Propensity score stratification using bootstrap aggregating classification trees analysis

Bambang Widjanarko Otok*, Marsuddin Musa, Purhadi, Septia Devi Prihastuti Yasmirullah

Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia

ARTICLE INFO

Keywords:
Mathematics
Statistics
Bootstrap aggregating
Classification trees analysis
Opportunistic infection
Propensity score stratification

ABSTRACT

Introduction: Observational research in the field of health often does not conduct randomized controlled trials on research subjects. A non-random selection process on research subjects can result in a biased treatment effect due to an imbalance between the treatment and control groups. The problem of bias effects can be dealt with by reducing the bias in the confounding variable using the propensity score method. Estimation of propensity score can use machine learning method with a classification tree analysis approach. The resulting single classification tree model is still unstable if there is a slight change in learning data. Therefore, the ensemble method is applied which is bootstrap aggregating the classification tree as a tool to improve the stability and predictive power of the classification tree.

Methods: The problem of bias effects can be dealt with by reducing the bias in the confounding variable using the propensity score method. Estimation of propensity score can use machine learning method with a classification tree analysis approach. The resulting single classification tree model is still unstable if there is a slight change in learning data. Therefore, the ensemble method is applied which is bootstrap aggregating the classification tree as a tool to improve the stability and predictive power of the classification tree.

Results: This study aims to determine the effect of giving treatment antiretroviral therapy and counseling to opportunistic infections in HIV AIDS patients. The result of propensity score stratification analysis using bootstrap aggregating classification trees analysis is able to reduce the bias by 89.54%, using 5 strata and having a balanced covariate in each stratum.

Conclusion: Testing the effect of treatment shows that there is a significant effect of giving antiretroviral therapy and counseling to opportunistic infections in HIV AIDS patients.

1. Introduction

Often in observational research does not conduct randomize the research subjects. Non-random selection process can result in an imbalance between treatment and control groups so that it can produce a biased treatment effect [1]. The problem of bias effects can be handled by reducing the bias in the confounding variable using the propensity score method so that an appropriate treatment effect can be obtained based on the Average Treatment Effect. The purpose of the propensity score method is eliminating the covariate imbalance between treatment and control groups. Covariates are balanced if the treatment and control groups have the same distribution of observed covariates [2].

Propensity score can be estimated using the machine learning method with a decision trees approach, namely Classification Trees Analysis (CTA). The propensity score estimation with the classification tree approach has a number of advantages, including: (i) the classification algorithm automatically selects variables for the model; (ii) the algorithm also automatically detects interactions in the data, so that those interactions do not have to be discovered and modeled explicitly as they do in logistic regression; and (iii) the tree's terminal nodes automatically supply the researcher with strata, eliminating the need to set stratification cut-points [1]. The model resulted using CTA is still unstable, because changes on a slight data learning can cause significant changes to the trees formed [3]. Therefore, the ensemble method is applied using bootstrap aggregating or bagging of the classification tree as a tool to improve the stability and predictive power of the classification tree by reducing the variance of a predictor variable [4].

Human Immunodeficiency Virus (HIV) and Acquired Immuno Deficiency Syndrome (AIDS) are among the diseases that become health problems in Indonesia. The HIV virus attacks the immune system in humans, so people affected by this virus become susceptible to various infections and then cause AIDS. Opportunistic infections (OIs) are a major cause of morbidity and mortality in people infected with HIV AIDS [5]. In an effort to reduce the high mortality rate due to OIs, research...
needs to be done on HIV AIDS patients. This study will examine the effect of giving antiretroviral therapy and counseling to opportunistic infections in HIV AIDS patients.

2. Methods

2.1. Propensity score

Propensity scores are defined as conditional probabilities for the observation \( i = 1, 2, \ldots, n \) of the treatment group \( Z_i = 1 \) compared to the control group \( Z_i = 0 \) based on \( x_i \) covariates observed, where randomization was not possible [6]. The propensity score method helps solve problems when the treatment group and controls are not appropriate for randomization and are able to provide valid estimates of the Average Treatment for Effect [7]. The propensity score can be written in the form of the following equation.

\[
e(x_i) = P(Z_i = 1|X_i = x_i)
\]

(1)

where \( 0 < e(x_i) = P(Z_i = 1|X_i = x_i) < 1 \), for each \( x \in X \) is a conditional probability value of a group based on the observed covariates [6]. This value can be used to reduce bias because covariate imbalances are observed.

Commonly used propensity score-based methods are Propensity Score Matching (PSM), Propensity Score Stratification (PSS), Inverse Probability of Treatment Weighting (IPTW) using the score proportion and Covariate Adjustment using the score propensity [8] and [9]. If applied appropriately a propensity score-based method can overcome the problem of selection bias and can reduce dimensions. If vector \( x \) has many covariates represented in many dimensions, the propensity score approach is able to reduce all dimensions to one-dimensional score, namely the propensity score [10].

