

Analisis Kehadiran Pemilih di Tempat Pemungutan Suara Pada Pemilihan Presiden dengan Menggunakan Algoritma C4.5 (Studi Kasus di Desa Wonokasian Sidoarjo)

Wahidiyah Kurniawati, Arif Senja Fitrani*, Sukma Aji, Suprianto

Universitas Muhammadiyah Sidoarjo

Abstrak: Pemilihan umum diadakan setiap lima tahun dimana seluruh warga negara Indonesia berpartisipasi sebagai bagian dari proses demokrasi. Partisipasi warga negara penting untuk menggunakan hak pilihnya dalam memilih pemimpin saat pelaksanaan pemilu. Meningkatnya jumlah pemilih yang mengikuti pemilu dapat dijadikan sebagai patokan keberhasilan pemilu. Peneliti akan menguji klasifikasi berdasarkan data penduduk pemilu yang diperoleh dari Desa Wonokasian tahun 2019 menggunakan metode data mining C4.5. Variabel yang digunakan adalah Jumlah KK, Asal Penduduk, Kategori, Status, RT, TPS, Lokasi, Hadir (keluarga) dan Hadir. Hasil dari penelitian berikut menggunakan weka sehingga memudahkan bagi masyarakat desa mengetahui data pemilu antara yang hadir maupun tidak hadir. Dari hasil klasifikasi pada metode C4.5 untuk mengklasifikasi status kehadiran dari dataset 1.700, diperoleh tingkat kebenaran sebesar 91,94% dan tingkat kesalahan sebesar 8,05%.

Kata kunci: Partisipasi, Pemilu, C4.5, Klasifikasi, Data Mining.

DOI:

<https://doi.org/10.47134/jte.v1i1.2475>

*Correspondence: Arif Senja Fitrani

Email: asfjim@umsida.ac.id

Received: 20-02-2024

Accepted: 15-03-2024

Published: 30-04-2024



Copyright: © 2024 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Abstract: General elections are held five years in which Indonesian citizens participate as part of the democratic. Citizen participation is important to exercise right to vote in selecting leaders during elections. The increasing number of voters participating in the election can be used as a benchmark for the success of the election. Researchers will test the classification based on election population data obtained from Wonokasian Village in 2019 using the C4.5 data mining method. The variables used are the number of households, population origin, category, status, RT, TPS, location, attendance (family) and presence. The results of the study use Weka to make it easier for village people to find out election data between those who are present and those who are not present. From the classification results in the C4.5 method for classifying the presence status of the 1,700 dataset, truth level of 91.94% and error rate of 8.05% were obtained.

Keywords: Participation, Election, C4.5, Classification, Data Mining.

Pendahuluan

Pemilihan umum ialah tahapan melaksanakan kedaulatan rakyat yang dilakukan secara langsung, umum, bebas, rahasia, jujur dan adil dalam negara kesatuan Indonesia menurut Pancasila dan Undang-Undang Dasar Negara Republik Indonesia tahun 1945. Dengan cara pemilihan umum, masyarakat dapat secara langsung menentukan kebijakan publik dengan mendukung calon yang memiliki tujuan dan cita-cita yang sama. Singkatnya, tujuan pemilu yaitu untuk memilih pejabat pemerintahan baik secara legislatif maupun eksekutif (Subiyanto, 2020).

Pemilihan umum diadakan setiap lima tahun dimana seluruh warga negara Indonesia berpartisipasi sebagai bagian dari proses demokrasi. Partisipasi warga negara penting untuk menggunakan hak pilihnya dalam memilih pemimpin saat pelaksanaan pemilu. Meningkatnya jumlah pemilih yang mengikuti pemilu dapat dijadikan sebagai patokan keberhasilan pemilu. Kebijakan voting pada pemilu dapat ditentukan dengan beberapa metode untuk mendapatkan nilai dari sekumpulan ide dan data dengan menggunakan ilmu data mining.

Ciri negara demokratis ialah negara yang melibatkan masyarakat dalam perencanaan maupun pelaksanaan pemilihan umum. Partisipasi masyarakat merupakan aspek penting dalam tatanan negara demokrasi. Setiap masyarakat memiliki preferensi dan kepentingan masing-masing untuk menentukan pilihan mereka dalam pemilu (Liando, 2016). Klasifikasi adalah sebuah pengklasifikasi yang dibuat dari sekumpulan data latih dengan kelas yang telah ditentukan (Amalia, 2013; F & others, 2019; Sindunata, 2014). Algoritma C4.5 ialah algoritma yang digunakan untuk mengelompokkan data dengan karakteristik kategori dan atribut numerik (Chen, 2023; Rahardja, 2022; Shamrat, 2022; Shanthi, 2022; Wang, 2022). Algoritma C4.5 disebut sebagai pohon keputusan (decision tree) yaitu metode klasifikasi yang menggunakan representasi struktur pohon, setiap node merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas (Bhardwaj, 2022; Bustillo, 2022; Mebawondu, 2022; Mohammad, 2022; Tempola, 2022).

Metode

A. Klasifikasi

Klasifikasi ialah proses penemuan model yang menggambarkan dan membedakan kelas data yang bertujuan agar digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Hendrian, 2018). Klasifikasi data terdiri dari 2 tahap. Pertama yaitu fase training, yang mana algoritma klasifikasi dibuat untuk menganalisa data training kemudian direpresentasikan dalam bentuk rule klasifikasi. Tahap kedua yaitu klasifikasi, yang mana data testing digunakan untuk memperkirakan akurasi dari rule klasifikasi (Han & Kamber, 2006)

B. Algoritma C4.5

Algoritma C4.5 adalah algoritma klasifikasi yang menghasilkan pohon keputusan yang mudah diinterpretasikan, serta memiliki tingkat akurasi yang dapat diterima (Kamagi, 2014). Dalam algoritma C4.5 terdapat perhitungan Entrophy dan Gain sebagai dasar penentuan nilai pada node pohon keputusan (Wanti, 2020):

1. Entrophy

Menurut Suyanto (2017), secara istilah Entrophy yaitu suatu keberagaman. Pada data mining, Entrophy memiliki definisi sebagai parameter guna mengukur keberagaman pada suatu himpunan data. Rumus Entrophy:

$$Entropy = \sum_{i=1}^n -p_i * \log_2 p_i \quad (1)$$

Keterangan:

N ialah jumlah nilai yang terdapat pada atribut target (jumlah kelas). Sedangkan Pi menyatakan rasio antara jumlah sampel di kelas i dengan jumlah sampel pada himpunan data.

