

# Implementasi *Data Mining Clustering* Penduduk Miskin Menggunakan Algoritma C4.5 Untuk Merekomendasikan Bansos Pada RT. 05/01 Kelurahan Jatimekar Bekasi

Sri Lestari<sup>1</sup>, Sony Agustiansyah<sup>1</sup>

<sup>1</sup> Program Studi Sistem Informasi, STIKOM CKI

Jl. Raden Inten II No.8, RT.5/RW.14, Duren Sawit, Kec. Duren Sawit, Kota Jakarta Timur, DKI Jakarta

E-mail: [Sri.lestari1203@gmail.com](mailto:Sri.lestari1203@gmail.com)

Naskah Masuk: 05 Agustus 2022; Diterima: 26 Agustus 2022 Terbit: 17 Maret 2023

---

## ABSTRAK

---

**Abstrak** - Kemiskinan merupakan persoalan dua dimensi yang berkaitan dengan ketidakmampuan individu secara finansial, budaya sosial, politik dan pergaulan sosial. Untuk mengatasi kemiskinan, otoritas publik melakukan berbagai cara dengan menyelenggarakan program bantuan seperti program keluarga harapan (PKH). Program keluarga harapan (PKH) merupakan program kesejahteraan sosial bersyarat pemerintah yang diperuntukkan kepada masyarakat kurang mampu yang telah terdaftar dalam kumpulan informasi tersusun yang memenuhi syarat dan kriteria. Pada penelitian ini dilakukan kualifikasi *data mining* memakai perhitungan C4.5 untuk menentukan kelayakan penerima bantuan program keluarga harapan (PKH). Tindakan dalam penelitian ini adalah pengumpulan informasi yang didapat dari kepala RT. 05/01 Kelurahan Jatimekar Bekasi. selanjutnya penentuan informasi yaitu dengan kriteria jumlah anak usia dini, ibu hamil, siswa sekolah dasar, siswa sekolah menengah pertama, siswa sekolah menengah atas, pekerjaan, lansia, cacat berat, status tempat tinggal, jenis dinding terluas, terima raskin, fasilitas toilet, tempat pembuangan sampah dengan kriteria layak dan kriteria tidak layak. Setelah penentuan informasi maka dilakukan pra-pemrosesan informasi sebanyak 2.244 catatan informasi yang mencakup pengecekan informasi kosong, informasi ganda dan kesalahan penulisan. selanjutnya transformasi untuk memudahkan pengolahan informasi. Selanjutnya proses *data mining* dengan perhitungan C4.5 hingga terbentuknya pohon keputusan. Pengujian memakai *confusion matrix* menghasilkan ketepatan senilai 86,37% dengan perbandingan 90% informasi latih dan 10% informasi uji. Berdasarkan hasil pengujian tersebut perhitungan C4.5 termasuk pada tingkat Klasifikasi Bagus.

**Kata kunci:** Perhitungan C4.5, Kemiskinan, Penataan, Pohon Keputusan, Program Keluarga Harapan (PKH)

---

## ABSTRACT

---

**Abstract** - Poverty is a two-dimensional problem related to the inability of individuals financially, socially culturally, politically and socially. To overcome poverty, public authorities take various ways by organizing assistance programs such as the Family Hope Program (PKH). The Family Hope Program (PKH) is a conditional government social welfare program intended for underprivileged communities who have been registered in a structured collection of information that meets the requirements and criteria. In this study, information mining qualifications were carried out using the C4.5 calculation to determine the eligibility of recipients of the Family Hope Program (PKH) assistance. The action in this research is the collection of information obtained from the head of the RT. 05/01 Jatimekar Bekasi Village. The next step is to determine the information based on the criteria for the number of early childhood children, pregnant women, elementary school students, junior high school students, high school students, occupations, elderly, severely disabled, residence status, widest type of wall, accept raskin, toilet facilities, place of residence. waste disposal with proper criteria and unfeasible criteria. After determining the information, a total of 2,244 information records were pre-processed, which included checking for blank information, duplicate information and writing errors. further transformation to facilitate information processing. Furthermore, the process of mining information by calculating C4.5 to form a decision tree. Testing using the confusion matrix resulted in an accuracy of 86.37% with a comparison of 90% of training information and 10% of test information. Based on the test results, the calculation of C4.5 is included in the Good Classification level.

**Keywords:** Calculation of C4.5, Poverty, Arrangement, Decision Tree, Family Hope Program (PKH)

Copyright © 2023 Jurnal Teknik Elektro dan Komputasi (ELKOM)

---

## 1. PENDAHULUAN

Persoalan kesejahteraan sosial di Indonesia yang harus dipelajari terus-menerus secara bersangkutan adalah kemiskinan. Kemiskinan menjadi persoalan pokok yang harus segera dihadapi sebagaimana menjadi tujuan nasional yang terdapat dalam UUD tahun 1945. Kemiskinan memang persoalan yang selalu dihadapi individu. Persoalan kemiskinan memang sama tuanya dengan individu itu sendiri dan keterlibatan persoalannya dapat melibatkan berbagai segi kehidupan individu. Kemiskinan merupakan persoalan dua dimensi yang berkaitan dengan ketidakmampuan individu, baik itu secara finansial, budaya sosial, politik dan pergaulan sosial. Dengan kata lain bahwa kemiskinan itu merupakan persoalan sosial yang jenisnya mendunia dan sudah menjadi sorotan dunia meskipun pengaruh dari kemiskinan tersebut berbeda-beda. Berdasarkan informasi yang dirilis oleh BPS pada tahun 2019 tercatat bahwa nilai kemiskinan di Indonesia senilai 9,22% dengan jumlah penduduk 24,74 juta orang pada september 2019, nilai tersebut menurun 0,36 juta orang pada bulan maret 2018 dan menurun menjadi 0,88 juta orang pada bulan september 2018. Sedangkan jumlah penduduk kurang mampu di kabupaten Bekasi sejumlah 72,21 ribu orang dengan persentase 10,53% penduduk kabupaten Bekasi tercatat sebagai penduduk kurang mampu.

