

Optimalizing Big Data in Reducing Miss-Targeting Family Hope Program (PKH) in Sidoarjo Disctrict with Approach Machine Learning

Aditama Azmy Musaddad¹, Arimurti Kriswibowo^{*2}

^{1,2} Public Administration, Faculty of Social and Political Sciences, *Universitas Pembangunan Nasional “Veteran” Jawa Timur*, Surabaya, Indonesia

e-mail: ¹aditamaazmy@gmail.com, ^{*2}arimurti.adne@upnjatim.ac.id

Abstrak

Pendekatan machine learning telah digunakan untuk memecahkan berbagai masalah. PKH seringkali mengalami miss-targeting. Penelitian ini bertujuan untuk membandingkan hasil Big Data PKH oleh SIKS-NG dan machine learning berdasarkan data dan indikator pengukuran kemiskinan yang sama. Diperoleh model algoritma Averaged Neural Network dengan hasil yang optimal dibandingkan algoritma-algoritma lain. Adapun hasil pengujian data yang diperoleh pada SIKS-NG dan Machine Learning yang menggunakan evaluasi confusion matrix dengan 3 indikator sebagai berikut: 1) Accuracy yang didapatkan SIKS-NG 72,40% meningkat menjadi 81,18% pada Machine Learning; 2) Precision pada SIKS-NG mendapatkan angka persentase tinggi 91,01%, akan tetapi hasil tersebut mampu meningkat setelah data diberi Machine Learning menjadi 95,37%; 3) Recall dengan SIKS-NG memperoleh hasil 75,49%, sedangkan Machine Learning memperoleh hasil yang lebih tinggi yakni 82,19%. Dengan demikian, machine learning terbukti dapat menurunkan angka miss-targeting dan dapat dijadikan rekomendasi alternatif dalam pengambilan keputusan otomatis dan praktik manajemen inovatif di lingkungan pemerintahan.

Kata kunci— *Program Keluarga Harapan, Miss-Targeting, Big Data, Machine Learning*

Abstract

Machine learning approaches have been used to solve various problems. PKH experienced miss-targeting. This study aims to compare the result of big data by SIKS-NG and machine learning based on the same data and measurement indicators. Obtained algorithms Averaged Neural Network with optimal output compared to others. As for data testing obtained on SIKS-NG and machine learning that uses elevated matrix evaluations with the following 3 indicators: 1) Accuracy obtained by SIKS-NG 72.40% increased to 81.18% for Machine Learning; 2) Precision at the center is getting a high percentage of 91,01%, but it is capable of increasing once the data is given Machine Learning to 95,37%; 3) Recall with the cycle was obtained at 75.49%, while Machine Learning obtained a higher yield of 82.19%. Thus, machine learning has been proven to reduce miss-targeting and can be used as an alternative recommendation in automatic decision making and innovative management practices in government circles.

Keywords— *Family Hope Program, Miss-Targeting, Big Data, Machine Learning*

1. INTRODUCTION

The Family of Hope Program (PKH) in the Regulation of the Minister of Social Affairs of the Republic of Indonesia Number 1 of 2018 is a well-planned, targeted and having sustainable goals Social Protection Program. Poverty level data is one of the considerations for determining PKH areas. An important aspect to support a poverty reduction strategy is accurate poverty data [1]. Therefore, the government needs to encourage regular data sharing and data transparency as a requirement for prospective PKH beneficiaries. Azizah, Mahmudah and Kriswibowo (2020) argue that the government's political will is very necessary to minimize the increase in poverty in the village[2]. However, the practice of PKH often experiences inaccurate data collection so that PKH does not reach the poor who really need it. In addition, even though the community has been registered in the Integrated Social Welfare Data (DTKS), there are still poor people who have not received PKH assistance. On the other hand, there are still rich people got PKH assistance. The inaccurate data has resulted in social jealousy among the community and the data has not been integrated systematically.

