

# Penerapan Data Mining Algoritma C4.5 Terhadap Prediksi Faktor Menurunnya Hasil Panen Padi

Nove Viktor Boyke Siahaan<sup>1</sup>, Poningsih<sup>2</sup>, Dedi Suhendro<sup>3</sup>, Dedy Hartama<sup>4</sup>, Suhada<sup>5</sup>

<sup>1,4</sup>STIKOM Tunas Bangsa, Pematangsiantar, Sumatera Utara, Indonesia

<sup>2,3,5</sup>AMIK Tunas Bangsa, Pematangsiantar, Sumatera Utara, Indonesia

email : <sup>1</sup>noveviktor@gmail.com, <sup>2</sup>poningsih@amiktunasbangsa.ac.id,

<sup>3</sup>dedi.su@amiktunasbangsa.ac.id, <sup>4</sup>dedyhartama@amiktunasbangsa.ac.id, <sup>5</sup>suhada.atb@gmail.com

## Abstract

*The aim of the study was to predict the factors causing the decline in rice yields. By knowing the factors of declining rice yields, business owners can further evaluate the causes of declining rice yields and then look for solutions on how to overcome them. The method used in this study is the C4.5 Algorithm, the source of the data used is primary data obtained by direct interviews with rice mill owners and farmers in Siborna Village, Kec. Panei Kab. Simalungun Prov. North Sumatra. The variables used include (1) Pests, (2) Rice Grains, (3) Leaf Color, (4) Planting Month and (5) Planting Method. The results obtained 8 rules for the classification of factors causing the decline in rice yields with 3 increasing decision rules and 5 decreasing decision rules with an accuracy rate of 93.33%. It can be concluded that the predictor of the decline in rice yields is based on the connectedness of the Attributes of Planting Month, Pests, Rice Grains, Leaf Color and Planting Methods.*

**Keywords:** Rice Harvest Results, Predictions, Data Mining, C4.5 Algorithm

## Abstrak

*Tujuan penelitian adalah untuk memprediksi faktor penyebab menurunnya hasil panen padi. Dengan mengetahui faktor menurunnya hasil panen padi, pemilik usaha dapat mengevaluasi lebih lanjut penyebab menurunnya hasil panen padi kemudian mencari solusi bagaimana cara menanggulanginya. Metode yang digunakan dalam penelitian ini adalah Algoritma C4.5, sumber data yang digunakan adalah data primer yang diperoleh dengan wawancara langsung kepada pemilik kilang padi dan petani di Desa Siborna Kec. Panei Kab. Simalungun Prov. Sumatera Utara. Adapun variabel yang digunakan diantaranya (1) Hama, (2) Bulir Padi, (3) Warna Daun, (4) Bulan Tanam dan (5) Cara Tanam. Hasil penelitian diperoleh 8 rules untuk klasifikasi faktor penyebab menurunnya hasil panen padi dengan 3 rules keputusan meningkat dan 5 rules keputusan menurun dengan tingkat akurasi 93,33%. Dapat disimpulkan bahwa prediksi faktor menurunnya hasil panen padi didasari oleh keterhubungan Atribut Bulan Tanam, Hama, Bulir Padi, Warna Daun dan Cara Tanam.*

**Kata kunci:** maksimal 5 kata terpenting dalam makalah

## 1. PENDAHULUAN

Tanaman padi merupakan tanaman budidaya yang memiliki peranan penting bagi umat manusia. Tanaman padi menjadi sumber bahan pangan utama hampir dari setengah penduduk dunia. Di Indonesia padi telah menjadi komoditas pangan utama sehingga banyak masyarakat pedesaan yang memilih untuk menjadi petani padi sebagai salah satu mata pencaharian mereka[1]. Para petani padi khususnya di Siborna Kec. Panei Kab. Simalungun Prov. Sumatera Utara mereka tidak menjual padinya dalam bentuk beras melainkan menjual padi secara borongan atau lebih dikenal dengan istilah sistem tebas. Berdasarkan hasil wawancara yang dilakukan kepada pemilik kilang padi dan petani di Desa Siborna Kec. Panei Kab. Simalungun Prov. Sumatera Utara yang merupakan objek dari penelitian ini, menerangkan

usaha mereka ini bergerak sebagai perusahaan penggilingan padi yang melakukan transaksi pembelian tanaman padi petani secara langsung.