In general, the steps of the propensity score analysis are (i) choosing the covariate as a confounding for the estimation of the propensity score; (ii) estimate the value of the propensity score and do matching (propensity score matching) or divide the strata based on the value of the propensity score (propensity score stratification) obtained in the previous step; (iii) testing whether the balance of covariates in confounding between treatment groups and control groups has been balanced; (iv) calculate the treatment effect [11] and [12].

The propensity score using CTA is obtained from the classification trees model which is in the form of a proportion value on the terminal nodes that are formed [1]. The estimated propensity score using CTA is formulated with the following Eq. (2).

\[
\hat{e}(x_i) = \frac{\exp(\hat{f}(x_i))}{1 + \exp(\hat{f}(x_i))} = \frac{\exp(\hat{\beta}(x_i))}{1 + \exp(\hat{\beta}(x_i))}
\]

(2)

Suppose there are \( p \) predictor variables \( x_1, x_2, \ldots, x_p \) affecting the response variable \( y \), thus forming a function \( y = g(x_1, x_2, x_3, \ldots, x_p) + \epsilon_1 \). This shows that each predictor variable \( x_1, x_2, \ldots, x_p \) is thought to influence the response variable \( y \). In a situation there is a confounding variable on the predictor variable, the variable is random and influenced by other predictor variables. For example, the confounding variable in the predictor variable is \( x_1 \), so the variable \( x_1 \) can then be denoted by \( z \). The relationship between the \( z \) variable and the other predictor variables \( x_2, x_3, \ldots, x_p \) can be written as a function \( z = f(x_2, x_3, \ldots, x_p) + \epsilon_2 \), so the predictor variable \( p - 1 \) can be reduced to one dimension score, which is the \( z \) variable. The \( z \) variable is used to estimate the propensity score. Propensity score can be estimated using several methods according to the scale of the confounding variable \( z \). If the confounding variable \( z \) is nominal or ordinal scale, then using methods: logistic regression, CART, Multivariate Adaptive Regression Splines (MARS) classification, or Support Vector Machine (SVM). Meanwhile, if the confounding variable \( z \) is interval or ratio scale, then using methods: linear regression, MARS prediction, or Support Vector Regression (SVR).

2.2. Evaluation of propensity score

One of the criteria the goodness of the propensity score method can be seen from how much bias can be reduced. A propensity score method is good if the method is able to reduce the bias that occurs in observational studies. Cochran explained the effectiveness of sub-classification in removing bias in observational studies [13]. The initial bias \( (B_i) \) in \( e(x) \) is shown in the following Eq. (3).

\[
B_i = E(e(x)|Z = 1) - E(e(x)|Z = 0)
\]

(3)

Bias in \( e(x) \) after sub-classification on the propensity score \( (B_k) \) and adjustments to the total weight of the subclass

\[
B_k = \sum_{i=1}^{K} \left( E(e(x)|Z = 1, e \in I_k) - E(e(x)|Z = 0, e \in I_k) \right) P(e \in I_k)
\]

(4)

there is \( K \) subclass with \( I_k \) is a set of fixed values of \( e \) that defines subclass \( k \). The magnitude of the bias reduced in \( e(x) \) after sub-classification of the propensity score is formulated in Eq. (5).

\[
PBR = \left( 1 - \frac{B_k}{B_i} \right) \times 100\%
\]

(5)

The covariate balance testing of the confounding variable can use the average difference test of two independent samples (t-test) if the data types used in the treatment and control groups are continuous, and use the difference test of the proportions of two independent samples (Z-test) if the data type used is categorical.

2.3. Bootstrap aggregating classification trees analysis

Breiman, Friedman, Olshen and Stone approximately 1980 developed the Classification and Regression Trees (CART) method for the topic of classification analysis, for categorical and continuous response variables. CART results in a classification tree if the response variable is categorical data type and result in a regression tree if the response variable is a continuous data type [14]. The response variables in this study are categorical data types, so the method used is Classification Trees Analysis (CTA). Classification using CTA involves learning data and testing data. Data learning is a dataset that is trained to make predictive models, while testing data is used to test the rules of classification trees resulted by data learning. The forming of the classification tree includes three things, that is classifier selection, terminal node determination, and class label marking [14].