The miss-targeting problem is the main challenge of the PKH program. The slow process of handling complaints about invalid data at the regional level is confirmed by statements by the SMERU Research Institute in the *katadata.co.id* media [3], Anwar Sadad, Deputy Chairperson of the Regional People's Representative Council of East Java Province 2019-2024 at *kominfo.jatimprov.go.id* [4] and M Dhamroni Chudlori, Deputy Chairman of the Sidoarjo DPRD Covid-19 Handling Committee at *republikjatim.com* media[5]. In addition, the Ombudsman of the Republic of Indonesia (ORI) on the *Kompas.com* [6] has received 817 reports of complaints from the public regarding data manipulation in the distribution of social assistance in order to tackle the Covid-19 pandemic. There are 2 types of errors in analyzing miss-targeting, namely under coverage and leakage [7].

The situation above indicates that poverty reduction strategies must be effective, efficient and transparent by the application of Big Data technology with a machine learning approach. Big Data Analytics assists in finding valuable decisions by understanding data patterns with the help of machine learning algorithms [8]. There are several opportunities to use Big Data in the public sector, including getting feedback and public response from government service information systems and from social media, as a basis for policy making and improving public services [9]. The term machine learning is used to show a systematic and unsystematic graph of the growth and availability of large amounts of data. Machine learning is a sub-field of artificial intelligence that is widely researched and used to solve various problems [10].

Research conducted by Fitriani aims to determine the eligibility of PKH beneficiaries by comparing the C4.5 and Naïve Bayes algorithms through the Rapidminer tools[11]. The total data were 1,109 residents. The results show that the C4.5 algorithm has an accuracy value of 91.25% and an AUC of 0.930 is the highest among the other methods, while the Naïve Bayes method has an accuracy of 87.11% and AUC of 0.923. In other studies, according to Sugianto and Maulana in their research using the classification method of the Naïve Bayeshasil algorithm, the accuracy is 58.29%, precision 92.90%, recall 21.84%, AUC 0.765 and F-measure 34.42%. and the Decision Tree Algorithm got an accuracy of 73.97%, precision 85.04%, recall of 61.92%, AUC 0.746, and F-Measure 71.17% [12].

Based on the exposure of previous research and ensuring the validity of the data used in machine learning, this study focuses to determine the performance of machine learning algorithms with deep learning characters and provide an overview of the comparison of the miss-targeting level of PKH recipients with data processed by the Indonesian Ministry of Social Affairs using the System Next-Generation Social Welfare Information (SIKS-NG).

The data and measurement indicators used are the same, which distinguishes the data processing tools. The tools used in this research are RStudio which is an integrated development environment (IDE) specifically for the R programming language and statistical analysis which is supported by many packages and functions as a translator. Machine learning mechanisms for large-scale multidimensional data from multiple sources are indispensable. It aims to facilitate a

more accurate determination of the poor [13]. Machine learning as a field of artificial intelligence in Indonesia is still filled by many actors in the business sector. Almost in line with big data, this happens because there are still many players in the business sector well-established data processing infrastructure [14].

There are strong reasons to believe that intelligent data analysis with machine learning will become more widespread as a necessary ingredient for technological advances, especially in the formulation of public policies. Therefore, this study aims to find a model that can reduce the error rate of PKH aid distribution in Sidoarjo Regency, so that it can contribute to national development.

2. METHODS

2.1 Problem

The problem that will be examined in this research is using big data as a source, namely PKH recipients in Sidoarjo Regency. However, it is often miss-targeting which has caused conflict empirically. Big Data analytics helps in finding valuable decisions by understanding data patterns with the help of machine learning algorithms [8]. The machine learning approach has become one of the mainstays of information technology, supported by the large amount of data available.

2.2 Data Collection

The data that will be the main processed material are only in Tanggulangin District with a number of poor people of 5,688 people, and Candi District with a number of poor people of 7,214 people. It is because of the poverty data of other sub-districts are not ready to be released considering the nature of the data is very sensitive, very confidential and already never anonymized (de-identified), therefore the confidentiality of personal data needs to be guaranteed.

There are 14 variables to determine poor households. The 14 variables used in this study are depicted in Table 1.