Harga yang terbentuk berdasarkan hasil taksiran pemborong yang dilakukan dengan melihat kondisi padi apakah terserang hama, apakah bulir pada padi banyak atau sedikit, cara tanam padi, musim tanam, warna daun padi. Berdasarkan kriteria tersebut, maka selanjutnya akan ditaksir hasil yang diperoleh dari tanaman padi tersebut untuk ditentukan harga yang sesuai dengan jumlah taksiran. Namun sering kali hasil panen padi mengalami statistik penurunan karena gagal panen, tidak diketahui apa faktor utama yang menyebabkan hasil panen padi petani kerap gagal, hal ini juga berimbas pada turunnya harga jual padi karena jumlah hasil taksiran tidak sesuai dengan harapan, jika hal ini terus berlanjut maka petani padi akan terus mengalami kerugian.

Berdasarkan permasalahan diatas, perlunya dilakukan pengukuran faktor penyebab menurunnya hasil panen padi dengan melakukan wawancara kepada pemilik kilang dan petani di Desa Siborna Kec. Panei Kab. Simalungun Prov. Sumatera Utara. Ada beberapa metode untuk mengetahui faktor penyebab menurunnya hasil panen padi salah satunya adalah melalui pendekatan *Data Mining*[2] dengan menggunakan Algoritma jenis pohon keputusan (*decision tree*) yaitu Algoritma C4.5. Algoritma C4.5 merupakan salah satu algoritma yang digunakan untuk melakukan klasifikasi atau segmentasi atau pengelompokan dan bersifat prediktif[3][4]. Algoritma C.45 adalah salah satu algoritma induksi pohon keputusan, yaitu ID3 (*Iterative Dichotomiser 3*). ID3 dikembangkan oleh J. Ross Quinlan. Dalam prosedur algoritma ID3, input berupa sampel *training*, label *training*, dan atribut. Algoritma C.45 merupakan pengembangan dari ID3[5]. Teknik prediktif melakukan prediksi terhadap data dengan menggunakan hasil-hasil yang telah diketahui dari data yang berbeda. Model ini dapat dibuat berdasarkan penggunaan data historis lain[6].

Tujuan dari penelitian ini adalah Mengklasifikasi faktor menurunnya hasil panen padi menggunakan Algoritma C4.5 dan menguji hasil dari prediksi faktor menurunnya hasil panen padi dengan Algoritma C4.5 menggunakan aplikasi *RapidMiner*. Diharapkan hasil penelitian ini dapat memberikan pengetahuan baru kepada pemilik kilang padi dan petani dari permasalahan faktor mengetahui faktor penyebab menurunnya hasil panen padi yang diperoleh. Dengan mendapatkan hasil tersebut pemilik kilang padi dan petani mendapatkan solusi bagaimana menyelesaikan masalah menurunnya hasil panen padi dan mengurangi resiko kerugian pada usahanya.

## 2. METODOLOGI PENELITIAN

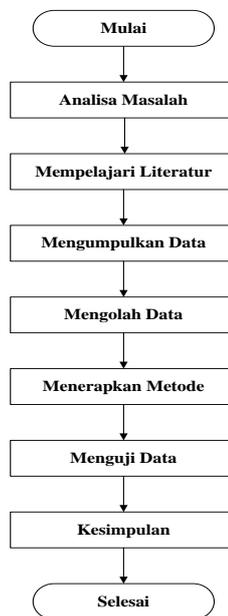
Pada bagian ini Pada bagian ini penulis menggambarkan bagaimana pengumpulan data dalam penelitian ini seperti lokasi dan waktu penelitian, rancangan penelitian, *prosedure* pengumpulan data dan menganalisis data yang akan digunakan pada penelitian ini.

## 2.1. Lokasi Penelitian

Dalam penyusunan skripsi ini penulis melakukan penelitian di Desa Siborna Kec. Panei Kab. Simalungun Prov. Sumatera Utara. Waktu pengumpulan data dan pengolahan data dilakukan selama satu minggu dari tanggal 22 Juni 2020 – 28 Juni 2020.