2. Gain

Secara istilah Gain yaitu proses perolehan informasi. Pada data mining Gain didefinisikan sebagai ukuran efektivitas suatu atribut dalam mengklasifikasi data. Secara matematis, gain dirumuskan sebagai berikut:

$$Gain(S, A) = Entropy(s) - \sum \frac{|S_i|}{|s|} Entropy(S_i) \quad (2)$$

Keterangan:

Atribut nilai gain yang paling tertinggi dipilih sebagai atribut tes untuk simpul.

C. Weka

“WEKA” merupakan singkatan dari “Waikato Environment for Knowledge Analysis”, dibuat di Universitas Waikato, New Zealand untuk penelitian, pendidikan dan berbagai aplikasi. WEKA mengandung tools untuk preprocessing data, klasifikasi, regresi, clustering, aturan asosiasi, dan visualisasi (Purnamasari, 2013). Contoh penggunaan WEKA ialah menerapkan sebuah metode pembelajaran ke dataset dan menganalisis hasil guna memperoleh informasi tentang data maupun menerapkan beberapa metode serta membandingkan performanya untuk dipilih (Aswendy, 2016).

D. Flowchart

Flowchart merupakan teknik analitis yang berfungsi untuk menjelaskan aspek-aspek sistem informasi dengan jelas, tepat dan logis (Liksha, 2018). Secara umum, flowchart adalah dari tahapan dan urutan prosedur sebuah program secara grafik. Flowchart digunakan oleh analis dan programmer dalam menyelesaikan masalah dengan membagi menjadi segmen yang lebih kecil serta digunakan untuk menganalisis alternatif lain dalam pengoperasian (Budiman, 2021).

E. Lokasi Dan Waktu Penelitian

Penelitian ini dilaksanakan pada 15 Februari 2022 dengan melakukan pengumpulan data pada desa Wonokasian di Kabupaten Sidoarjo. Penelitian tersebut dilakukan setelah mendapat persetujuan dari dosen pembimbing. Data yang digunakan pada penelitian ialah data pemilu di desa Wonokasian Kabupaten Sidoarjo.

F. Teknik Pengumpulan Data

Guna memperoleh data yang dapat menunjang penelitian, maka peneliti melakukan pencarian dan pengumpulan data. Beberapa metode yang dibutuhkan dalam pengumpulan data ialah sebagai berikut (Susanto, 2014):

1. Studi Pustaka (Literatur)

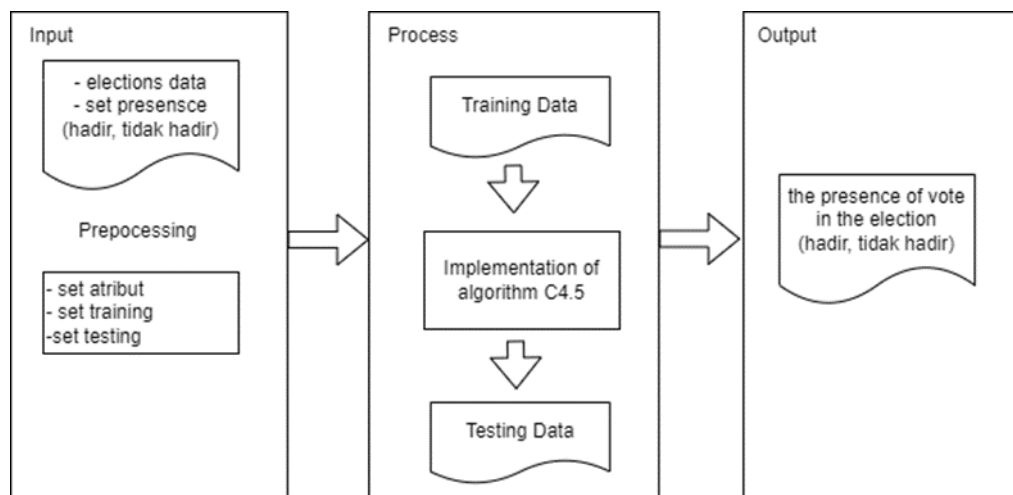
Dalam metode ini menggunakan literatur yang digunakan dalam pencarian data yang berhubungan dengan data pemilu pada desa Wonokasian Kabupaten Sidoarjo. Metode pencarian data ini diperoleh melalui jurnal, buku dan browsing di internet. Literatur yang didapatkan dalam memperoleh informasi ialah berupa teori dasar yang berkaitan dengan penelitian ini.

2. Observasi

Metode pengumpulan data berupa observasi yang dilaksanakan dengan pengamatan secara langsung di desa Wonokasian Kabupaten Sidoarjo berdasarkan objek penelitian dengan pengambilan data yang akan digunakan. Sehingga akan didapatkan data yang cukup relevan dan informasi yang lebih jelas. Data yang didapat oleh peneliti ialah data pemilu.

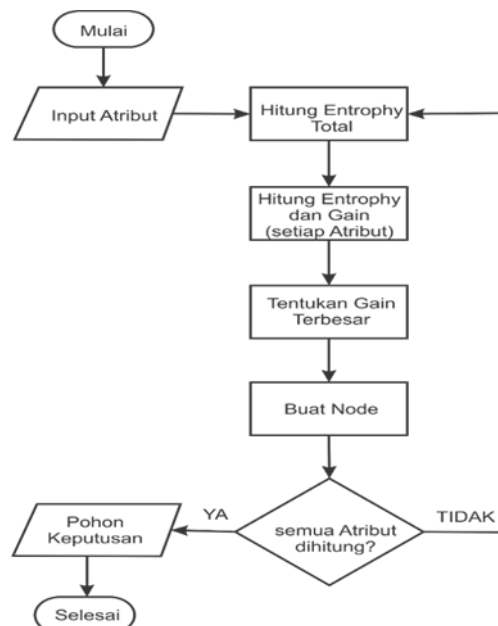
G. Diagram Klasifikasi

Perancangan diagram klasifikasi menggunakan algoritma C4.5 ialah sebagai berikut:



Gambar 1. Diagram Klasifikasi

H.Flowchart

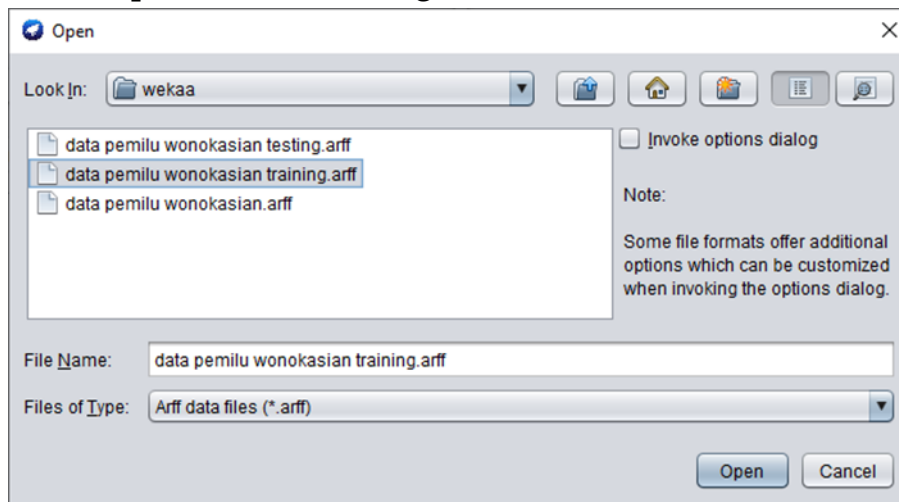


Gambar 2. Flowchart

Hasil dan Pembahasan

Data yang digunakan sebagai penelitian ialah data pemilu di desa Wonokasian Kabupaten Sidoarjo. Digunakan untuk memprediksi kehadiran saat pemilu. Data yang diperoleh ialah sebanyak 4.249 data pada tahun 2019. Penggunaan dataset sejumlah 4.249 dataset diambil 60% yaitu sebanyak 2.549 instances untuk dijadikan sebagai data training dan 40% yaitu sebanyak 1.700 instances untuk dijadikan sebagai data testing. Data ini akan diolah menggunakan metode C4.5 untuk mengetahui kehadiran masyarakat pada pemilu.

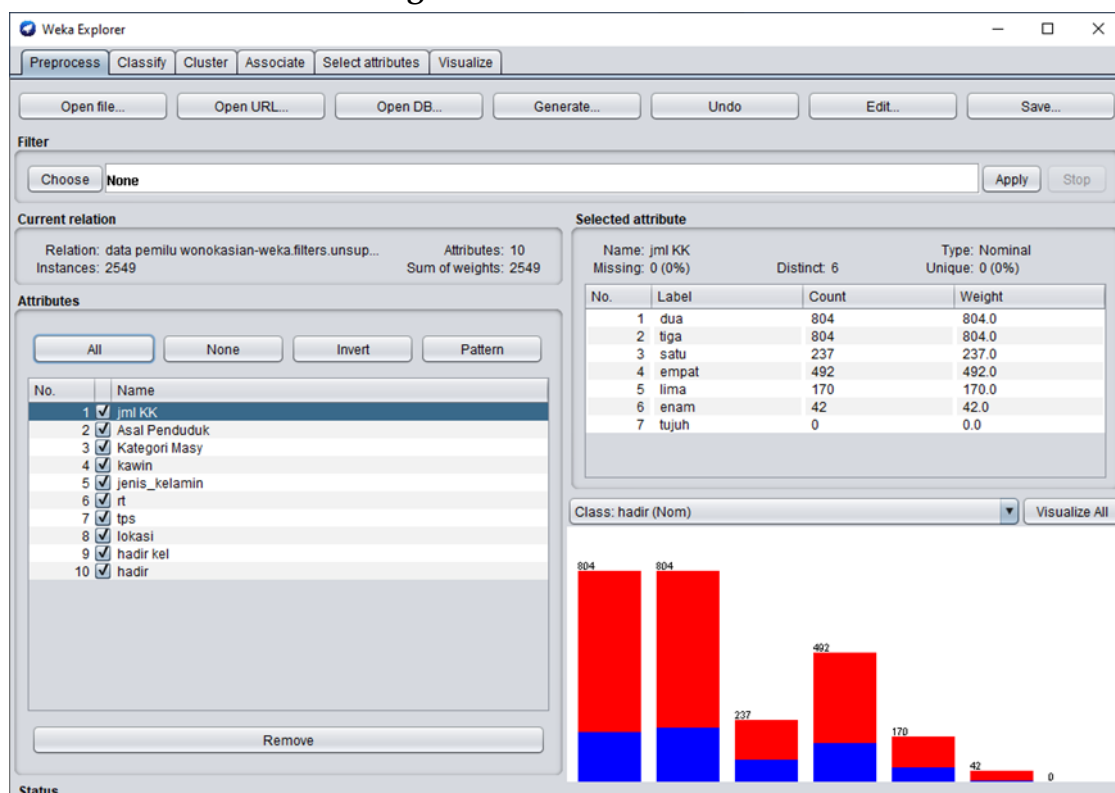
A. Halaman Menu Explorer Data Training



Gambar 3. Menu Explorer Data Training

Peneliti akan menguji data training menggunakan algoritma C4.5 pada weka dengan format arff.

B. Halaman Atribut Data Training

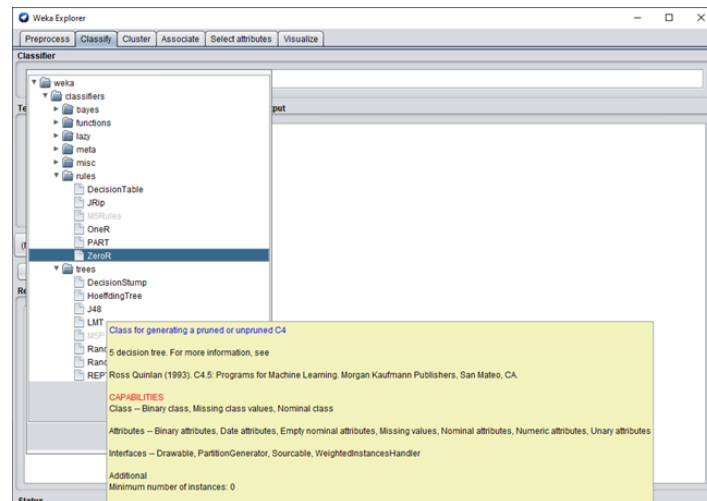


Gambar 4. Halaman Atribut Data Training

Peneliti menampilkan atribut yang terdapat pada data training. Terdapat 10 atribut yang digunakan pada dataset data training, diantaranya adalah Jumlah KK, Asal

Penduduk, Kategori Masyarakat, Kawin, Jenis Kelamin, RT, TPS, Lokasi, Hadir Kel, Hadir. Class yang digunakan untuk pengujian yaitu Jumlah kehadiran setiap KK dengan nilai seperti yang ditampilkan di gambar 4 dari 2.549 data training.