Table 1 Criteria for Determining the Poor

Number	Variable
1.	Residential Building Floor Area Less than 8 m ² per person
2.	Types of residential floors made of cheap soil / bamboo / wood
3.	Types of residential walls made of bamboo / thatch / low quality wood / walls without plaster
4.	Do not have defecation facilities / together with other households
5.	Household lighting sources do not use electricity
6.	The source of drinking water comes from wells / unprotected springs / rivers / rainwater
7.	Daily cooking fuel is firewood / charcoal / kerosene
8.	Consuming Meat / Milk / Chicken once a week
9.	Purchasing a set of clothes only once year
10.	Having meals once or twice a day
11.	Unable to Pay Medical Costs at the community Health centers / Polyclinic
12.	Sources of income for the head of a household are: Farmers with a land area of 500 M2, farm workers, fishermen, construction workers, plantation workers and / or other occupations with an income below Rp. 600,000, - per month
13.	Highest Education Head of Household: No School / Not Completed Elementary School / Completed Elementary School
14.	Do not have savings / items that are easy to sell with a minimum of Rp. 500,000, - such as credit / non-credit motorbikes, gold, livestock, motor boats, or other capital goods

Note: If at least 9 variables are met, it is called a poor household.

Source: [15]

2.3 Identification of Required Data

Based on data obtained from the Social Service Office of Sidoarjo Regency, it includes 14 these variables, however, when verifying and validating data on a predetermined form, these variables are branched into 70 sub variables.

Table 2 Sub variables on verification and data validation forms

Number	Sub Variable	Number	Sub Variable
1.	IDBDT	36.	There's a gas cylinder
2.	RUTA6	37.	There is a fridge
3.	KDPROP	38.	There is a AC
4.	KDKAB	39.	There is a heater
5.	KDKEC	40.	There is a phone
6.	KDDESA	41.	There is a tv
7.	Address	42.	There is a gold
8.	Participants' number PKH	43.	There is a laptop
9.	Participants' number KKS 2016	44.	There is a bicycle
10.	Participants' number PBI	45.	There is a motor
11.	Participants' KIP	46.	There is a car
12.	Name SLS	47.	There is a boat
13.	Name KRT	48.	There is a motor temple
14.	Total ART	49.	There is a motor boat
15.	Total families	50.	There is a ship
16.	Sta building	51.	Immovable assets
17.	Sta land	52.	ATB area
18.	Floor area	53.	Another house
19.	Floor	54.	Total cow
20.	Wall	55.	Total buffalo
21.	Wall Condition	56.	Total horse
22.	Roof	57.	Total pig
23.	Roof condition	58.	Total goats
24.	Total rooms	59.	Sta art business
25.	Source of drinking water	60.	Sta KKS
26.	Water meter number	61.	Sta KIP
27.	How to get drinking water	62.	Sta KIS
28.	Source of light	63.	Sta BPJS Independently
29.	Energy	64.	Sta JAMSOSTEK
30.	PLN number	65.	Sta Insurance
31.	Cooking fuel	66.	Sta PKH
32.	Gas number	67.	Sta rastra
33.	Fasbab	68.	Sta kur
34.	Bidet	69.	Sta existence RT
35.	Remove stool	70.	percentile

As for the 70 sub variables that have been presented, both the poverty data of Tanggulangin District and Candi District, only 55 sub variables will be used in machine learning. It is because of the data is sensitive and does not affect.

Furthermore, the classification model is obtained as follows: Class decile 1,2,3,4, and 4+. Decile 1 includes the PKH recipient class which is the focus of this research study, decile 2 includes the Non-Cash Food Assistance recipient class (BPNT), decile 3 and decile 4 including the Healthy Indonesian Card (KIS) class, the last class is decile 4+ for backup data if you have not met the quota for social assistance. Deciles 2, 3, 4, and 4+ are only used as a comparison considering that the data used are poverty data which is the determining measure for all social protection programs.

Table 3 Class Data

	Decile 1	Decile 2	Decile 3	Decile 4	Decile 4+
Percentile	1-10	11-20	21-30	31-40	41-50

2.4 Data Pre-processing

After going through a series of data preparation then it comes to the data pre-processing stage. Before the data is ready to be trained and tested, pre-processing is needed so that the classifier works better [16]. At this stage, a statistical analysis is needed, namely the PCA (Principal Component Analysis) method technique.

This stage is used to visualize multivariate data. The PCA working pattern detects target errors by seeing how many points overlap in the data set and the goal is to avoid the errors in the target data obtained. In addition, PCA functions to compress variable branches without removing the original characters from the parent variable.