Banyaknya jumlah penduduk kurang mampu, otoritas publik berupaya menghadapi persoalan kemiskinan dengan berbagai program bantuan seperti: bantuan langsung tunai (BLT), beras miskin (Raskin), bantuan pokok dan *family hope program* (PKH). *Family hope program* (PKH) merupakan program kesejahteraan sosial bersyarat otoritas publik yang diperuntukkan kepada masyarakat kurang mampu yang telah terdaftar dalam kumpulan informasi tersusun program penanganan masyarakat kurang mampu yang memenuhi syarat dan kriteria. Kriteria keluarga penerima manfaat *family hope program* (PKH) meliputi komponen sekolah, kesehatan dan kesejahteraan sosial dengan ketentuan yang telah ditetapkan yaitu terdapat siswa tingkat SD-SMA, ibu hamil, ibu menyusui, lansia dan cacat berat. Sebagai sebuah program bansos bersyarat, *family hope program* membuka jalan bagi keluarga kurang mampu agar dapat memanfaatkan berbagai fasilitas administrasi sekolah, pergaulan dan fasilitas administrasi kesehatan yang tersedia disekeliling mereka guna untuk mempertahankan tingkat kesejahteraan sosial sebagaimana sesuai dengan amanat susunan dan nawacita presiden republik Indonesia.

Sebelum keluarga penerima manfaat *family hope program* mendapatkan bantuan, keluarga penerima manfaat terlebih dahulu melakukan proses penentuan berupa pemeriksaan dan pengesahan informasi yang dilakukan oleh wakil atau petugas yang mencatat keluarga penerima manfaat *family hope program* tersebut. Pada proses pemeriksaan dan pengesahan informasi yang dilakukan oleh petugas masih dilakukan manual dengan mengunjungi rumah keluarga penerima manfaat *family hope program* sehingga memakan waktu yang panjang dalam menentukan keluarga yang layak mendapatkan bantuan *family hope program*. Maka dibutuhkan sebuah sistem tiruan yang dapat melakukan kualifikasi keluarga penerima manfaat bantuan *family hope program* dengan laju. Dengan bantuan *data mining* sehingga memudahkan pihak instansi dalam menentukan keluarga yang masih layak atau tidak layak mendapatkan bantuan *family hope program*.

*Data mining* adalah suatu kata yang digunakan untuk menjabarkan pengetahuan di dalam kumpulan informasi. *Data mining* adalah proses yang memakai cara statistik, matematika, *artificial intelligence* dan *machine learning* untuk mengutip dan mengenali informasi bermanfaat dan pengetahuan yang dibangun dari kumpulan informasi besar. Ada beberapa penelitian *data mining* yang dipakai untuk menelaah kelayakan penerima bantuan *family hope program* yaitu penelitian tentang sistem pendukung penetapan kelayakan penerima *family hope program* (PKH) memakai perhitungan *analytic network process*. Berdasarkan *confusion matrix* penelitian ini menghasilkan tingkat ketepatan sistem pendukung penetapan senilai 39,64%, pengulangan senilai 39,29% dan nilai ketepatan senilai 69,12%. Sedangkan pengujian kegunaan sistem memakai *blackbox testing* menghasilkan 93,34% sistem berjalan dengan bagus.

Perhitungan C4.5 merupakan salah satu dari tipe *decision tree* yang sangat terkenal. *Decision tree* sangat bermanfaat untuk menjelajah informasi, menemukan hubungan tersulit antara sejumlah kandidat faktor masukan dengan sebuah faktor tujuan. Perhitungan C4.5 dapat menghadapi informasi penomoran dan berlainan dengan memakai perbandingan perolehan (*gain ratio*) sebelum mendapatkan nilai perbandingan perolehan, perlu dilakukan perhitungan nilai informasi dalam satuan dari suatu kumpulan objek, yaitu dengan memakai konsep entropi. Dengan memakai *decision tree*, peneliti tidak perlu melakukan perkiraan pada pembagian dimensi tinggi ataupun pengukuran tertentu dari pembagian kelas tersebut. Karena strategi ini memakai kriteria yang nilainya lebih sedikit pada setiap *internal node* tanpa banyak mengurangi mutu ketetapan yang dihasilkan.

Berdasarkan latar belakang diatas maka pada penelitian ini akan dilakukan penerapan perhitungan C4.5 untuk mengkualifikasi kelayakan penerima bantuan *family hope program* berdasarkan kriteria-kriteria yang telah ditetapkan. Penelitian ini bertujuan untuk membangun sistem kualifikasi sebagai penunjang dalam penentuan kelayakan anggota penerima bantuan *family hope program*. Hal ini dapat dimanfaatkan sebagai strategi pendataan pelaksanaan *Family Hope Program* kota Bekasi dan unsur-unsur yang merusak pelaksanaan *Family Hope Program*.