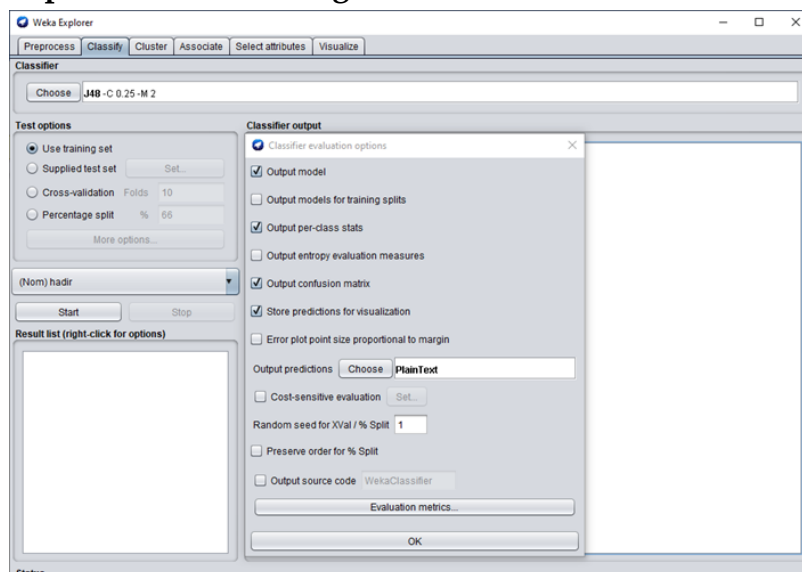
C. Halaman Menu Klasifikasi



Gambar 5. Halaman Menu Klasifikasi

Peneliti memilih metode klasifikasi yang akan digunakan untuk pengujian. Pengujian klasifikasi data pemilu ini akan diuji menggunakan algoritma C4.5. Tahap berikutnya ialah memasukkan algoritma C4.5 yang akan digunakan pada bagian Filter Choose pilih folder trees lalu pilih J48. Berikut ini ialah rincian dari hasil klasifikasi dari 2.549 data training menggunakan Test Option berupa Use Training Set.

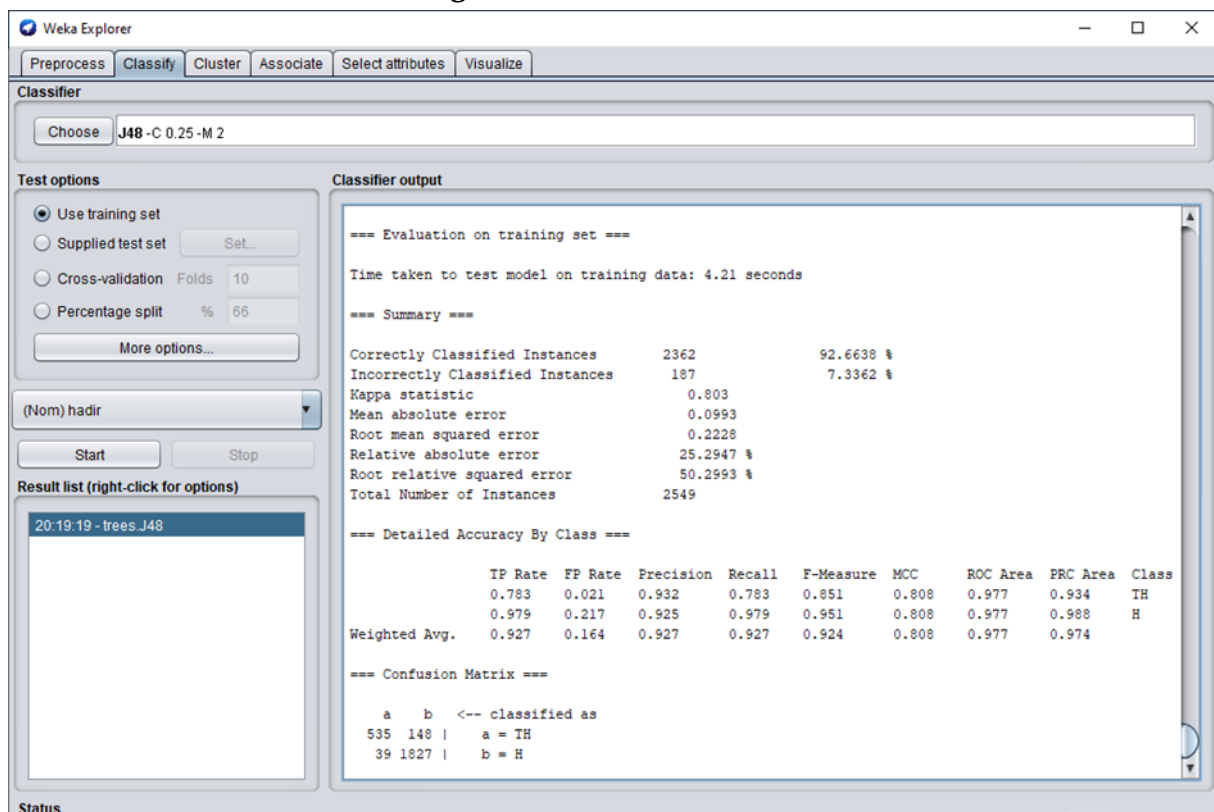
D. Halaman Test Options Data Training



Gambar 6. Halaman Test Options Data Training

Peneliti akan memunculkan basis aturan yang digunakan pada data testing. Pada halaman test options, klik use training set lalu klik more options. Pada output predictions pilih folder prediction klik choose lalu pilih PlainText, setelah itu klik Ok. Kemudian akan muncul basis aturan. Klik Start untuk melakukan klasifikasi pada data training.

E. Hasil Klasifikasi Data Training



Gambar 7. Hasil Klasifikasi Data Training

Berikut ini ialah hasil evaluasi data training, yaitu sebagai berikut: a). Skema berikut diolah menggunakan weka, relasinya data pemilu wonokasian-weka.filters.unsupervised.instance.RemovePercentage-P40.0, jumlah data sebanyak 2.549, jumlah atribut sebanyak 10. b). Hasil menunjukkan basis aturan pada pohon keputusan dengan jumlah data training 2.549 data dan daun (*number of leaves*) sebanyak 85 dan ukuran pohon (*size of the tree*) sebanyak 98. Node diawali atribut (hadir kel) dan diteruskan dengan pencabangan lengan sampai perkabelan daun pada akhir keputusan. c) waktu yang digunakan untuk pengujian data training yaitu 3,89 detik. d) correctly classified instances adalah jumlah hasil klasifikasi yang benar sebanyak 2.362 dengan perolehan presentase sebesar 92,66%. e) incorrectly classified instances adalah jumlah hasil klasifikasi yang salah sebanyak 187 dengan perolehan persentase sebesar 7,33%.