## 2.2. Rancangan Penelitian

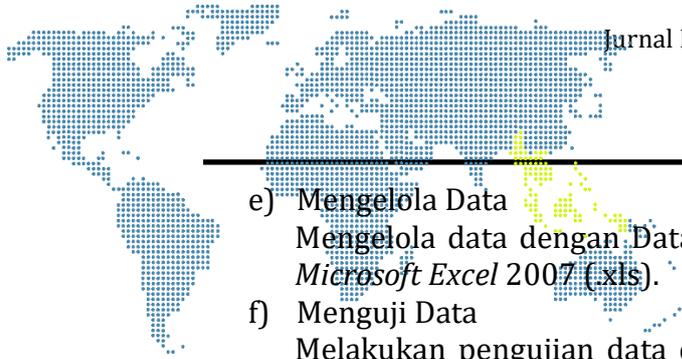
Pada rancangan penelitian ini penulis menguraikan metodologi dan kerangka penelitian kerja yang digunakan dalam menyelesaikan masalah penelitian. Rancangan atau model penelitian ini disajikan dalam rancangan *Folwchart* pada Gambar 1 berikut :



**Gambar 1.** *Flowchart* Rancangan Penelitian

Pada Gambar 1 menjelaskan rancangan penelitian yang dilakukan untuk mengklasifikasi faktor menurunnya hasil panen padi menggunakan Algoritma C4.5 yang terdiri dari :

- a) **Analisa Masalah**  
Menganalisis masalah terkait faktor menurunnya hasil panen padi dan menentukan parameter atau atribut penilaian apa yang digunakan dalam penelitian.
- b) **Mempelajari Literatur**  
Penelitian ini harus didasari rujukan yang digunakan untuk mendapatkan informasi dan teori yang mendukung dalam penelitian.
- c) **Menetapkan Metode**  
Pada penelitian ini penulis menggunakan teknik *data mining* klasifikasi dengan algoritma C4.5 untuk mencari solusi permasalahan penelitian.
- d) **Mengumpulkan Data**  
Data dikumpulkan dengan menggunakan teknik observasi dan wawancara langsung dengan pemilik kilang padi di desa siborna.



- e) **Mengelola Data**  
Mengelola data dengan *Data Transformation Servis (DTS)* kedalam bentuk *Microsoft Excel 2007 (.xls)*.
- f) **Menguji Data**  
Melakukan pengujian data dengan menggunakan aplikasi *RapidMinerversi 5.3*.
- g) **Kesimpulan**  
Kesimpulan yang diperoleh dari klasifikasi faktor menurunnya hasil panen padi memberikan keputusan yang dapat dijadikan sebagai rekomendasi dalam mengurangi resiko kerugian dan meningkatkan hasil panen padi.

### 2.3. Analisis Data

Analisis data dilakukan setelah semua informasi yang dibutuhkan terkumpul. Pada penelitian ini digunakan 5 variabel dalam melakukan klasifikasi terhadap faktor menurunnya hasil panen padi yang dibuat dalam bentuk Kuesioner. Kriteria yang digunakan adalah sebagai berikut :

- a) Hama (C1), yaitu jenis hama yang menyerang tanaman padi.
- b) Bulir Padi (C2), berkaitan dengan kepadatan bulir padi, apakah padi memiliki bulir yang banyak atau sedikit.
- c) Warna Daun (C3), dapat menjadi indikator kondisi padi, apakah tergolong kondisi yang normal atau tidak normal.
- d) Bulan Tanam (C4), menggambarkan kondisi iklim dan musim pada masa tanam padi.
- e) Cara Tanam (C5), yakni cara tanam yang digunakan petani apakah dengan sistem tradisional atau dengan sistem yang modern.

Data yang digunakan akan diolah dari hasil kuesioner, yang telah diberikan kepada para petani di desa siborna. Jumlah Populasi dihitung dari bayak nya kepala keluarga yang ada didesa tersebut adalah sebanyak 116 Kartu Keluarga. Untuk menentukan jumlah sampel data yang dibutuhkan maka dilakukan perhitungan menggunakan Rumus Slovin [7].

$$n = \frac{N}{1+Ne^2} \tag{1}$$

Keterangan :