The classifier selection aims to reduce heterogeneity on the node. The heterogeneity function used is the Gini Index shown in the following equation.

\[
i(t) = 1 - \sum_{j=1}^{J} p^2(j|t)
\]

(6)

where \( i(t) \) is the heterogeneity function of the Gini index and \( p(j|t) \) the proportion of class \( j \) members in node \( t \). The classifier selection is done by the Goodness of Split criteria using the following equation.

\[
\Delta i(s, t) = i(t) - p_i l(t_{iL}) - p_R l(t_{iR})
\]

(7)

where \( \Delta i(s, t) \) is the value of the Goodness of Split. The biggest Goodness of Split value indicates the best split. \( t \) and \( t_R \) as threshold value \( \Delta i(s, t) \) represents the change in heterogeneity in node \( t \) caused by splitting \( s \), where \( s \) is an element of \( S \). The node is developed by determining \( s^* \) which gives the greatest heterogeneous derivative value.

\[
\Delta i(s^*, t_{i}) = \max_{s \in S} \Delta i(s, t_{i})
\]

(8)
therefore $t_3$ has chosen to be $t_2$ and $t_3$ using $z^*$. Using the same way is determined the best split on $t_2$ and $t_3$ separately and so on [14].

The second stage of forming a classification tree is determining the terminal node. The criteria used to decide whether a node will not be split again or become a terminal node is (i) the development of the classification tree will stop if at the existing node there are only less than 5 observations ($n_t < 5$), where $n_t$ is number of observation; (ii) If the limit of the number of levels has been reached or the level of depth in the classification tree is maximized.

The last stage of forming a classification tree is class labels marking. Labeling the terminal node class is based on the rules of the highest number using the following equation.

$$p[j \mid t] = \max_{j} \frac{N_j(t)}{N(t)}$$

(9)

where $N_j(t)$ the number of observations of class $j$ at the terminal node $t$ and $N(t)$ number of observations at terminal node $t$.

Bootstrap aggregating or bagging is one of the ensemble methods introduced by Breiman in 1996 which aims to reduce the variance of a predictor so that it can improve the quality of predictions. Bootstrap is a method of random sampling with resampling. The results obtained through the bootstrap process are followed by an aggregating process by making a combined prediction (for example using a majority vote for the classification case, or the average for the regression case in the CART method).

The process of making bagging estimates using a classification tree is as follows [3]:

1. Bootstrap stages
   a. Take a random sample of size $n$ from data learning (L).
   b. Optimum classification tree shape from random samples that have been taken.
   c. Repeat steps a-b $R$ times to get the classification tree $R$. $R$ values used are 25, 50, 100, 125 and 150. Generally Bagging gives good results from low replication. If when replication is low the bagging gives poor results, bagging is carried out with replication up to 100 or more [3].
2. Aggregating Stages
   Make a combined prediction based on the R classification tree formed using the majority vote rule.

2.4. Average Treatment effect

If the balance covariate has been fulfilled, then the estimation of ATE is carried out. The stages of estimating ATE using the propensity score are as follows [15].

1. Divide the subject in a homogeneous subclass with $k = 1, \ldots, K$.
2. Determine the response average of the treatment and control groups in the subclass.

$$\overline{Y}_d = \frac{\sum_{i \in d} Y_{di}}{n_d}; \quad \overline{Y}_c = \frac{\sum_{i \in c} Y_{ci}}{n_c}; \quad n_t = \sum_{k=1}^{K} n_{tk}; \quad n_c = \sum_{k=1}^{K} n_{ck}$$

(10)

3. Calculate the estimated value of ATE ($\hat{\theta}$).

$$\hat{\theta} = \frac{\sum_{k=1}^{K} n_{tk} + n_{ck} (\overline{Y}_d - \overline{Y}_c)}{n_t + n_c}$$

(11)

4. Calculate the standard error of ATE.

$$SE(\hat{\theta}) = \sqrt{\sum_{k=1}^{K} \left( \frac{n_{tk} + n_{ck}}{n_t + n_c} \right)^2 \left( \frac{s_{tk}^2}{n_{tk}} + \frac{s_{ck}^2}{n_{ck}} \right)}$$

(12)

where $n_t$ and $n_c$ are many treatment and control group subjects, $n_k$ and $n_{tk}$ many treatment group subjects and control in the k-strata, $s_{tk}^2$ and $s_{ck}^2$ response variance for treatment and control groups in the k-strata.

Significant testing of the $\theta$ parameter is done to determine the effect of treatment given on the response variable.

$$H_0: \theta = 0 \text{ (There is no effect of treatment given to the response)}$$

$$H_1: \theta \neq 0 \text{ (There is effect of treatment given to the response)}$$

obtained test statistics based on Eqs. (11) and (12) are as follows.

$$Z = \frac{\hat{\theta}}{SE(\hat{\theta})}$$

(13)

The testing criteria is reject $H_0$, if $|Z| > Z_2$ or $p$-value < $\alpha$.

