyaroh, “Determinants of the Stunting of Children in Indonesia : A Multilevel Analysis of the 2013 Indonesia Basic Health Survey,” *Nutrients*, vol. 11, p. 1160, 2013.

[7] A. Muche, L. D. Gezie, A. G. egzabher Baraki, and E. T. Amsalu, “Predictors of stunting among children age 6–59 months in

Ethiopia using Bayesian multi-level analysis," *Sci. Rep.*, vol. 11, no. 1, pp. 1–12, 2021, doi: 10.1038/s41598-021-82755-7.

[8] F. H. Bitew, C. S. Sparks, and S. H. Nyarko, "Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia," *Public Health Nutr.*, vol. 25, no. 2, pp. 269–280, 2022, doi: 10.1017/S1368980021004262.

[9] S. M. J. Rahman *et al.*, "Investigate the risk factors of stunting, wasting, and underweight among under-five Bangladeshi children and its prediction based on machine learning approach," *PLoS One*, vol. 16, no. 6 June 2021, pp. 1–11, 2021, doi: 10.1371/journal.pone.0253172.

[10] Kementerian Kesehatan Republik Indonesia, "Buletin Jendela Data dan Informasi Kesehatan: Situasi Balita Pendek (Stunting) di Indonesia," Jakarta, 2018.

[11] S. van der Berg, L. Patel, and G. Bridgman, "Food insecurity in South Africa: Evidence from NIDS-CRAM wave 5," *Dev. South. Afr.*, vol. 0, no. 0, pp. 1–16, 2022, doi: 10.1080/0376835X.2022.2062299.

[12] The Ministry of Health of The Republic of Indonesia, "Report on the Implementation of the March 2019 Susenas Integration and the 2019 Indonesian Toddler Nutritional Status Survey," 2019.

[13] E. Retnoningsih and R. Pramudita, "Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python," *BINA Insa. ICT J.*, vol. 7, no. 2, p. 156, Dec. 2020, doi: 10.51211/biict.v7i2.1422.

[14] P. D. Kusuma, *Machine Learning Teori, Program, Dan Studi Kasus*. Yogyakarta: Deepublish, 2020.

[15] S. Lonang and D. Normawati, "Klasifikasi Status Stunting Pada Balita Menggunakan K-Nearest Neighbor Dengan Feature Selection Backward Elimination," *J. Media Inform. Budidarma*, vol. 6, no. 1, p. 49, 2022, doi: 10.30865/mib.v6i1.3312.

[16] A. Byna, "Comparative Analysis of Machine Learning Algorithms for classification about Stunting Genesis," 2020, doi: 10.4108/eai.23-11-2019.2298349.

[17] A. Y. Perdana, R. Latuconsina, and A. Dinimaharawati, "Prediksi Stunting Pada Balita Dengan Algoritma Random Forest," *ISSN 2355-9365 e-Proceeding Eng. Vol.8, No.5 Oktober 2021*, vol. 8, no. 5, pp. 6650–6656, 2021.

[18] M. Anshori, F. Mar'i, and F. A. Bachtiar, "Comparison of Machine Learning Methods for Android Malicious Software Classification based on System Call," *Proc. 2019 4th Int. Conf. Sustain. Inf. Eng. Technol. SIET 2019*, pp. 343–348, 2019, doi: 10.1109/SIET48054.2019.8985998.

[19] N. Khasanah, R. Komarudin, N. Afni, Y. I. Maulana, and A. Salim, "Skin Cancer Classification Using Random Forest Algorithm," *Sisfotenika*, vol. 11, no. 2, p. 137, 2021, doi: 10.30700/jst.v11i2.1122.

[20] L. Wan, K. Gong, G. Zhang, X. Yuan, C. Li, and X. Deng, "An efficient rolling bearing fault diagnosis method based on spark and improved random forest algorithm," *IEEE Access*, vol. 9, pp. 37866–37882, 2021, doi: 10.1109/ACCESS.2021.3063929.

[21] S. A. Hemo and M. I. Rayhan, "Classification tree and random forest model to predict under-five malnutrition in Bangladesh," *Biom Biostat Int J*, vol. 10, no. 3, pp. 116–123, 2021, doi: 10.15406/bbij.2021.10.00337.