N = Jumlah Populasi

e = Margin of error 5%

n = Sampel

Perhitungannya adalah :

$$n = \frac{116}{1 + 116x(0,05)^2}$$

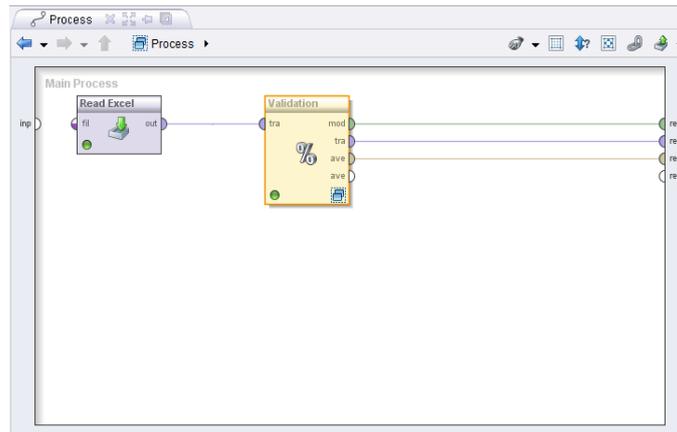
n = 89,922 dibulatkan menjadi 90.

Berdasarkan perhitungan di atas maka penulis memberikan sebanyak 90 form kuesioner kepada masing – masing petani yang ada di Desa Siborna. Kemudian dilakukan analisis data untuk mendapatkan hasil agar peneliti

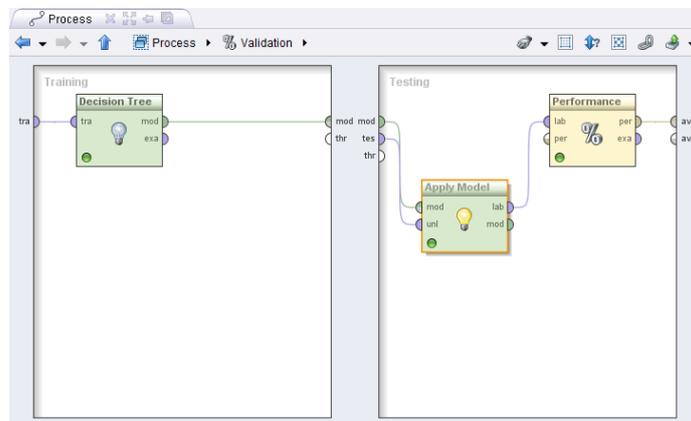
mengetahui prediksi faktor menurunnya hasil panen padi, informasi terkait penelitian ini menggunakan data *kuantitatif* dengan teknik analisis data yang menggunakan jenis statistik *deskriptif*.

#### 2.4. Pemodelan RapidMiner

Pengujian menggunakan Algoritma *C4.5* dengan menggunakan *RapidMiner* dapat dilihat pada Gambar 2 dan Gambar 3 berikut :



Gambar 2. Desain Model Algoritma *C4.5*



Gambar 3. Desain Model Algoritma *C4.5* Didalam Komponen *Split Validation*

Dari Gambar 2 dan Gambar 3 dapat dijelaskan pemodelan Algoritma *C4.5* pada aplikasi *RapidMiner* yaitu :

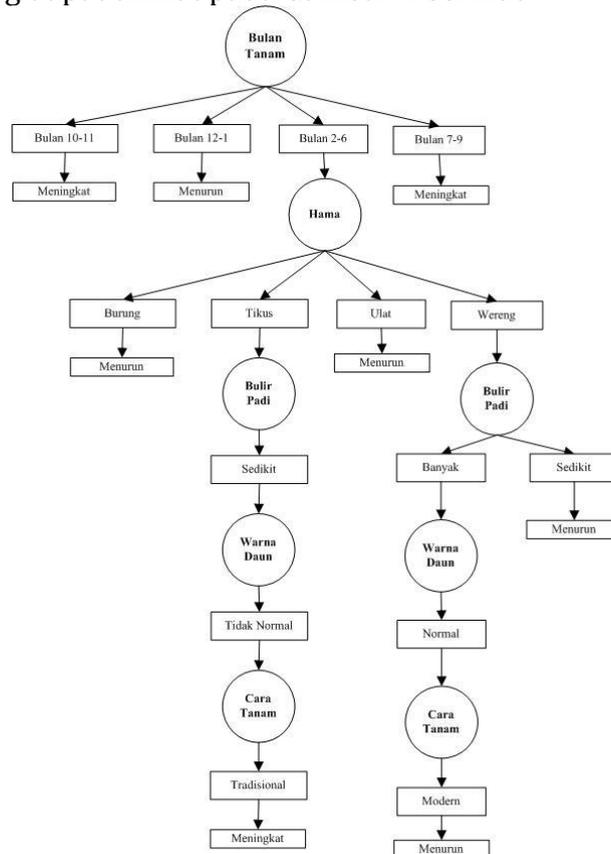
- ReadExcel* merupakan operator yang digunakan untuk mengimport *dataset* dengan tipe file excel (.xls, .xlsx).
- Validation* merupakan operator yang membagi total dari *dataset* menjadi data *training* dan data *testing*. Metode *validation* yang digunakan adalah *Split Validation* yang membagi data *training* dan *testing* berdasarkan nilai *split ratio* yang ditentukan,
- Decision Tree* merupakan metode klasifikasi yang digunakan dalam penelitian ini