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
 Relation: data pemilu wonokasian-weka.filters.unsupervised.instance.RemovePercentage-P40.0
 Instances: 2549
 Attributes: 10
 jml_KK
 Asal Penduduk
 Kategori Masy
 kawin
 jenis_kelamin
 rt
 tps
 lokasi
 hadir_kel
 hadir
 Test mode: evaluate on training data

=== Classifier model (full training set) ===

J48 pruned tree

```

hadir_kel = TSH
| kawin = S: H (357.0/109.0)
| kawin = B
| | tps = TPS-1: TH (0.0)
| | tps = TPS-2: TH (0.0)
| | tps = TPS-3: TH (0.0)
| | tps = TPS-4: TH (0.0)
| | tps = TPS-5: TH (0.0)
| | tps = TPS-6: TH (0.0)
| | tps = TPS-7
| | | jenis_kelamin = P: H (8.0/2.0)
| | | jenis_kelamin = L: TH (20.0/6.0)
| | tps = TPS-8
| | | Asal_Penduduk = SIDOARJO: TH (13.0/4.0)
| | | Asal_Penduduk = LUAR: H (6.0/2.0)
| | | Asal_Penduduk = LUAR_PROV: TH (0.0)
| | tps = TPS-9
| | | Kategori_Masy = Milenial: H (16.0/4.0)
| | tps = TPS-16: H (23.0/11.0)
| kawin = P: H (25.0/11.0)
hadir_kel = THS: TH (425.0)
hadir_kel = HS: H (1497.0)
  
```

Number of Leaves : 85

Size of the tree : 98

```

=== Evaluation on training set ===

Time taken to test model on training data: 3.89 seconds

=== Summary ===

Correctly Classified Instances   2362      92.6638 %
Incorrectly Classified Instances  187       7.3362 %
Kappa statistic                  0.803
Mean absolute error              0.0993
Root mean squared error          0.2228
Relative absolute error          25.2947 %
Root relative squared error      50.2993 %
Total Number of Instances       2549

=== Detailed Accuracy By Class ===

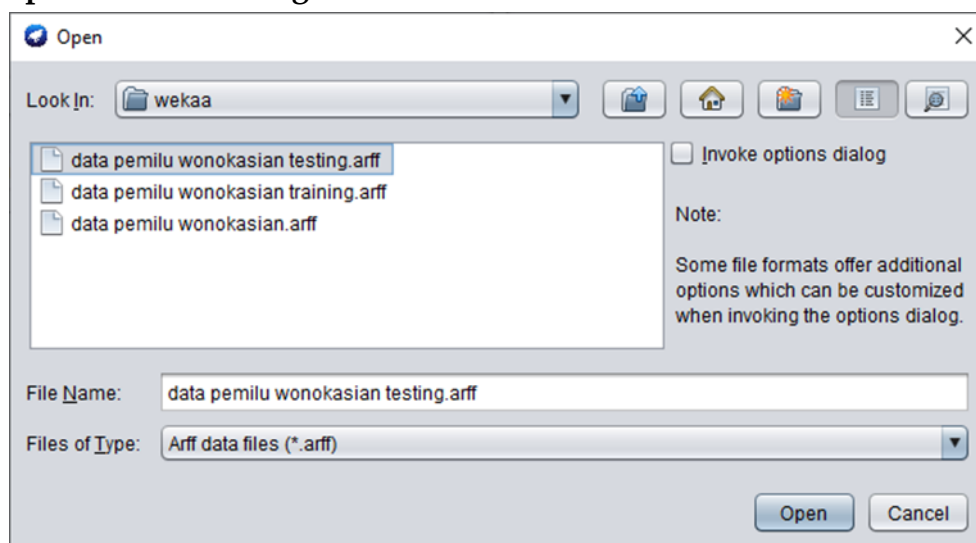
TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
0.783    0.021    0.932     0.783   0.851     0.808  0.977    0.934    TH
0.979    0.217    0.925     0.979   0.951     0.808  0.977    0.988    H
Weighted Avg. 0.927  0.164    0.927     0.927   0.924     0.808  0.977    0.974

=== Confusion Matrix ===

  a  b  <-- classified as
535 148 | a = TH
39 1827 | b = H

```

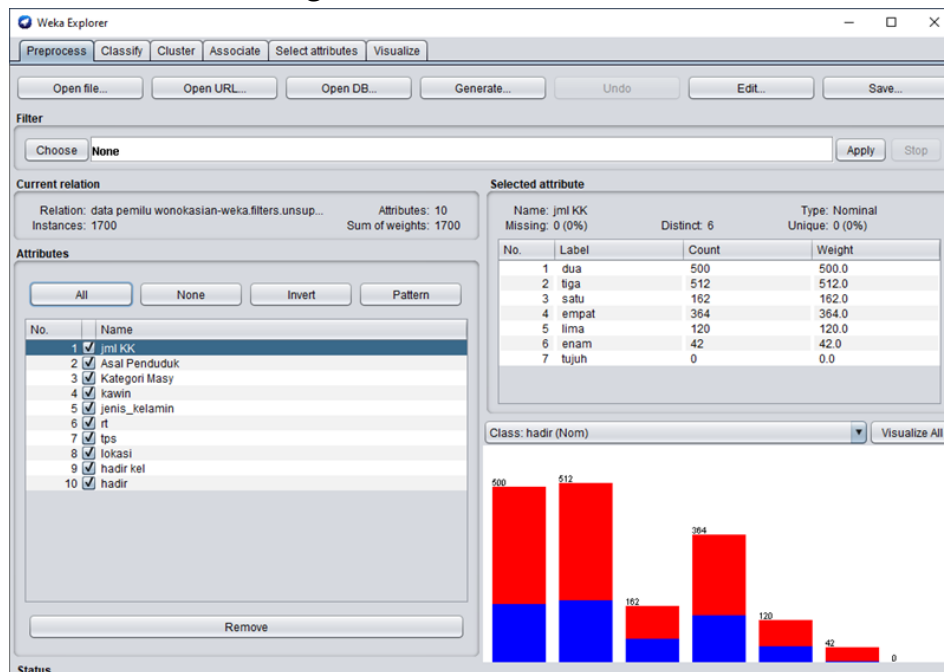
F. Menu Explorer Data Testing



Gambar 8. Menu Explorer Data Testing

Peneliti akan menguji data testing menggunakan algoritma C4.5 pada weka dengan format arff.