2.5. Research variables

This study used secondary data, which is observational data obtained from Grati Public Health Center in Pasuruan Regency. The data following 150 respondent that have response variable is the presence or absence of an infectious disease that accompanies people with HIV/AIDS, which is opportunistic infections. There are seven predictors variables, where one of them was used to confounding variable after empirical proof. The explanation and operational definition of research variables were shown in Table 1.

2.6. Analysis method

The steps of analysis in this study are the following:

1. Estimating the propensity score is using the CART method and CART bagging.
   a. Exploring data to find out the description of the variables that support the preparation of classification models.
   b. Determine the confounding variable ($z$) derived from the predictor variable ($x$).
   c. Build a function between the confounding variable ($z$) and the predictor variable ($x$) besides the confounding variable, that is $z = f(x_1, x_2, ..., x_{p-1}) + e_2$, with $z$ is confounding variable, and $(x_1, x_2, ..., x_{p-1})$ are predictor variables.
   d. Divide data into two parts, i.e. learning data and testing data.
   e. Estimating propensity score using the CART method.

   i. Determine the best sorting of predictor variables with Eqs. (6), (7), and (8).

   ii. Determine terminal node

   iii. Class labeling using Eq. (9)

2) Termination forms the classification tree.
3) Pruning the classification tree using formula:

$$R_{c}(T) = R(T) + a|\overline{T}|$$

Where,

$R_{c}(T)$: cost complexity measure, the complexity measure of a classification tree $T$ on complexity $a$

$R(T)$: re-substitution estimate (the errors proportion on the sub tree)

$a$: complexity parameter (complexity parameter)

$|\overline{T}|$: the number of terminal nodes in the classification tree $T$

4) The selection of the optimal classification tree uses a cross-validation V-fold estimate based on this formula.
Table 1. Research variables.

| No | Variables                        | Operational Definition                                                                 | Scale  |
|----|----------------------------------|----------------------------------------------------------------------------------------|--------|
| 1  | Opportunistic Infection (Y)      | Infectious diseases that accompany HIV/AIDS sufferers, such as pneumonia, tuberculosis, hepatitis, etc. Data is categorized by: 0 = There is opportunistic infection 1 = There is not opportunistic infection | Nominal |
| 2  | Age (X₁)                         | The length of life someone has lived and is calculated based on the last birthday. The data unit is in the form of years. | Ratio  |
| 3  | Knowledge (X₂)                   | Something that is known by patients about HIV/AIDS includes understanding, signs, symptoms, treatment, and the ways to prevent transmission of HIV. Data is categorized by: 0 = Not Good 1 = Good | Nominal |
| 4  | Self Conpect (X₃)                | Attitudes or acceptance towards oneself (individuals with HIV/AIDS). Data is categorized by 0 = Negative 1 = Positive | Nominal |
| 5  | Attitudes towards HIV/AIDS (X₄)  | Patient’s perception of the quality of life, which includes social relationships, physical well-being, psychological, and spiritual. Data is categorized by: 0 = Negative 1 = Positive | Nominal |
| 6  | Family Support (X₅)              | Patient’s perceptions of family support include emotional support, moral, information, and social. Data is categorized by: 0 = Does not support 1 = Support | Nominal |
| 7  | Time suffering from HIV/AIDS (X₆)| The duration of suffering from HIV/AIDS starts from the first diagnosis until the study time. The data unit is in the form of months. | Ratio  |
| 8  | Treatments (X₇)                  | Giving different treatments for patients. Data is categorized by: 0 = Only get ARV treatment 1 = Get ARV treatment and accompaniment | Nominal |

\[ R^CV (T_i) = \frac{1}{V} \sum_{i=1}^{V} R^CV (T_i^{(v)}) \]

Where,  
\( R^CV (T_i) \) : the errors total proportion of the folding cross validation estimator \( V \)  
\( V \) : the number of folds used

5) Find the accuracy of the CART classification tree model following the formula below.

Sensitivity = \( \frac{n_{11}}{n_{10} + n_{11}} \)

Specificity = \( \frac{n_{00}}{n_{00} + n_{01}} \)

Accuracy = \( \frac{n_{00} + n_{11}}{n_{10} + n_{11} + n_{01} + n_{11}} \)

AUC = \( \frac{1}{2} (\text{sensitivity} + \text{specificity}) \)

6) Getting the best CART model is used to estimate the value of Propensity Score Stratification (PSS).
7) Evaluate the PSS CART method by testing the balance of the covariate and calculating the Percent Bias Reduction (PBR) based on Eqs. (3), (4), and (5).

f. Estimating propensity score using the CART bagging method.