2.5 Algorithm Selection

At this stage the machine learning algorithm will be selected. There are several types of algorithms in machine learning such as Nearest Neighbor, Naive Bayes, KNN Classification, Support Vector Machine, Ada Boost, Random Forest, Decision Tree, Neural Network, Bayesian Networks, K-Means Clustering and others [17]; [18]. This is held to obtain the best comparison results from the existing algorithms. It takes a caret library call available on RStudio which functions to determine the automatic turning of algorithm parameters with the highest accuracy value.

2.6 Training

Serves as a guide in the algorithm selection stage so that it can study the patterns given to the selected algorithm model. In other words, it provides directions using an algorithm so that the trained machine can look for its own correlation or learn patterns from the given data.

2.7 Evaluation with test set

The evaluation carried out aims to determine the classification model. The evaluation in this study uses confusion matrix. The confusion matrix used are accuracy, precision, and recall. According to [19] confusion matrix is information about actual classification results that can be predicted by a classification system. Confusion matrix accuracy and precision for measuring the performance of the model generated from the ANN algorithm, while confusion matrix recall measures the level of miss-targeting on PKH.

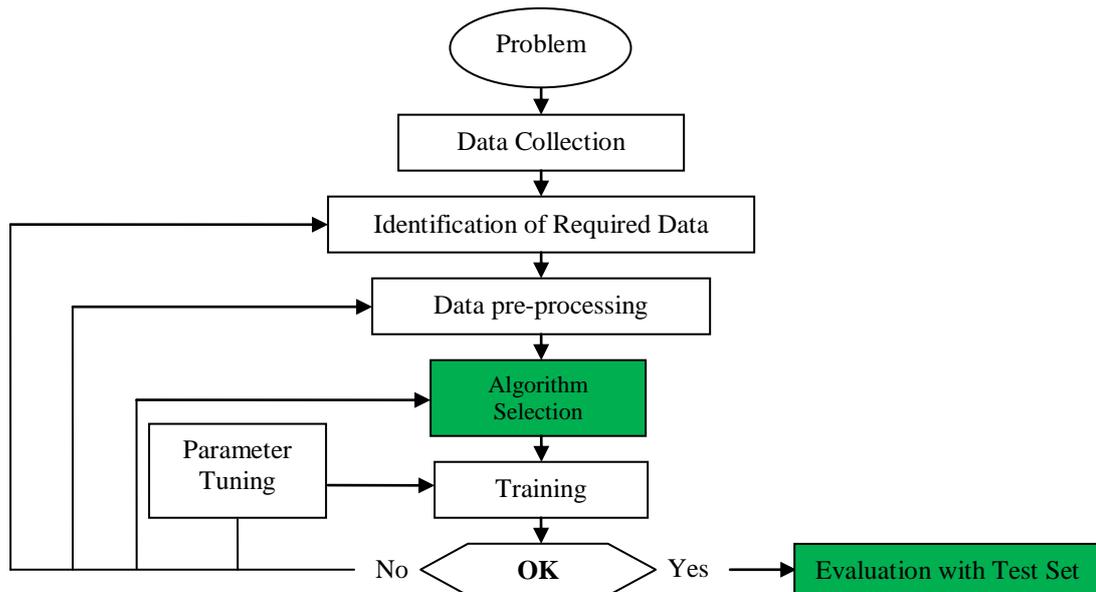


Figure 1 Machine Learning Process

3. RESULTS AND DISCUSSION

3.1. Implementation Process at RStudio

The stages of implementing the program code on RStudio have several outlines, namely calling the library package in RStudio, importing data sets, and using PCA; the built machine learning process including determining parameters: size, decay and bag, training, testing the dataset then evaluating the data. The detailed implementation of the program code in RStudio is as follows:

Program

```

library(tidyverse)
library(caret)
library(plotly)

data <- read.csv("~/Smt 7/Skripsi/datafull.csv")

set.seed(1202)
bpa = prcomp(data[,seq(52)], scale = T, center = T)

bpa$x[,seq(11)] -> metadata

metadata = data.frame(metadata, data[,c(53:56)])

metadata %>%
  rename(
    "penerima.dinsos" = Penerima
  ) -> metadata

metadata %>%
  mutate(penerima.ygbenar =
    ifelse(Desil == "desil1", "ya","tidak")) -> metadata
  