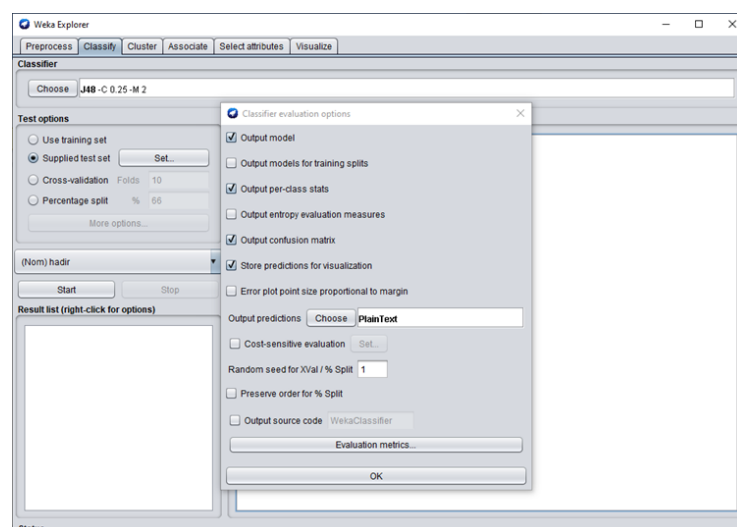
G. Halaman Atribut Data Testing



Gambar 9. Halaman Atribut Data Testing

Peneliti menampilkan atribut yang terdapat pada data testing. terdapat 10 atribut yang digunakan, diantaranya adalah Jumlah KK, Asal Penduduk, Kategori Masyarakat, Kawin, Jenis Kelamin, RT, TPS, Lokasi, Hadir Kel. Class yang digunakan untuk pengujian yaitu Jumlah kehadiran setiap KK dengan nilai seperti yang ditampilkan di gambar 9 dari 1.700 data pemilu.

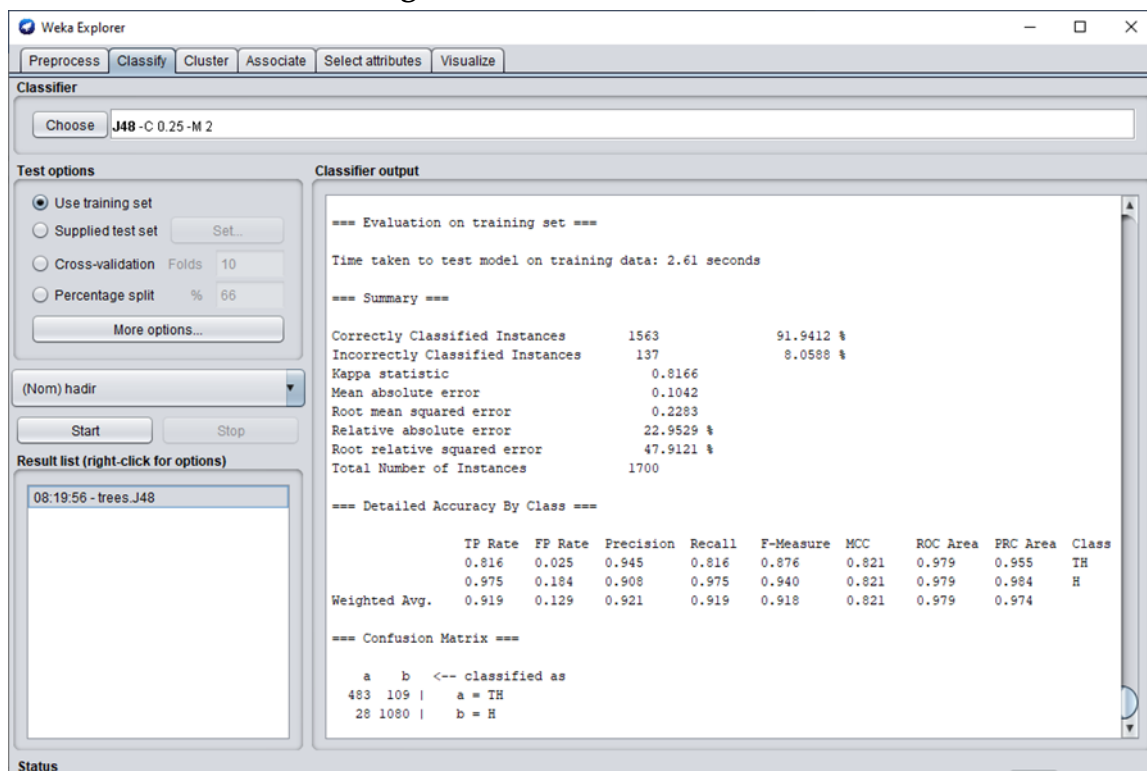
H. Halaman Test Options Data Testing



Gambar 10. Halaman Test Options Data Testing

Peneliti akan memunculkan basis aturan yang digunakan pada data testing. Pada halaman test options, klik supplied tes set lalu klik more options. Pada output predictions pilih folder prediction klik choose lalu pilih PlainText, setelah itu klik Ok. Kemudian akan muncul basis aturan. Klik Start untuk melakukan klasifikasi pada data testing.

I. Hasil Klasifikasi Data Testing



Gambar 10. Hasil Klasifikasi Data Testing

Berikut ini ialah hasil evaluasi data testing, yaitu sebagai berikut: a). Skema berikut diolah menggunakan weka, relasinya data pemilu wonokasian-weka.filters.unsupervised.instance.RemovePercentage-P60.0, jumlah data sebanyak 1.700, jumlah atribut sebanyak 10. b). Hasil menunjukkan basis aturan pada pohon keputusan dengan jumlah data training 1.700 data dan daun (*number of leaves*) sebanyak 60 dan ukuran pohon (*size of the tree*) sebanyak 69. Node diawali atribut (hadir kel) dan diteruskan dengan pencabangan lengan sampai perkabelan daun pada akhir keputusan. c) waktu yang digunakan untuk pengujian data training yaitu 2,61 detik. d) correctly classified instances adalah jumlah hasil klasifikasi yang benar sebanyak 1.563 dengan perolehan presentase sebesar 91,94%. e) incorrectly classified instances adalah jumlah hasil klasifikasi yang salah sebanyak 137 dengan perolehan presentase sebesar 8,05%.