## 2. TINJAUAN PUSTAKA

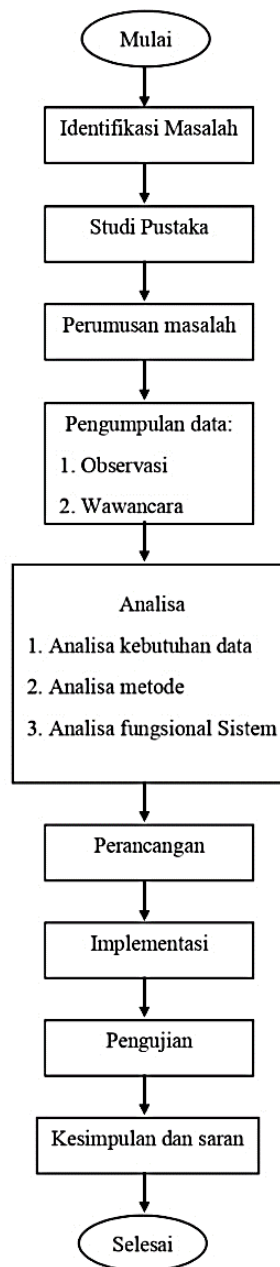
Berdasarkan penelitian sebelumnya terkait kualifikasi *data mining* dengan strategi perhitungan C4.5 oleh Muhammad Amirsyah Tanjung [1] yang berjudul “Pemeriksaan Kelayakan Penerima *Family Hope Program* (PKH) Memakai Perhitungan C4.5” menghasilkan nilai ketepatan sistem senilai 96,67%, yaitu bahwa aturan yang dihasilkan tingkat ketepatan mendekati 100%. Dimana strategi yang telah direncanakan dan diuji tingkat ketepatannya dengan memasukan dan pengujian yang berasal dari informasi latih dengan memakai validasi terpisah pada *software RapidMiner 5.3* untuk menguji tingkat ketepatan. Penelitian lain dikerjakan oleh Rajes Wasimson Sinaga [2] berjudul “Pemeriksaan *Data Mining* Penentuan Penerima BLT pada Desa Pamatang Purba dengan Perhitungan C4.5” menghasilkan kualifikasi penggunaan perhitungan C4.5 dan pengujian dengan *software RapidMiner* maka diperoleh pengaruh yang paling berpengaruh terhadap kelayakan penerima program bantuan langsung tunai (BLT). Penelitian lain dikerjakan oleh Rika Nur Adiha [3] berjudul “*Application of Data Mining in Determining Social Assistance Recipients with C4.5 Algorithm*” dengan hasil dapat mempermudah dan mempercepat aparatur otoritas publik dalam menentukan penerima bantuan sosial pada Gunung Kantor Kecamatan Maligas. Oleh karena itu, dapat diperoleh kesimpulan perhitungan C4.5 sesuai untuk dipergunakan pada penelitian penentuan kelayakan penerima bantuan *family hope program* (PKH).

Penelitian lain dikerjakan oleh Yuris Alkhalifi [4] berjudul “Perbandingan Perhitungan *Naive Bayes* Dan Perhitungan C.45 Pada Kualifikasi Masyarakat Miskin Penerima Bantuan Pangan Non Tunai Di Desa Wanasari Kabupaten Karawang” menghasilkan tingkat ketepatan pada strategi *Naive Bayes Classifier* adalah senilai 98,89% dan untuk tingkat ketepatan pada strategi *Decision Tree C.45* adalah senilai 95,00%. Maka dihasilkan kesimpulan bahwa dalam penelitian ini perhitungan dengan tingkat ketepatan tertinggi adalah strategi perhitungan *Naive Bayes Classifier* dengan perbandingan tingkat ketepatan senilai 3,89%. Penelitian lain dikerjakan oleh Laila Qadrini [5] berjudul “Pohon Keputusan Dan *Adaboost* Pada Kualifikasi Penerima Program Bansos” dengan hasil strategi kualifikasi Pohon Keputusan dan *Adaboost* sama-sama memperoleh hasil ketepatan yang bagus yaitu senilai 94% dan 95% dan strategi kualifikasi Pohon Keputusan dan *Adaboost* bagus digunakan untuk penentuan penerima bansos secara bagus. Penelitian lain dikerjakan oleh Siti Badriah [6] berjudul “Kualifikasi Perhitungan C4.5 Dalam Penentuan Penerima Bantuan Covid-19” menghasilkan aturan dari perhitungan C4.5 dengan nilai ketepatan yang terbagus terdapat pada perbandingan 90% informasi latih dan 10% informasi pengujian dengan nilai ketepatan senilai 79,54%. Penelitian lain dikerjakan oleh Muafi [7] berjudul “Kualifikasi Pemberian Bantuan Kepada Masyarakat Terkena Covid-19 Dengan Memakai Perhitungan C4.5 Di Desa Demung Kecamatan Besuki Kabupaten Situbondo” dengan hasil aturan atau *decision tree C4.5*, dengan ketepatan yang terbagus 90% latih dan 10% pengujian, dengan nilai ketepatan 74.09%.

Penelitian lain dikerjakan oleh Nabilah Iftah Nella [8] berjudul “Kualifikasi Penerima Bantuan *Family Hope Program* Memakai Perhitungan Pohon Keputusan C4.5” menghasilkan *confusion matrix* menghasilkan ketepatan 71,1%, ketelitian 90,9%, pengulangan 70,4%, dan f1-skor 79,3%. Hasil dari proses kualifikasi digambarkan dengan penggunaan dasbor yang menghasilkan *SUS* Skor senilai 87,5 yang memperlihatkan dasbor dapat diterima pemangku kepentingan. Penelitian lain dikerjakan oleh Fuji Astuti [9] berjudul “Kualifikasi Penerima Bansos Memakai Perhitungan C.45 Kantor Kuwu Desa Bangodua” dengan hasil pengujian berhasil mengkualifikasikan 311 informasi yang dites. Sehingga berhasil memperkirakan penerima bansos dengan tingkat ketepatan senilai 55.59%. Penelitian selanjutnya dikerjakan oleh Hariati [10] berjudul “Penerapan Perhitungan C4.5 pada Penentuan Penerima Program Bantuan Pemerintah Daerah Di Kabupaten Kutai Kartanegara”. Pada penelitian ini, hasil kualifikasi memperlihatkan bahwa ketepatan yang diperoleh meraih 97,14% pada informasi latih 85% dengan informasi 105 dan nilai *AUC* 0,913 yang termasuk dalam klasifikasi bagus sekali, hal ini memperlihatkan bahwa perhitungan C4.5 sesuai dipakai untuk penentuan penerima program bantuan dan aturan hasil kualifikasi dapat diterapkan dalam produksi sistem kualifikasi penerima bantuan.