```

```
rep(c(1:5),2454) -> as
as[-1] -> as

set.seed(1202)
metadata[sample(nrow(metadata),)] -> metadata
metadata$no.urut = as

ctrl = trainControl(method = "cv", number = 10)
newdata = NULL

for(i in seq(5)) {

train = metadata %>%
filter(no.urut != i)

test = metadata %>%
filter(no.urut == i)

set.seed(1202)
fit = train(paste("penerima.ygbenar ~", paste(names(train)[seq(11)],
collapse = "+")) %>% as.formula(), data = train,
method = 'avNNet',
trControl = ctrl)

set.seed(1202)
prediksi = predict(fit, test[,seq(11)])

test = data.frame(test, prediksi)
newdata = rbind(newdata, test)

train = NULL
test = NULL
}
newdata$penerima.ygbenar = as.factor(newdata$penerima.ygbenar)

newdata %>%
count(Desil, penerima.dinsos,
name = "Jumlah") %>%
filter(penerima.dinsos == "ya" &
Desil %in% "desil1") %>%
select(Jumlah) %>% sum()-> ya.dinsos

newdata %>%
count(Desil, penerima.dinsos,
name = "Jumlah") %>%
filter(penerima.dinsos == "ya") %>%
select(Jumlah) %>% sum() -> total.dinsos

newdata %>%
count(Desil, prediksi,
name = "Jumlah") %>%
filter(prediksi=="ya" &
```

```

Desil %in% "desil1") %>%
select(Jumlah) %>% sum() -> ya.modbaru

newdata %>%
count(Desil, prediksi,
name = "Jumlah") %>%
filter(prediksi=="ya") %>%
select(Jumlah) %>% sum() -> total.modbaru

newdata %>%
count(Desil) %>%
select(Desil) -> desil

newdata %>%
count(Desil, penerima.dinsos,
name = "Jumlah") %>%
filter(penerima.dinsos == "ya") %>%
select(Jumlah) %>%
rename("Metode.Dinsos" = Jumlah) -> dinsos

newdata %>%
count(Desil, prediksi,
name = "Jumlah") %>%
filter(prediksi == "ya") %>%
select(Jumlah) %>%
rename("Metode.Baru" = Jumlah) -> baru

dataset = data.frame(desil, dinsos, baru)
dataset %>%
mutate(Metode.Dinsos = -Metode.Dinsos) -> dataset

ggplot(dataset, aes(x = Desil)) +
geom_bar(aes(y = Metode.Dinsos),fill = "red", stat = "identity") +
geom_bar(aes(y = Metode.Baru), fill = "green", stat = "identity") +
coord_flip() + scale_x_discrete(limits =
c("desil4+","desil4","desil3","desil2","desil1")) + ylim(-2000,2000) +
theme_minimal() +
labs(x = "", y = "", title = "SIKS-NG vs Machine Learning") +
theme(plot.title =element_text(h = .49)) -> grafik

plotly::ggplotly(grafik)

print(
print(paste(
"Metode SIKS-NG Tepat Sasaran hingga",
c(((ya.dinsos / total.dinsos) * 100) %>% round(2), "%")),
print(paste(
"namun memiliki rasio kesalahan hingga",
c(((total.dinsos - ya.dinsos) / total.dinsos) * 100) %>% round(2), "%")),
print(' '),
print(paste(
"Metode Machine Learning Tepat Sasaran hingga",
c(((ya.modbaru / total.modbaru) * 100) %>% round(2), "%")),

```

```

print(paste(
"namun memiliki rasio kesalahan hingga",
c(((total.modbaru - ya.modbaru) / total.modbaru) * 100) %>% round(2, "%")
)
)

table(newdata$Desil, newdata$penerima.dinsos)
table(newdata$Desil, newdata$prediksi)
confusionMatrix(newdata$penerima.ygbenar, newdata$prediksi)

```

3.2. Results Evaluation

The final result of selecting the algorithm is focused on the Averaged Neural Network algorithm model because of the dozens of existing algorithms, Averaged Neural Network has the highest accuracy among others such as Naive Bayes Classifier, Support Vector Machines (SVM), KNN Classification, Algorithm C5.0, Random Forest Algorithm, and many others. The results of turning parameters of this model are size = 5, decay = 0.1, and bag = FALSE.