1) Take bootstrap samples \( L_b \) from learning data to build a classification tree. Repeat \( R \) times this step to get the classification tree \( \phi_1(x), ..., \phi_R(x) \). Recommendations from previous researchers, the determination of the number of bootstrap sample replications performed were 25, 50, 75,100, 125, 150, and 175 times.

2) Do the aggregating process to get the best classification tree.

3) Find the accuracy of the CART bagging classification tree model according to formula on Step 1.e.5

4) Getting the best CART bagging model is used to estimate the value of propensity score with Propensity Score Stratification (PSS).

5) Evaluate the PSS CART bagging method by testing the balance of the covariate and calculating the Percent Bias Reduction (PBR) based on Eqs. (3), (4), and (5).

2. Applying the estimation of propensity score is using the PSS CART and CART bagging method PSS. In the case of this study used opportunistic infections of HIV/AIDS patients based on the previous steps, so that the Average Treatment Effect (ATE) can be obtained and then test the significance of the ATE.

3. Comparing the accuracy of the classification results and Percent Bias Reduction (PBR) for the PSS CART method and the PSS bagging CART method.

3. Results and discussion

The variables have ratio scale in this study, i.e. age and time suffering of HIV/AIDS patient. The characteristics were shown in Table 2.
Table 2 shows the youngest patients with OIs aged 23 years and the oldest aged 61 years, with an average age of 34.82 years and a standard deviation of 8.57 years. It shows that the risk of developing OIs does not only occur in patients with advanced age, but patients of productive age are also at risk of developing OIs. Meanwhile, the duration of suffering from patients who had the most rapid opportunistic infections was 11 months, and the longest was 147 months, with an average of 37.95 months and a standard deviation of 26.21 months. The average duration of suffering from patients who did not have an opportunistic infection was 40.16 months, on average longer than patients who had an opportunistic infection, which was 37.95 months. It means HIV/AIDS patients have a relatively long time to suffer, but they are still at risk of opportunistic infections.

The formation of CART classification tree to estimate the value of propensity score is carried out on the confounding variable (administration of ARV therapy (Z)). The initial step in forming a CART classification tree is to divide research data into learning data and testing data. The division of data into two parts aims to obtain an optimal model.
Sharing the proportion of learning data and testing data does not have specific provisions, but learning data must be more than testing data. The formation of classification tree models is done by using several alternative proportions of learning data, i.e. 70%, 75%, 80%, 85% and 90% of total data. The proportion of learning data with the highest accuracy of the classification results based on testing data will be chosen for further analysis (see Table 3).

The estimated PSS CART value is obtained based on the optimal classification tree model formed in Figure 1. The prediction results of the CART classification tree in the form of proportion values at the terminal nodes are the value of the propensity score. The classification tree automatically provides strata through homogeneous terminal nodes, eliminating the need to set stratification cut points. Five terminal nodes are obtained the optimal classification tree, so the number of strata is five.

The accuracy results of the optimal classification tree CART method using learning and testing data are shown in the table below. Table 4 shows 77 observations from the learning data appropriately classified, i.e. 63 observations appropriately classified as patients who received ARV therapy without assistance, and 14 observations appropriately classified as patients who received ARV therapy with assistance.

### Table 2. Patient Characteristics based on Opportunistic Infections Status.

| Variable                  | Measure       | Opportunistic Infections (OIs) Status |
|---------------------------|---------------|---------------------------------------|
|                           |              | There is OIs | There is not OIs |
| **Age**                   | Mean | 34.82        | 33.22        |
|                           | StDev | 8.57         | 7.89         |
|                           | Min  | 23.00        | 23.00        |
|                           | Max  | 61.00        | 56.00        |
| **Time suffering from HIV/AIDS** | Mean | 37.95        | 40.16        |
|                           | StDev | 26.21        | 27.43        |
|                           | Min  | 11.00        | 3.00         |
|                           | Max  | 147.00       | 146.00       |

### Table 3. Classification results of testing data based on the learning data.

| Proportion of Learning Data | Classification Results | AUC (%) |
|-----------------------------|------------------------|---------|
| 70%                         | 63.04                  | 59.73   |
| 75%                         | 55.26                  | 49.36   |
| 80%                         | 54.84                  | 50.68   |
| 85%                         | 52.17                  | 45.53   |
| 90%                         | 50.00                  | 28.57   |

The highest accuracy of the classification results is 63.04%, so the proportion of learning data is 70%. The number of research data points is 150, which is the learning data was used 104 data points, while the testing data was used 46 data points.