## 3. METODE PENELITIAN

Metodologi penelitian adalah proses penelitian yang dilakukan secara sistematis. Metodologi penelitian pada penelitian ini mencakup identifikasi persoalan, kajian pustaka, perumusan persoalan, pengumpulan informasi, pemeriksaan, penerapan sistem, pengujian, kesimpulan dan ide seperti yang dapat dilihat pada bagan *flow chart* berikut:



Gambar 1. Proses metodologi penelitian

### 3.1 Identifikasi Masalah

Identifikasi persoalan adalah proses awal pada penelitian ini untuk mendapatkan keadaan dari persoalan yang dihadapi selanjutnya penyelesaian dari persoalan tersebut. Adapun persoalan yang ditetapkan pada penelitian ini adalah bagaimana penerapan kualifikasi *data mining* dengan strategi perhitungan C4.5 untuk penentuan kelayakan penerima bantuan *family hope program* (PKH).

### 3.2 Studi Pustaka

Pada proses ini peneliti akan mengerjakan pencarian referensi, ulasan wawasan dan pengembangan dari teori yang berasal dari buku harian, naskah, buku dan internet yang akan diajukan sebagai landasan dalam perhitungan dan penulisan skripsi ini. Selanjutnya, dengan adanya kepustakaan peneliti akan mempelajari dan memahami referensi sebagai landasan dasar yang harus dipatuhi untuk menyelesaikan penelitian ini. Referensi yang dipakai pada penelitian ini adalah kajian pustaka mengenai *family hope program*, perhitungan C4.5, *decision tree*, *k-fold cross validation*, *confusion matrix* serta referensi terkait yang dapat memudahkan proses penelitian.

### 3.3 Perumusan Masalah

Perumusan persoalan adalah proses yang dikerjakan setelah mengerjakan pencarian referensi, ulasan wawasan dan pengembangan dari teori yang berasal dari buku harian, naskah, buku dan internet yang akan diajukan sebagai landasan dalam perhitungan dan penulisan. Pada proses ini maka ditetapkan suatu persoalan yang akan dijadikan unsur untuk penelitian ini yaitu penerapan perhitungan C4.5 untuk menetapkan kelayakan penerima bantuan *family hope program* (PKH).

### 3.4 Pengumpulan Data

Pada proses ini dilakukan pengumpulan informasi, informasi yang dikumpulkan antara lain adalah informasi keluarga penerima manfaat *family hope program* berupa jumlah anak usia dini, ibu hamil, siswa sekolah dasar, siswa sekolah menengah pertama, siswa sekolah menengah atas, pekerjaan, lansia, cacat berat, status tempat tinggal, jenis dinding terluas, terima raskin, fasilitas toilet, tempat pembuangan sampah. Informasi yang dipakai merupakan informasi penerima manfaat *family hope program* kabupaten bekasi tahun 2017 sd 2019. Cara pengumpulan informasi dalam penelitian ini adalah:

- a. Pertemuan langsung, peneliti melihat langsung kondisi masyarakat penerima bantuan *family hope program*.
- b. Tanya jawab, peneliti melakukan tanya jawab dengan wakil desa *family hope program* sebagai orang yang terkait langsung dalam penelitian guna memperoleh informasi tentang kriteria yang dibutuhkan untuk menetapkan keluarga penerima bantuan *family hope program* dan orang dinas sosial sebagai orang yang akan memberikan informasi yang dibutuhkan dalam penelitian.

### 3.5 Pemeriksaan

Pada penelitian ini dilakukan tiga proses pemeriksaan yaitu pemeriksaan kebutuhan informasi dan pemeriksaan strategi. Berikut adalah paparan dari setiap proses:

- a. Pemeriksaan Kebutuhan Informasi
- b. Pada proses ini dilakukan pemeriksaan pada sejumlah kriteria yang berkaitan dengan penelitian ini. Kriteria yang berkaitan dengan bantuan *family hope program* jumlah anak usia dini, ibu hamil, siswa sekolah dasar, siswa sekolah menengah pertama, siswa sekolah menengah atas, pekerjaan, lansia, cacat berat, status tempat tinggal, jenis dinding terluas, terima raskin, fasilitas toilet, tempat pembuangan sampah. Bagian dan kriteria penerima bantuan *family hope program* yang meliputi daftar kesehatan, pendidikan, dan kesejahteraan sosial. Setelah kriteria ditetapkan dan informasi yang dibutuhkan telah didapat maka dilakukan proses penentuan informasi, pra-pemrosesan, perubahan dan *data mining*. Pada proses penentuan informasi akan dilakukan pemilihan faktor serta atribut yang akan dipakai dalam penelitian ini. Pada proses pra-pemrosesan akan dilakukan proses untuk mengatasi persoalan informasi yang hilang, informasi ganda, dan memperbaiki persoalan yang terdapat pada informasi yang tidak sesuai. Pada proses perubahan adalah proses dalam proses pemrograman yaitu melakukan perubahan pada informasi yang telah dipilih sesuai kebutuhan untuk dilanjutkan pada proses *data mining*. pada proses *data mining* dilakukan penerapan perhitungan C4.5 untuk kualifikasi informasi yang telah melalui proses-proses yang telah dilakukan sebelumnya.
- c. Pemeriksaan Strategi
- d. Pada proses pemeriksaan strategi dilakukan pemeriksaan terhadap strategi yang dipakai dalam penelitian ini yaitu perhitungan C4.5. Pemeriksaan perhitungan C4.5 yaitu membuat *decision tree* berdasarkan nilai perolehan tertinggi. Proses penggunaan perhitungan C4.5 adalah sebagai berikut:
  1. Menghitung total *case* dari setiap atribut
  2. Menghitung nilai entropi dari setiap *case* dengan memakai persamaan.
  3. Menghitung nilai perolehan dari setiap atribut dengan memakai persamaan.
  4. Memilih atribut sebagai sumber berdasarkan dari atribut yang memiliki nilai perolehan yang tertinggi.
  5. Mengulangi perhitungan entropi dan perolehan hingga atribut tidak memiliki cabang *case* sehingga terbentuknya *decision tree*.