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
 Relation: data pemilu wonokasian-weka.filters.unsupervised.instance.RemovePercentage-P60.0
 Instances: 1700
 Attributes: 10
 jml KK
 Asal Penduduk
 Kategori Masy
 kawin
 jenis_kelamin
 t
 tps
 lokasi
 hadir kel
 hadir
 Test mode: evaluate on training data

=== Classifier model (full training set) ===

J48 pruned tree

```

hadir kel = TSH
| kawin = S: H (241.0/82.0)
| kawin = B
| | tps = TPS-1: TH (0.0)
| | tps = TPS-2: TH (0.0)
| | tps = TPS-3: TH (0.0)
| | tps = TPS-4: TH (0.0)
| | tps = TPS-5: TH (0.0)
| | tps = TPS-6: TH (0.0)
| | tps = TPS-7: TH (0.0)
| | tps = TPS-8: TH (0.0)
| | tps = TPS-9: TH (0.0)
| | tps = TPS-10
| | | jenis_kelamin = P: H (5.0/1.0)
| | | jenis_kelamin = L: TH (9.0/2.0)
| | tps = TPS-15
| | | jml KK = dua
| | | | jenis_kelamin = P: TH (2.0)
| | | | jenis_kelamin = L: H (3.0)
| | | jml KK = tiga: TH (7.0/1.0)
| | | jml KK = satu: TH (0.0)
| | | jml KK = empat: TH (8.0)
| | | jml KK = lima: TH (0.0)
| | | jml KK = enam: TH (0.0)
| | | jml KK = tujuh: TH (0.0)
| | tps = TPS-16: H (23.0/11.0)
| kawin = P: H (17.0/8.0)
hadir kel = THS: TH (404.0)
hadir kel = HS: H (870.0)
  
```

Number of Leaves : 60

Size of the tree : 69

```

=== Evaluation on training set ===

Time taken to test model on training data: 2.91 seconds

=== Summary ===

Correctly Classified Instances   1563      91.9412 %
Incorrectly Classified Instances  137      8.0588 %
Kappa statistic                  0.8166
Mean absolute error              0.1042
Root mean squared error          0.2283
Relative absolute error          22.9529 %
Root relative squared error      47.9121 %
Total Number of Instances       1700

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
      0.816   0.025   0.945    0.816   0.876    0.821  0.979   0.955    TH
      0.975   0.184   0.908    0.975   0.940    0.821  0.979   0.984    H
Weighted Avg.  0.919   0.129   0.921    0.919   0.918    0.821  0.979   0.974

=== Confusion Matrix ===

  a  b  <-- classified as
483 109 | a = TH
28 1080 | b = H

```

J. Hasil Klasifikasi Data 4.249

Weka Explorer

Preprocess | **Classify** | Cluster | Associate | Select attributes | Visualize

Classifier

Choose **J48 - C 0.25 - M 2**

Test options

- ☒ Use training set
- ☐ Supplied test set **Set...**
- ☐ Cross-validation **Folds** 10
- ☐ Percentage split **%** 66
- More options...**

(Nom) hadir

Start **Stop**

Result list (right-click for options)

10:03:39 - trees.J48

Classifier output

```

=== Evaluation on training set ===

Time taken to test model on training data: 0.25 seconds

=== Summary ===

Correctly Classified Instances   3935      92.61 %
Incorrectly Classified Instances  314      7.39 %
Kappa statistic                  0.7577
Mean absolute error              0.1013
Root mean squared error          0.2251
Relative absolute error          30.5178 %
Root relative squared error      55.2495 %
Total Number of Instances       4249

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
      0.713   0.017   0.917    0.713   0.802    0.767  0.971   0.895    TH
      0.983   0.287   0.928    0.983   0.955    0.767  0.971   0.989    H
Weighted Avg.  0.926   0.230   0.926    0.926   0.923    0.767  0.971   0.969

=== Confusion Matrix ===

  a  b  <-- classified as
637 256 | a = TH
58 3298 | b = H

```

Status

Gambar 11. Hasil Klasifikasi Data 4.249

Berikut ini ialah hasil evaluasi data testing, yaitu sebagai berikut: a). Skema berikut diolah menggunakan weka, relasinya data pemilu wonokasian, jumlah data sebanyak 4.249, jumlah atribut sebanyak 10. b). Hasil menunjukkan basis aturan pada pohon keputusan dengan jumlah keseluruhan data 4.249 data dan daun (*number of leaves*) sebanyak 156 dan ukuran pohon (*size of the tree*) sebanyak 177. Node diawali atribut (hadir kel) dan diteruskan dengan pencabangan lengan sampai perkabelan daun pada akhir keputusan. c) waktu yang

digunakan untuk pengujian data training yaitu 0,03 detik. d) correctly classified instances adalah jumlah hasil klasifikasi yang benar sebanyak 3.935 dengan perolehan persentase sebesar 92,61%. e) incorrectly classified instances adalah jumlah hasil klasifikasi yang salah sebanyak 314 dengan perolehan persentase sebesar 7,39%.

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: data pemilu wonokasari

Instances: 4249

Attributes: 10

iml KK

Asal Penduduk

Kategori Masy

kawin

jenis_kelamin

rt

tps

lokasi

hadir kel

hadir

Test mode: evaluate on training data

=== Classifier model (full training set) ===

J48 pruned tree

hadir kel = TSH

| kawin = S: H (618.0/179.0)

| kawin = B

| | tps = TPS-1

| | | iml KK = dua: H (4.0/1.0)

| | | iml KK = tiga: TH (5.0)

| | | iml KK = satu: TH (0.0)

| | | iml KK = empat: H (6.0/2.0)

| | | iml KK = lima: TH (6.0/2.0)

| | | iml KK = enam: TH (0.0)

| | | iml KK = tujuh: TH (0.0)

| | | tps = TPS-2

| | | jenis_kelamin = P

| | | | rt = rt01: TH (0.0)

| | | | rt = rt02: TH (8.0/2.0)

| | | | rt = rt03: H (3.0/1.0)

| | | | rt = rt04: TH (0.0)

| | | | rt = rt05: TH (0.0)

| | | | rt = rt06: TH (0.0)

| | | | rt = rt07: TH (0.0)

| | | | rt = rt08: TH (0.0)

| | | | rt = rt09: TH (0.0)

| | | | rt = rt10: TH (0.0)

| | | | rt = rt11: TH (0.0)

| | | | rt = rt12: TH (0.0)

| | | | rt = rt13: TH (0.0)

| | | | rt = rt14: TH (0.0)

| | | | rt = rt16: TH (0.0)

| | | | rt = rt15: TH (0.0)

| | | | rt = rt17: TH (0.0)

| | | | rt = rt19: TH (0.0)

| | | | rt = rt18: TH (0.0)

| | | | rt = rt20: TH (0.0)

```

| | | rt = rt21: TH (0.0)
| | | jenis_kelamin = L: H (13.0/3.0)
| kawin = P: H (46.0/21.0)
hadir_kel = THS: TH (485.0/1.0)
hadir_kel = HS: H (2696.0)

Number of Leaves : 156
Size of the tree : 177