### Table 4. Classification results of testing data based on the learning data.

| Proportion of Learning Data | Classification Results | AUC (%) |
|-----------------------------|------------------------|---------|
| 70%                         | 63.04                  | 59.73   |
| 75%                         | 55.26                  | 49.36   |
| 80%                         | 54.84                  | 50.68   |
| 85%                         | 52.17                  | 45.53   |
| 90%                         | 50.00                  | 28.57   |

The highest accuracy of the classification results is 63.04%, so the proportion of learning data is 70%. The number of research data points is 150, which is the learning data was used 104 data points, while the testing data was used 46 data points.
The accuracy of the classification results obtained in the learning data is 74.04% with AUC of 73.90%. The accuracy of the optimal classification tree constructed from learning data when applied to 30% of the testing data shows that the number of observations that were correctly classified was 28 observations, with details of 24 observations appropriately classified as patients receiving ARV therapy without assistance, and 4 observations appropriately classified as patients receiving ARV therapy with assistance. The accuracy of the classification results obtained in the testing data is 60.87%, with the AUC of 53.33%.

The application of bagging techniques aims to improve the classification accuracy and consistency of a single classification tree. The bootstrap sample resampling process has made, which is 125 times, and the aggregating process is carried out by making a combined prediction to obtain the results of the bagging CART classification tree. Where, the accuracy results of the optimal bagging CART method are shown Table 6.

The number of classification trees formed in accordance with the replication of bootstrap samples have made, which is 125 times, and the aggregating process is carried out by making a combined prediction to obtain the results of the bagging CART classification tree. Tree pruning can be done by using cost complexity pruning in order to get a better clas-

| Table 4. Accuracy of optimal classification tree. |
|-----------------------------------------------|
| Actual | Prediction | Accuracy (%) | AUC (%) |
|        |            |              |        |
|        | 0 | 1 |
| Learning Data | 63 | 22 | 74.04 | 73.90 |
| Testing Data | 5 | 14 | 60.87 | 53.33 |

| Table 5. Accuracy of classification. |
|--------------------------------------|
| Number of Replication | Accuracy (%) |
|                        | Training Data | Testing Data |
| 25 | 96.15 | 60.87 |
| 50 | 97.11 | 56.52 |
| 75 | 97.11 | 58.70 |
| 100 | 97.11 | 60.87 |
| 125 | 97.11 | 60.87 |
| 150 | 97.11 | 60.87 |
| 175 | 97.11 | 60.87 |

| Table 6. Accuracy of bagging CART. |
|-------------------------------------|
| Actual | Prediction | Accuracy (%) | AUC (%) |
|        |            |              |        |
|        | 0 | 1 |
| Learning Data | 66 | 20 | 78.85 | 82.81 |
| Testing Data | 26 | 13 | 63.04 | 54.76 |

The number of classification trees formed in accordance with the replication of bootstrap samples have made, which is 125 times, and the aggregating process is carried out by making a combined prediction to obtain the results of the bagging CART classification tree. Where, the accuracy results of the optimal bagging CART method are shown Table 6.

The classification tree resulted from the CTA bagging algorithm is a very complex classification tree because this tree is formed by all predictor variables. The number of classification trees formed according to many bootstrap sample replications was made 125 times, and aggregating processes were carried out by making a combined prediction to obtain the results of the CTA bagging classification tree. Tree pruning can be done by using cost complexity pruning in order to get a better clas-

| Table 7. Propensity score estimation results. |
|-----------------------------------------------|
| No | $\hat{x}(x_i)$ |
|----|----------------|
| 1-7 | 0.552 | 0.992 | 0.808 | 0.768 | 0.920 | 0.864 | 0.632 |
| 8-14 | 0.952 | 0.968 | 0.960 | 1.000 | 0.808 | 0.952 | 0.976 |
| 15-21 | 0.976 | 0.984 | 0.984 | 0.944 | 0.984 | 0.970 | 0.984 |
| 22-28 | 0.992 | 0.896 | 0.720 | 0.792 | 0.968 | 0.496 | 0.816 |
| 29-35 | 1.000 | 0.752 | 0.976 | 0.816 | 0.992 | 0.944 | 0.992 |
| 36-42 | 0.808 | 0.936 | 0.352 | 0.936 | 0.800 | 0.984 | 0.752 |
| 43-49 | 0.976 | 0.816 | 0.960 | 0.976 | 0.784 | 0.928 | 0.968 |
| 50-56 | 0.928 | 0.968 | 0.968 | 0.752 | 0.720 | 0.672 | 0.736 |
| 57-63 | 0.504 | 0.616 | 0.728 | 0.984 | 0.632 | 0.936 | 0.984 |
| 64-70 | 0.992 | 0.984 | 0.912 | 0.992 | 0.968 | 0.120 | 0.448 |
| 71-77 | 0.880 | 0.996 | 0.520 | 0.472 | 0.512 | 0.192 | 0.120 |
| 78-84 | 0.312 | 0.232 | 0.504 | 0.672 | 0.384 | 0.304 | 0.720 |
| 85-91 | 0.504 | 0.680 | 0.184 | 0.440 | 0.056 | 0.648 | 0.632 |
| 92-98 | 0.520 | 0.112 | 0.424 | 0.328 | 0.688 | 0.672 | 0.880 |
| 99-104 | 0.080 | 0.728 | 0.160 | 0.600 | 0.296 | 0.024 |
sification of new data (data testing). The CTA bagging model from the pruning results is then used to estimate the value of the propensity score. The results of the CTA bagging classification tree prediction in the form of a proportion value which is also an estimated value of the propensity score. The estimated results of the propensity score \( \hat{b}(x_i) = \hat{p}(Z = j|x_i) \) of the CTA bagging classification tree are shown in Table 7.