### 3.6 Penerapan

Pada proses ini dilakukan untuk pengoperasian sistem yang telah direncanakan dan dibuat sebelumnya. Kemudian dilakukan pemeriksaan sistem apakah sistem yang telah dibuat sudah sesuai

dengan apa yang telah direncanakan sebelumnya dan telah mencapai sasaran yang dituju pada penelitian ini. Berikut adalah piranti yang dipakai dalam penelitian ini:

1. Piranti keras (*Hardware*)
  - a. Prosesor: *Intel Core i3*
  - b. *RAM*: 4 GB
  - c. *Hard Disk*: 500 GB
2. Piranti Lunak (*Software*)
  - a. Sistem Operasi: *Ms. Windows 8.1 64-bit*
  - b. *Library*: *pandas, numpy, pydotplus, Ipython.display*.
  - c. Peramban: *Google Chrome*
  - d. Bahasa Program: *Python*

### 3.7 Pengujian

Pada proses ini pengujian ini dilakukan untuk melihat apakah perhitungan sudah akurat, Pengujian yang dilakukan dengan sistem yang telah dibuat memakai strategi *confusion matrix* dengan sasaran menghitung tingkat ketepatan perhitungan C4.5 dalam menetapkan kelayakan penerima bantuan *family hope program* (PKH).

### 3.8 Kesimpulan dan Ide

Pada proses ini berisi kesimpulan dan ide. Kesimpulan adalah penetapan hasil pengujian yang telah dikerjakan, apakah hasil dari penelitian yang telah dikerjakan sesuai dengan apa yang diharapkan atau tidak dan level keberhasilan pada suatu penelitian. Pada bagian ide berisi peluang pengembangan yang akan dikerjakan terhadap penelitian ini dan tentang penyempurnaan yang diharapkan dapat menghasilkan penelitian yang lebih bagus sehingga menjadi referensi yang bagus untuk penelitian selanjutnya.

## 4. HASIL DAN PEMBAHASAN

### 4.1 Penerapan

Penerapan adalah proses yang dikerjakan berdasarkan proses pemeriksaan proses sebelumnya. Pada proses ini dilakukan untuk pengoperasian sistem yang telah direncanakan dan dibuat sebelumnya. Kemudian dilakukan apakah sistem yang telah dibuat telah sesuai dengan apa yang telah direncanakan sebelumnya dan telah mencapai sasaran yang diharapkan dalam penelitian ini.

#### a. Lingkungan Penerapan

Lingkungan penerapan adalah lingkungan sistem yang telah direncanakan dan dibuat. Penerapan perhitungan dan sistem ini terbagi dalam dua bidang yaitu lingkungan piranti keras dan piranti lunak. Berikut adalah perincian lingkungan tersebut.

1. Piranti keras (*Hardware*)
  - i. Prosesor: *Intel Core i3*
  - ii. *RAM*: 4 GB
  - iii. *Hard Disk*: 500 GB
2. Piranti Lunak (*Software*)
  - i. Sistem Operasi: *Ms. Windows 8.1 64-bit*
  - ii. *Library*: *pandas, numpy, sklearn, pydot, pydotplus, Ipython.display*.
  - iii. Peramban: *Google Chrome*
  - iv. Bahasa Program: *Python*

#### b. Masukan *Library* dan alat

Sebelum memasuki proses masukan informasi dan pra-prosesing, masukan *library* dilakukan terlebih dahulu. *Library* yang dipakai adalah *pandas, numpy, sklearn, pydot, pydotplus, dan Ipython.display*. Masukan *library* dan alat dapat dilihat pada Gambar 2 berikut:

```
[61] from sklearn.tree import DecisionTreeClassifier
      from sklearn.model_selection import train_test_split
      from sklearn.metrics import plot_confusion_matrix
      from sklearn.metrics import confusion_matrix
      from sklearn.tree import export_graphviz
      from IPython.display import Image
      from pydot import graph_from_dot_data
      from sklearn.tree import plot_tree
      from sklearn import tree
      import pydotplus
      import pandas as pd
      import numpy as np
```

Gambar 2. Masukkan *library*

c. Masukan Informasi

Informasi yang ditampilkan adalah informasi dengan atribut anak usia dini, ibu hamil, siswa sekolah dasar, siswa sekolah menengah pertama, siswa sekolah menengah atas, pekerjaan, lansia, cacat berat, status tempat tinggal, jenis dinding terluas, terima raskin, fasilitas toilet, tempat pembuangan sampah dan kelas. Kemudian diberi tanda atribut sesuai dengan atribut yang telah ditetapkan. Proses masukan informasi dapat dilihat pada gambar berikut:

|   | balita | hamil | lansia | disabilitas | anak sd | anak smp | anak sma | pekerjaan     | status tempat tinggal | dinding terluas | raskin                  | jamban       | pembuangan akhir | kelas |
|---|--------|-------|--------|-------------|---------|----------|----------|---------------|-----------------------|-----------------|-------------------------|--------------|------------------|-------|
| 0 | 1      | TIDAK | 0      | 0           | 0       | 1        | 0        | BURUHKARYAWAN | MLIK SENDIRI          | TEMBOK          | TIDAK TERIMA            | MLIK SENDIRI | TANGKUSPAL       | LAYAK |
| 1 | 0      | TIDAK | 0      | 0           | 1       | 1        | 0        | TIDAK BEKERJA | MLIK SENDIRI          | TEMBOK          | TERIMA KURANG DARI 15KG | MLIK SENDIRI | TANGKUSPAL       | LAYAK |
| 2 | 0      | TIDAK | 0      | 0           | 2       | 1        | 0        | TIDAK BEKERJA | MLIK SENDIRI          | TEMBOK          | TERIMA 15KG             | MLIK SENDIRI | TANGKUSPAL       | LAYAK |
| 3 | 0      | TIDAK | 0      | 0           | 0       | 1        | 2        | TIDAK BEKERJA | MLIK SENDIRI          | TEMBOK          | TERIMA KURANG DARI 15KG | MLIK SENDIRI | TANGKUSPAL       | LAYAK |
| 4 | 0      | TIDAK | 0      | 0           | 1       | 1        | 0        | PETANI        | MLIK SENDIRI          | TEMBOK          | TIDAK TERIMA            | MLIK SENDIRI | TANGKUSPAL       | LAYAK |

Gambar 3. Masukan informasi

d. Transformasi

Setelah memasukkan informasi dengan struktur informasi berdasarkan atribut yang telah ditetapkan. Informasi yang ditampilkan adalah informasi dengan atribut anak usia dini, ibu hamil, siswa sekolah dasar, siswa sekolah menengah pertama, siswa sekolah menengah atas, pekerjaan, lansia, cacat berat, status tempat tinggal, jenis dinding terluas, terima raskin, fasilitas toilet, tempat pembuangan sampah. Selanjutnya dilakukan transformasi informasi agar proses kualifikasi dapat terbaca oleh sistem. Proses transformasi dapat dilihat pada gambar berikut.

|   | balita | lansia | disabilitas | anak sd | anak smp | anak sma | hamil_TIDAK | hamil_YA | pekerjaan_BURUHKARYAWAN | pekerjaan_PEDAGANG | pekerjaan_PETANI | pekerjaan_TIDAK BEKERJA | tinggal_1 | 5 |
|---|--------|--------|-------------|---------|----------|----------|-------------|----------|-------------------------|--------------------|------------------|-------------------------|-----------|---|
| 0 | 1      | 0      | 0           | 0       | 1        | 0        | 1           | 0        | 1                       | 0                  | 0                | 0                       | 0         | 0 |
| 1 | 0      | 0      | 0           | 1       | 1        | 0        | 1           | 0        | 0                       | 0                  | 0                | 0                       | 1         | 1 |
| 2 | 0      | 0      | 0           | 2       | 1        | 0        | 1           | 0        | 0                       | 0                  | 0                | 0                       | 1         | 1 |
| 3 | 0      | 0      | 0           | 0       | 1        | 2        | 1           | 0        | 0                       | 0                  | 0                | 0                       | 1         | 1 |
| 4 | 0      | 0      | 0           | 1       | 1        | 0        | 1           | 0        | 0                       | 0                  | 1                | 0                       | 1         | 0 |

Gambar 4. Transformasi

e. Penerapan Proses Pra-Prosesing

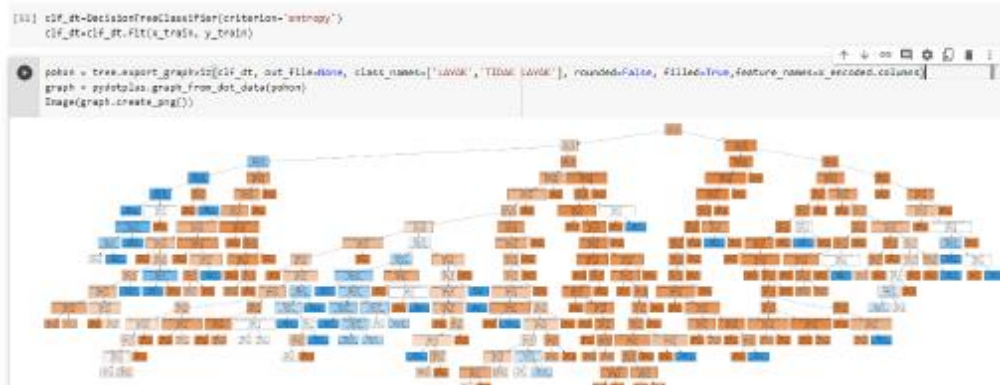
Setelah proses masukan informasi berhasil selanjutnya dilakukan proses pra-prosesing. Proses ini dilakukan pengambilan informasi latih dan informasi uji untuk dilakukan pra-prosesing. Proses pra-prosesing dapat dilihat pada gambar berikut.

```
x_train, x_test, y_train, y_test = train_test_split(x_encoded, y,)
```

Gambar 5. Penerapan pra-prosesing

f. Penerapan Kualifikasi

Penerapan kualifikasi menghasilkan informasi baru yang telah melalui proses pra-prosesing. Informasi tersebut akan diproses dengan perhitungan C4.5 untuk dikualifikasi sesuai dengan kualifikasi yang ditetapkan hingga terbuat *decision tree*. Proses kualifikasi dapat dilihat pada Gambar 6 berikut.



Gambar 6. Penerapan kualifikasi

Setelah strategi kualifikasi perhitungan C4.5 dilakukan maka didapat hasil *decision tree*. Kemudian akan dilakukan transformasi *decision tree* menjadi rupa aturan-aturan. Berikut aturan-aturan dari *decision tree* yang telah dibuat.