```

```

=== Evaluation on training set ===

Time taken to test model on training data: 0.03 seconds

=== Summary ===

Correctly Classified Instances   3935      92.61 %
Incorrectly Classified Instances  314       7.39 %
Kappa statistic                  0.7577
Mean absolute error              0.1013
Root mean squared error          0.2251
Relative absolute error          30.5178 %
Root relative squared error      55.2495 %
Total Number of Instances       4249

=== Detailed Accuracy By Class ===

    TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
    0.713    0.017    0.917    0.713    0.802    0.767  0.971    0.895    TH
    0.983    0.287    0.928    0.983    0.955    0.767  0.971    0.989    H
Weighted Avg. 0.926    0.230    0.926    0.926    0.923    0.767  0.971    0.969

=== Confusion Matrix ===

  a  b  <-- classified as
637 256 | a = TH
58 3298 | b = H

```

Simpulan

Berdasarkan hasil klasifikasi pada algoritma C4.5 untuk mengklasifikasi status kehadiran dari dataset 1.700, diperoleh tingkat kebenaran sebesar 91,94% dan tingkat kesalahan sebesar 8,05%. Pada penelitian selanjutnya dapat menggunakan metode lain untuk digunakan sebagai perbandingan.

Daftar Pustaka

- Amalia, E. d. H. (2013). Algoritma C4.5 Untuk Prediksi Hasil Pemilihan Legislatif DPRD DKI Jakarta. *Techno Nusa Mandiri*, Vol. IX(1).
- Aswendy. (2016). *Analisis Data Iklim Indonesia Menggunakan Aplikasi Weka Dengan Metode Klasifikasi Naive Bayes*.
- Bhardwaj, R. B. (2022). Use of ANN, C4.5 and Random Forest Algorithm in the Evaluation of Seismic Soil Liquefaction. *Journal of Soft Computing in Civil Engineering*, 6(2), 92–106. <https://doi.org/10.22115/SCCE.2022.314762.1380>
- Budiman, S. S. N. A. F. Y. P. (2021). *Analisis Pengendalian Mutu di Bidang Industri Makanan*.
- Bustillo, J. C. M. (2022). Predictive Hybridization Model integrating Modified Genetic Algorithm (MGA) and C4.5. *6th International Conference on Electronics, Communication*

- and Aerospace Technology, ICECA 2022 - Proceedings, 1500–1507. <https://doi.org/10.1109/ICECA55336.2022.10009532>
- Chen, X. (2023). Design and research of MOOC teaching system based on TG-C4.5 algorithm. *Systems and Soft Computing*, 5. <https://doi.org/10.1016/j.sasc.2023.200064>
- F, A. S., & others. (2019). Classification Using C4.5 Algorithm in Election Participation Prediction. *International Conference on Engineering, Technologies, and Applied Sciences (ICETsAS)*.
- Han, J., & Kamber, M. (2006). *Data Mining Concepts and Techniques* (Second). Morgan Kaufmann.
- Hendrian, S. (2018). Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan. *Akt. Exacta*, 11(3), 266–274.
- Kamagi, D. & H. S. (2014). Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa. *Ultimatics : Jurnal Teknik Informatika*, 6(1), 15–20.
- Liando, D. M. (2016). Pemilu dan Partisipasi Politik Masyarakat (Studi Pada Pemilihan Anggota Legislatif Dan Pemilihan Presiden Dan Calon Wakil Presiden Di Kabupaten Minahasa Tahun 2014). *J. LPPM Bid. EkoSosBudKum*, 3(2), 14–28.
- Liksha, V. a. P. D. (2018). Aplikasi Akuntansi Pengolahan Data Jasa Service Pada Pt. Budi Berlian Motor Lampung. *J. Sist. Inf. Akunt*, 1(1).
- Mebawundu, O. J. (2022). Network Intrusion Detection Models based on Naives Bayes and C4.5 Algorithms. *Proceedings of the 2022 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development, NIGERCON 2022*. <https://doi.org/10.1109/NIGERCON54645.2022.9803086>
- Mohammad, G. B. (2022). Classification of normal and anomalous activities in a network by cascading C4.5 decision tree and K-means clustering algorithms. *Social Network Analysis: Theory and Applications*, 109–132. <https://doi.org/10.1002/9781119836759.ch7>
- Purnamasari, J. H. Y. P. S. F. I. a. I. W. S. W. (2013). Machine Learning 'Get Easy Using WEKA. *Dapur Buku*.
- Rahardja, U. (2022). Application of the C4.5 Algorithm for Identifying Regional Zone Status Using A Decision Tree in the Covid-19 Series. *APTISI Transactions on Technopreneurship*, 4(2), 164–173. <https://doi.org/10.34306/att.v4i2.234>
- Shamrat, F. M. J. M. (2022). Performance Evaluation Among ID3, C4.5, and CART Decision Tree Algorithm. *Lecture Notes in Networks and Systems*, 317, 127–142. https://doi.org/10.1007/978-981-16-5640-8_11
- Shanthi, J. (2022). A C4.5 decision tree classifier based floorplanning algorithm for System-on-Chip design. *Microelectronics Journal*, 121. <https://doi.org/10.1016/j.mejo.2022.105361>
- Sindunata, R. D. I. S. d. Y. (2014). Penerapan Data Mining untuk Analisa Pola Perilaku Nasabah dalam Pengkreditan Menggunakan Metode C4.5 Studi Kasus pada KSU Insan Kamil Demak. *Jurnal Ilmiah Teknologi Dan Informasi ASIA*, Vol. 8(2).
- Subiyanto, A. E. (2020). *Pemilihan Umum Serentak Yang Berintegritas Sebagai Pembaruan Demokrasi Indonesia*. Mahkamah Konstitusi Republik Indonesia.

- Susanto, S. H. (2014). *Data Mining Untuk Memprediksi Prestasi Siswa Berdasarkan Sosial Ekonomi, Motivasi, Kedisiplinan dan Prestasi Masa Lalu*.
- Suyanto. (2017). *Data Mining Untuk Klasifikasi dan Klasterisasi Data*.
- Tempola, F. (2022). Rule Formation Application based on C4.5 Algorithm for Household Electricity Usage Prediction. *Trends in Sciences*, 19(3).
<https://doi.org/10.48048/tis.2022.2167>
- Wang, J. (2022). Application of C4.5 Decision Tree Algorithm for Evaluating the College Music Education. *Mobile Information Systems*, 2022.
<https://doi.org/10.1155/2022/7442352>
- Wanti, K. D. R. S. P. F. I. R. Z. D. H. S. (2020). *Penerapan Algoritma C4.5 Dalam Meningkatkan Kemampuan Bahasa Inggris Pada Mahasiswa*. PETIR.