After estimating the propensity score, the stratification process is then carried out by dividing the research subjects into several sub-classes (strata) based on the homogeneity of the resulting propensity score. The division of research subjects into several strata aims to get the strata group which shows no difference between the treatment group and the control group. The results of the stratification of the propensity score are shown in the following table.

The results of the PSS bagging CTA analysis need to be evaluated to find out whether the method used is able to provide the appropriate results or not. One of the criteria for the goodness of the PSS bagging CTA method can be seen from the results of the covariate balance test. The results of the balance test at each covariate in each stratum are shown in Table 8.

Next, test the significance of the ATE (\( \theta \)) parameters to determine the effect of ARV Getting ARV and Counseling (Z) treatment on opportunistic infections (Y) when the influence of other covariates has reduced. The hypothesis used is as follows.

H\(_0\): \( \theta = 0 \) (there is not a treatment effect given to the response)

H\(_1\): \( \theta \neq 0 \) (there is a treatment effect given to the response)

The estimated results of the ATE value and the ATE standard error are shown in Table 10.

Table 9 shows that prior to PSS bagging CTA there are two covariates that are not balanced, namely attitude to HIV AIDS (X\(_3\)) and family support (X\(_5\)) because they have a p-value < 0.05. After PSS bagging CTA divided into 5 strata, the test results are balanced against each covariate. There are covariates that cannot be tested for balance, namely the family support covariate (X\(_5\)) in strata 1 and strata 5, while the other covariates balance in each stratum with p-value > 0.05.

The estimated results of the ATE value and the ATE standard error are shown in Table 10.

Table 11 shows the results of the ATE estimation using the PSS bagging CTA method. Obtained ATE value of 0.222 with SE ATE of 0.038.

### Table 8. Propensity score stratification results.

| Strata | \( \hat{e}(x_i) \) |
|--------|------------------|
| Strata 1 | 0.024 0.056 0.080 0.096 0.112 0.120 0.120 0.160 0.184 0.200 0.212 0.232 0.296 0.304 0.312 0.328 0.352 |
| Strata 2 | 0.472 0.496 0.504 0.504 0.504 0.512 0.520 |
| Strata 3 | 0.720 0.720 0.720 0.720 0.728 0.728 0.728 0.728 0.728 0.728 0.728 0.728 0.728 0.728 0.728 0.728 0.728 0.728 0.728 |
| Strata 4 | 0.864 0.880 0.880 0.880 0.880 0.880 0.880 0.880 0.880 0.880 0.880 0.880 0.880 0.880 0.880 0.880 0.880 0.880 0.880 |
| Strata 5 | 0.968 0.976 0.976 0.976 0.976 0.976 0.976 0.976 0.976 0.976 0.976 0.976 0.976 0.976 0.976 0.976 0.976 0.976 1.000 |

### Table 9. Results of covariate balance testing.

| Covariate | Before PSS (p-value) | After PSS (p-value) |
|-----------|----------------------|---------------------|
|           | Strata 1 | Strata 2 | Strata 3 | Strata 4 | Strata 5 |
| Age (X\(_1\)) | 0.552    | 0.129    | 0.883    | 0.877    | 0.992    | 0.992    |
| Knowledge (X\(_2\)) | 0.460    | 0.798    | 0.067    | 0.588    | 0.650    | 0.684    |
| Attitude (X\(_3\)) | 0.046\(^*\) | 0.717    | 0.456    | 0.154    | 0.427    | 0.629    |
| Self Concept (X\(_4\)) | 0.090    | 0.717    | 0.199    | 0.639    | 0.427    | 0.175    |
| Family Support (X\(_5\)) | 0.002\(^*\) | NaN      | 0.402    | 0.127    | 0.158    | NaN      |
| Time Infected (X\(_6\)) | 0.514    | 0.878    | 0.220    | 0.843    | 0.121    | 0.121    |

\(^*\) declare the covariate is not balanced.