- a. R1 jika (siswa sekolah menengah pertama = ' $\leq 0$ ') ^ (siswa sekolah dasar = ' $\leq 0$ ') ^ (siswa sekolah menengah atas = ' $\leq 0$ ') ^ (cacat berat = ' $\leq 0$ ') ^ (anak usia dini = ' $\leq 0$ ') maka kelas = Tidak Layak
- b. R2 jika (siswa sekolah menengah pertama = ' $\leq 0$ ') ^ (siswa sekolah dasar = ' $\leq 0$ ') ^ (siswa sekolah menengah atas = ' $\leq 0$ ') ^ (cacat berat = ' $\leq 0$ ') ^ (anak usia dini = ' $\leq 2$ ') Maka kelas = Layak
- c. R3 jika (siswa sekolah menengah pertama = ' $\leq 0$ ') ^ (siswa sekolah dasar = ' $\leq 0$ ') ^ (siswa sekolah menengah atas = ' $\leq 0$ ') ^ (cacat berat = ' $\leq 0$ ') ^ (anak usia dini = ' $> 0$ ') ^ (jenis dinding terluas = 'tembok') ^ (pekerjaan = 'tidak bekerja') Maka kelas = Tidak Layak
- d. R4 jika (siswa sekolah menengah pertama = ' $\leq 0$ ') ^ (siswa sekolah dasar = ' $\leq 0$ ') ^ (siswa sekolah menengah atas = ' $\leq 2$ ') ^ (pekerjaan = 'petani') ^ (tempat pembuangan sampah = 'sungai/danau') ^ (terima raskin = 'terima 15kg') ^ (pekerjaan = 'pedagang') Maka kelas = Layak
- e. R5 jika (siswa sekolah menengah pertama = ' $\leq 0$ ') ^ (siswa sekolah dasar = ' $\leq 2$ ') ^ (siswa sekolah menengah atas = ' $\leq 2$ ') ^ (pekerjaan = 'petani') Maka kelas = Layak

**4.2 Pengujian**

Pengujian dilakukan untuk melihat ketepatan kualifikasi penerima bantuan *family hope program* (PKH). Mengukur ketepatan dilakukan dengan memakai *confusion matrix* selanjutnya perbandingan hasil kualifikasi yang memakai perhitungan C4.5 yang sudah dilakukan secara manual sebelumnya pada setiap informasi pengujian. Hasil perhitungan ketepatan menjadi standar keberhasilan kualifikasi penggunaan strategi tersebut. Adapun tingkat kualifikasi yaitu:

- 1. 90-100 = Klasifikasi bagus sekali
- 2. 80-90 = Klasifikasi bagus
- 3. 70-80 = Klasifikasi cukup
- 4. 60-70 = Klasifikasi buruk
- 5. 0-50 = Kegagalan

a. Pengujian dengan perbandingan 90%:10%

Pengujian dilakukan memakai informasi penerima bantuan *family hope program* (PKH) dengan atribut anak usia dini, ibu hamil, siswa sekolah dasar, siswa sekolah menengah pertama, siswa sekolah menengah atas, pekerjaan, lansia, cacat berat, status tempat tinggal, jenis dinding terluas, terima raskin, fasilitas toilet, tempat pembuangan sampah. Berikut perhitungan ketepatan perhitungan C4.5 dengan perbandingan informasi latih 90% sejumlah 1.117 catatan dan informasi uji 10% sejumlah 125 catatan.

Tabel 1. *Confusion matrix*

| Kelas        | Kelas Perkiraan |       |             |
|--------------|-----------------|-------|-------------|
|              | Kelas           | Layak | Tidak layak |
| Sesungguhnya | Layak           | 180   | 17          |
|              | Tidak layak     | 21    | 61          |



Berdasarkan Tabel 1 *Confusion matrix* diatas maka didapat hasil sebagai berikut:

$$akurasi = \frac{180 + 61}{180 + 17 + 21 + 61} \times 100 = \frac{241}{279} \times 100 = 86,37 \quad (1)$$

Berdasarkan perhitungan yang telah dilakukan, maka didapat ketepatan penggunaan perhitungan C4.5 senilai 86,37%.

b. Pengujian dengan perbandingan 80%:20%

Berikut perhitungan ketepatan perhitungan C4.5 dengan perbandingan informasi latih 80% sejumlah 993 catatan dan informasi uji 20% sejumlah 249 catatan.

Tabel 2. *Confusion matrix*

| Kelas<br>Sesungguhnya | Kelas Perkiraan |       |             |
|-----------------------|-----------------|-------|-------------|
|                       | Kelas           | Layak | Tidak layak |
|                       | Layak           | 170   | 10          |
|                       | Tidak layak     | 29    | 47          |

Berdasarkan Tabel 2. *Confusion matrix* diatas maka didapat hasil sebagai berikut:

$$akurasi = \frac{170 + 47}{170 + 10 + 29 + 47} \times 100 = \frac{217}{256} \times 100 = 84,76 \quad (2)$$

Berdasarkan perhitungan yang telah dilakukan, maka didapat ketepatan penggunaan perhitungan C4.5 senilai 84,76%.

c. Pengujian dengan perbandingan 70%:30%

Berikut perhitungan ketepatan perhitungan C4.5 dengan perbandingan informasi latih 70% sejumlah 865 catatan dan informasi uji 30% sejumlah 377 catatan.

Tabel 3. *Confusion matrix*

| Kelas<br>Sesungguhnya | Kelas Perkiraan |       |             |
|-----------------------|-----------------|-------|-------------|
|                       | Kelas           | Layak | Tidak layak |
|                       | Layak           | 140   | 13          |
|                       | Tidak layak     | 23    | 49          |

Berdasarkan Tabel 3. *Confusion matrix* diatas maka didapat hasil sebagai berikut:

$$akurasi = \frac{140 + 49}{140 + 13 + 23 + 49} \times 100 = \frac{189}{225} \times 100 = 84,00 \quad (3)$$

Berdasarkan perhitungan yang telah dikerjakan, maka didapat ketepatan penggunaan perhitungan C4.5 sejumlah 84%.