Turning parameters on the Averaged Neural Network Algorithm above results in a comparison of the number of poor people who receive PKH and those who do not receive PKH in Tanggulangin District and Candi District between the versions of SIKS-NG and Machine Learning which are represented in the attributes 'ya' and 'tidak'.

```

Model Averaged Neural Network
9815 samples
 11 predictor
 2 classes: 'tidak', 'ya'

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 8834, 8832, 8834, 8834, 8833, 8834, ...
Resampling results across tuning parameters:

```

size	decay	Accuracy	Kappa
1	0e+00	0.7467134	0.01846268
1	1e-04	0.7722878	0.19308290
1	1e-01	0.7948028	0.37343019
3	0e+00	0.7678108	0.13628766
3	1e-04	0.7951100	0.32051836
3	1e-01	0.8042804	0.39595740
5	0e+00	0.7617849	0.10096401
5	1e-04	0.7979687	0.32698745
5	1e-01	0.8121231	0.42662269

```

Tuning parameter 'bag' was held constant at a value of FALSE
Accuracy was used to select the optimal model using the largest value.
The final values used for the model were size = 5, decay = 0.1 and bag = FALSE.

```

Figure 2 Model Averaged Neural Network

As explained by the author in chapter III, the focus of this research is on decile 1 category, namely PKH aid recipients. Based on 5,688 poor people in Tanggulangin Subdistrict and 7,214 poor people in Candi Subdistrict, which the total population of the two sub-districts is 12,902 people. If you look at the picture in 4.8. The total population is 12,269 inhabitants. This is because when the data were identified, in the percentile sub-instrument, there were several residents who were detected as 'NULL' or the population did not have a percentile value (not zero but empty) which indicated that the population was not included in the category of beneficiaries. As for the population who do not have a percentile value of 633 people, so that the data processing given the action is as many as 12,269 people.

In the SIKS-NG data processing in the decile 1 category, the poor recommended to receive PKH assistance were 1,946 people, while those who were not recommended for PKH assistance were 1,195 people. In contrast to the results of Machine Learning data processing, the recommended recipients of PKH assistance were 1,255 people, while the recommended recipients of PKH assistance were 1886 people. The comparison of the results is shown in Figure 2 and Figure 3.

```
> table(newdata$Desil, newdata$penerima.dinsos)
      tidak ya
desil1 1195 1946
desil2 1997  821
desil3 2442  451
desil4 1708  168
desil4+ 1443  98
> table(newdata$Desil, newdata$prediksi)
      tidak ya
desil1 1886 1255
desil2 2554  264
desil3 2778  115
desil4 1843  33
desil4+ 1530  11
```

Figure 3 Results of the Averaged Neural Network Algorithm Implementation

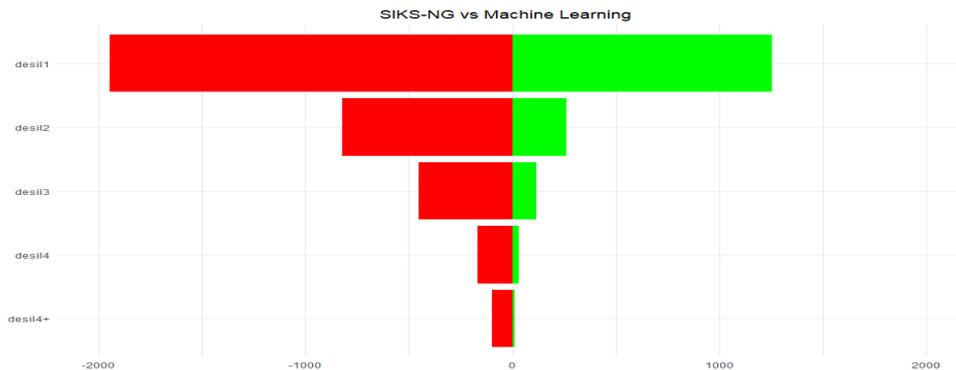


Figure 4 Comparison of PKH Recipient Versions SIKS-NG and Machine Learning

In addition, when there are high-dimensional data with a large number of variables, machine learning is able to show its ability to recognize informative patterns in the data even though the data has quite complex variables. This is evidenced by the higher accuracy and lower level of miss-targeting than the data processed using the SIKS-NG application.