### Table 10. ATE estimation results.

| ATE (\( \hat{\theta} \)) | SE (\( \hat{\theta} \)) | \( |Z| \) | p-value |
|--------------------------|----------------------|-------|--------|
| 0.222                    | 0.038                | 5.815 | 0.000  |

### Table 11. PBR using PSS Bagging CTA.

| Bias Before PSS Bagging CTA | Bias After PSS Bagging CTA | PBR (%) |
|-----------------------------|---------------------------|--------|
| 0.445                       | 0.046                     | 89.54  |
The p-value < 0.05, which means there is a significant effect of ARV therapy and counseling to opportunistic infections in HIV AIDS patients. The goodness of the propensity score method can be seen from how much bias can be reduced by that method. Obtained percent bias reduction (PBR) calculation results for the PSS bagging CTA method are shown in the following table.

Table 11 shows the results of the PBR calculation for the PSS bagging CTA method using 5 strata. CTA bagging method is able to reduce the bias higher that is equal to 89.54%.

4. Conclusion

Some things that can be concluded in this study are Propensity Score Stratification (PSS) using Bagging Classification Trees Analysis (CTA) divided into 5 strata and balanced in each stratum. Percent Bias Reduction (PBR) produced is 89.54%. The results of the Average Treatment Effect (ATE) test showed a significant effect on the provision of ARV therapy and counseling to opportunistic infections. Patients receiving ARV therapy and assistance can reduce opportunistic infections in HIV AIDS patients.

Declarations

Author contribution statement

B. W. Otok, Purhadi: Conceived and designed the experiments. M. Musa, S. D. P. Yasmirullah: Analyzed and interpreted the data; Wrote the paper.

Funding statement

This research received financial support from Ministry of Research and Technology/National Research and Innovation Agency of Republic Indonesia (RISTEK-BRIN). No: .3/AMD/E1/KP.PTNNBH/2020

Competing interest statement

The authors declare no conflict of interest.

References

[1] J.K. Luellen, W.R. Shadish, M.H. Clark, Propensity scores an introduction and experimental test, Eval. Rev. 29 (2005) 530–558.
[2] S. Litfin, J. Jarkovsky, T. Pavlik, J. Spinar, L. Dusek, Why to use propensity score in observational studies? Case study based on data from the Czech clinical database AHEAD 2006-09, Cor Vasa 55 (2013) 383–390.
[3] D.B. Rubin, Classification and regression trees, bagging, and boosting, Handb. Stat. 24 (2005) 303–329.
[4] L. Breiman, Bagging predictors, Mach. Learn. 24 (1996) 123–140.
[5] P.R. Shahapur, R.C. Bidri, Recent trends in the spectrum of opportunistic infections in human immunodeficiency virus infected individuals on antiretroviral therapy in South India, J. Nat. Sci. Biol. Med. 5 (2014) 392–396.
[6] P.R. Rosenbaum, D.B. Rubin, The central role of the propensity score in observational studies for causal effects, J. Biometrika 70 (1983) 41–55.
[7] S. Guo, M. Fraser, Propensity Score Analysis: Statistical Methods and Applications, SAGE, CA, 2010.
[8] P.C. Austin, M.M. Mamandi, A comparison of propensity score methods: a case-study estimating the effectiveness of post-AMI statin use, Stat. Med. 25 (2006) 2084–2106.
[9] P.C. Austin, An introduction to propensity score methods for reducing the effects of confounding in observational studies, Multivariate Behav. Res. 46 (2011) 399–424.
[10] F. Kamangar, Confounding variables in epidemiologic studies: basics and beyond, Arch. Iran. Med. 15 (2012) 508–516.
[11] B.W. Otok, A. Aisyah, Purhadi, S. Andri, Propensity score matching of the gymnastics for diabetes mellitus using logistic regression, in: AIP Conference Proceedings, 1913, 2017, pp. 1–8.
[12] I. Yanovitzky, E. Zanutto, R. Hornik, Estimating causal effects of public health education campaigns using propensity score methodology, Eval. Progr. Plann. 28 (2005) 209–220.
[13] P.R. Rosenbaum, D.B. Rubin, Reducing bias in observational studies using subclassification on the propensity score, J. Am. Stat. Assoc. 79 (1984) 516–524.
[14] L. Breiman, J. Friedman, R. Olshen, C. Stone, Classification and Regression Trees, Wadsworth, CA, 1984.
[15] W. Tu, X. Zhou, A bootstrap confidence interval procedure for the treatment effect using propensity score subclassification, Health Serv. Outcome Res. Methodol. 3 (2012) 135–147.

No additional information is available for this paper.