Berdasarkan pengujian ketepatan didapat kesimpulan yang dikerjakan dengan memakai informasi penerima bantuan *family hope program*. Hasil perhitungan ketepatan perhitungan C4.5 dengan perbandingan informasi latih 90% sejumlah 1.117 catatan dan informasi uji 10% sejumlah 125 catatan, maka didapat ketepatan sejumlah 86,37%. Perbandingan informasi latih 80% sejumlah 993 catatan dan informasi uji 20% sejumlah 249 catatan. Maka didapat ketepatan penggunaan perhitungan C4.5 sejumlah 84,76%. Perbandingan informasi latih 70% sejumlah 865 catatan dan informasi uji 30% sejumlah 377 catatan. Maka didapat ketepatan penggunaan perhitungan C4.5 sejumlah 84%.

## 5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dikerjakan, maka didapat suatu kesimpulan dari penerapan perhitungan C4.5 untuk mengkualifikasi penerima bantuan *family hope program* (PKH) adalah sebagai berikut:

- Penerapan perhitungan C4.5 untuk mengkualifikasi penerima bantuan *family hope program* (PKH) telah sukses dilakukan.

- b. *Decision tree* yang dihasilkan dengan perhitungan C4.5 membuat atribut siswa sekolah menengah pertama memiliki nilai perolehan tertinggi sebagai referensi penentuan kelayakan penerima bantuan.
- c. Pengujian perhitungan C4.5 memakai *confusion matrix* mendapat ketepatan terbaik dengan perbandingan 90% informasi latih dan 10% informasi uji sejumlah 86,37%.
- d. Berdasarkan hasil pengujian tersebut perhitungan C4.5 termasuk pada tingkat Klasifikasi Bagus.

Adapun saran untuk perluasan penelitian ini di waktu yang akan datang sejenis dengan penelitian ini yaitu:

- a. Memakai lebih banyak atribut yang termasuk dalam kriteria penerima bantuan *family hope program* agar memperoleh ketepatan yang lebih tinggi lagi.
- b. Menambah total informasi menjadi lebih banyak sehingga dapat mempengaruhi tingkat ketepatan pada proses kualifikasi.
- c. Berdasarkan penelitian ini dibuatnya perangkat lunak berdasar web atau android yang dapat menetapkan kelayakan penerima bantuan *family hope program*.
- d. Memakai strategi kualifikasi *data mining* lainnya sebagai unsur perbandingan dalam penentuan kelayakan penerima bantuan *family hope program*.

## REFERENSI

- [1] Hariati, M. Wati and B. Cahyono, "Penerapan Algoritma C4.5 pada Penentuan Penerima Program Bantuan Pemerintah Daerah di Kabupaten Kutai Kartanegara," *JURTI*, pp. 1-9, 2018.
- [2] F. Astuti, N. R. A. I. Purnamasari, I. Ali and D. A. Kurnia, "Klasifikasi Penerima Bantuan Sosial Menggunakan Algoritma C.45 Kantor Kuwu Desa Bangodua," *Jurnal Data Science & Informatika (JDSI)*, pp. 1-8, 2021.
- [3] N. I. Nella, N. Y. Setiawan and D. E. Ratnawati, "Klasifikasi Penerima Bantuan Program Keluarga Harapan menggunakan Algoritme Decision Tree C4.5 (Studi Kasus: Desa Mlirip Kabupaten Mojokerto)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, pp. 1-8, 2022.
- [4] Muafi, I. Ardiansyah and H. Basori, "Klasifikasi Pemberian Bantuan Kepada Masyarakat Terdampak Covid-19 Dengan Menggunakan Al-Goritma C4.5 Didesa Demung Kecamatan Besuki Kabupaten Situbondo," *Jurnal Kecerdasan Buatan, Komputasi dan Teknologi Informasi*, pp. 1-9, 2021.
- [5] S. Badriah, M. F. E. Nugroho, N. Sanjaya, I. Rismawati, B. N. Sari and C. Rozikin, "Klasifikasi Algoritma C4.5 dalam Menentukan Penerima Bantuan Covid-19 (Studi Kasus: Desa di Karawang)," *JIP (Jurnal Informatika Polinema)*, pp. 1-6, 2021.
- [6] L. Qadrini, A. Seppewali and A. Aina, "Decision Tree Dan Adaboost Pada Klasifikasi Penerima Program Bantuan Sosial," *Jurnal Inovasi Penelitian*, pp. 1-8, 2021.
- [7] Y. Alkhalifi, A. Zumarniansyah, R. Ardianto, N. Hardi and A. E. Augustia, "Comparison Of Naive Bayes Algorithm And C.45 Algorithm In Classification Of Poor Communities Receiving Non Cash Food Assistance In Wanasari Village Karawang Regency," *Jurnal TECHNO Nusa Mandiri*, pp. 1-6, 2020.
- [8] R. N. Adiha, S. R. Andani and W. Saputra, "Application Of Data Mining In Determining Social Assistance Recipients With C4.5 Algorithm," *Journal of Artificial Intelligence and Engineering Applications*, pp. 1-10, 2022.
- [9] R. W. Sinaga, R. Winanjaya and Susianti, "Analisis Data Mining Menentukan Penerima Bantuan Langsung Tunai pada Desa Pamatang Purba dengan Algoritma C 4.5," *BRAHMANA: Jurnal Penerapan Kecerdasan Buatan*, pp. 1-9, 2021.
- [10] M. A. Tanjung, Poningsih and H. Qurniawan, "Analisa Kelayakan Penerima Program Keluarga Harapan (PKH) Menggunakan Algoritma C4.5," *Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)*, pp. 1-9, 2021.