The results of the evaluation use 3 main indicators, namely accuracy, precision, and recall. The results of the comparison of the three can be seen in Figure 4 and Figure 5. Each of the SIKS-NG and Machine Learning tests. The accuracy results obtained have increased in each indicator. The first indicator, the accuracy obtained by SIKS-NG 72.40% increases to 81.18% in Machine Learning. The second indicator, precision in SIKS-NG gets a high percentage of 91.01%, but this result increases in the Machine Learning to be 95.37%. The third indicator, recall with SIKS-NG obtained 75.49% results, while Machine Learning obtained higher results, namely 82.19%. The three indicator results are shown in Figure 7.

```
Confusion Matrix and Statistics
Reference
Prediction tidak ya
tidak 7995 790
ya 2596 888

Accuracy : 0.724
95% CI : (0.716, 0.7319)
No Information Rate : 0.8632
P-Value [Acc > NIR] : 1

Kappa : 0.1955
Mcnemar's Test P-Value : <2e-16

Precision : 0.9101
Recall : 0.7549
F1 : 0.8252
Prevalence : 0.8632
Detection Rate : 0.6516
Detection Prevalence : 0.7160
Balanced Accuracy : 0.6420

'Positive' Class : tidak
```

Figure 5 Confusion Matrix SIKS-NG

```
Confusion Matrix and Statistics
Reference
Prediction tidak ya
tidak 8705 423
ya 1886 1255

Accuracy : 0.8118
95% CI : (0.8048, 0.8187)
No Information Rate : 0.8632
P-Value [Acc > NIR] : 1

Kappa : 0.4169
Mcnemar's Test P-Value : <2e-16

Precision : 0.9537
Recall : 0.8219
F1 : 0.8829
Prevalence : 0.8632
Detection Rate : 0.7095
Detection Prevalence : 0.7440
Balanced Accuracy : 0.7849

'Positive' Class : tidak
```

Figure 6 Confusion Matrix Machine Learning

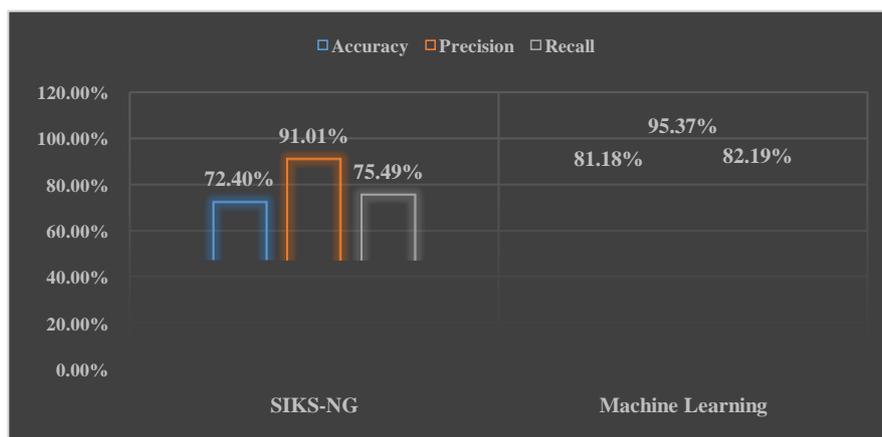


Figure 7 Graph accuracy, precision, and recall values at SIKS-NG and Machine Learning

4. CONCLUSIONS

Based on confusion matrix evaluation with the Averaged Neural Network algorithm model assigned, the machine learning approach is superior in every indicator compared to SIKS-NG. The first indicator, Accuracy produced by Machine Learning is 81.18%, while SIKS-NG is 72.40%, better 8.78%. Second, the Precision produced by Machine Learning is 95.37%, while SIKS-NG is 91.01%, better 4.36%. The third indicator, Recall with 82.19% results for Machine Learning, 75% for SIKS-NG, better 7.19%. In addition, PKH recipients also produced less from 1,946 people to 1,255. Thus, in public governance, Machine Learning is present to provide an innovative alternative and government management efficiency, and is able to select PKH aid recipients more accurately.

Regardless of the confidentiality of SIKS-NG using the same or different approaches, Machine Learning with the Averaged Neural Network algorithm model which has a high level of accuracy can be an alternative recommendation for automatic decision making and innovative management practices.

REFERENCES

- [1] D. V. Ferezagia, "Analisis Tingkat Kemiskinan di Indonesia Jurnal Sosial Humaniora Terapan," *J. Sos. Hum. Terap.*, vol. 1, no. 1, pp. 1–6, 2018.
- [2] W. I. Azizah, Z. Mahmudah, and A. Kriswibowo, "Political Will Pemerintah Kabupaten Jombang Terhadap Penanggulangan Kemiskinan Di Masyarakat Desa," *J. Sos. Ekon. dan Polit.*, vol. 1, no. 1, 2020.
- [3] Katadata.co.id, "Hasil Studi: PKH dan Bantuan Sembako Tak Tepat Sasaran dan Ganjal."
- [4] Kominfo.jatimprov.go.id., "DPRD Siap Terima Aduan Masyarakat di Jawa Timur Terkait Bansos Covid-19 Yang Tidak Tepat Sasaran."
- [5] republikjatim.com, "Panja Covid-19 Dewan Desak Pemkab Sidoarjo Tandai Rumah Penerima Bantuan PKH Dengan Tulisan Cat."
- [6] Kompas.com, "Distribusi Bansos PKH Tak Tepat Sasaran, Mensos Siapkan Aturan Baru."
- [7] A. Fiszbein *et al.*, "Conditional Cash Transfers: Reducing Present and Future Poverty, A World Bank Policy Research Report," The International Bank for Reconstruction and

- Development / The World Bank, Washington, 2009.
- [8] J. Archenaa and E. A. M. Anita, "A Survey of Big Data Analytics in Healthcare and Government," *Procedia Comput. Sci.*, vol. 50, pp. 408–413, 2015.
- [9] E. R. E. Sirait, "Implementasi Teknologi Big Data di Lembaga Pemerintahan Indonesia," *J. Penelit. Pos dan Inform.*, vol. 6, no. 2, pp. 113–136, 2016.
- [10] A. Roihan, P. A. Sunarya, and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review Paper," *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 5, no. 1, pp. 75–82, 2020.
- [11] E. Fitriani, "Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan," *Sist. J. Sist. Inf.*, vol. 9, no. 1, pp. 103–115, 2020.
- [12] C. A. Sugianto and F. R. Maulana, "Algoritma Naïve Bayes Untuk Klasifikasi Penerima Bantuan Pangan Non Tunai (Studi Kasus Kelurahan Utama)," *Techno.Com*, vol. 18, no. 4, pp. 321–331, 2019.
- [13] Z. Chang, L. Lei, Z. Zhou, S. Mao, and T. Ristaniemi, "Learn to Cache: Machine Learning for Network Edge Caching in the Big Data Era," *IEEE Wirel. Commun.*, vol. 25, no. 3, pp. 28–35, 2018.
- [14] H. A. Ramadhan and D. A. Putri, "Big Data, Kecerdasan Buatan, Blockchain, dan Teknologi Finansial di Indonesia (Usulan Desain, Prinsip, dan Rekomendasi Kebijakan)," Jakarta, 2018.
- [15] *Keputusan Menteri Sosial Republik Indonesia Nomor 146/ HUK/ 2013 Tentang Penetapan Kriteria dan Pendataan Fakir Miskin dan Orang Tidak Mampu.* .
- [16] N. Nofriani, "Comparations of Supervised Machine Learning Techniques in Predicting the Classification of the Household's Welfare Status," *J. Pekommas*, vol. 4, no. 1, pp. 43–52, 2019.
- [17] B. Buchanan and T. Miller, *Machine Learning for Policy Makers What It Is and Why It Matters*, no. June. Cambridge: President and Fellows of Harvard College, 2017.
- [18] T. O. Ayodele, *Types of Machine Learning Algorithms, New Advances in Machine Learning*, Yagang Zha. University of Portsmouth United Kingdom, 2010.
- [19] L. Farokhah, "Implementasi K-Nearest Neighbor Untuk Klasifikasi Bunga Dengan Ekstraksi Fitur Warna RGB," *J. Teknol. Inf. Dan Komun.*, vol. 7, no. 6, pp. 1129–1136, 2020.