- d) *Apply Model* yaitu operator yang menjalankan algoritma *C4.5* yang digunakan dalam penelitian ini
- e) *Performance* digunakan untuk mengukur performa akurasi dari model

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Hasil Pengolahan

Hasil yang diperoleh dari penelitian ini adalah dapat dilihat pada tabel hasil model asosiasi yang dapat dilihat pada Gambar 4 berikut :



**Gambar 4.** Pohon Keputusan Akhir Perhitungan Algoritma *C4.5*

Berdasarkan Gambar 4 dapat terbentuk aturan atau *rule* yaitu 3 (tiga) *rules* keputusan Meningkatkan dan 5 (lima) *rules* keputusan Menurun dijelaskan melalui teks narasi sebagai berikut:

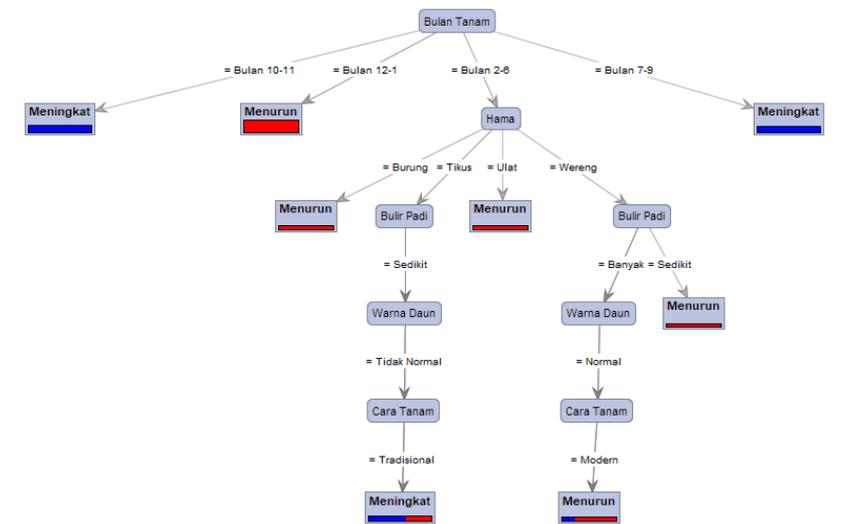
- a) Jika Bulan Tanam = Bulan 10-11, maka hasilnya Meningkatkan {Meningkat = 15, Menurun = 0},
- b) Jika Bulan Tanam = Bulan 12-1, maka hasilnya Menurun {Meningkat = 0, Menurun = 37},
- c) Jika Bulan Tanam = Bulan 2-6 dan Hama = Burung, maka hasilnya Menurun {Meningkat = 0, Menurun = 6},
- d) Jika Bulan Tanam = Bulan 2-6 dan Hama = Tikus dan Bulir Padi = Sedikit dan Warna Daun Tidak Normal dan Cara Tanam = Tradisional, maka hasilnya Meningkatkan {Meningkat = 6, Menurun = 4},



- e) Jika Bulan Tanam = Bulan 2-6 dan Hama = Ulat, maka hasilnya Menurun {Meningkat = 0, Menurun = 5},
- f) Jika Bulan Tanam = Bulan 2-6 dan Hama = Wereng dan Bulir Padi = Banyak dan Warna Daun Normal dan Cara Tanam = Modern, maka hasilnya Menurun {Meningkat = 1, Menurun = 3},
- g) Jika Bulan Tanam = Bulan 2-6 dan Hama = Wereng dan Bulir Padi = Sedikit, maka hasilnya Menurun {Meningkat = 0, Menurun = 1},
- h) Jika Bulan Tanam = Bulan 7-9, maka hasilnya Meningkatkan {Meningkat = 12, Menurun = 0}

### 3.2. Pengujian Dengan Menggunakan Tools *RapidMiner*

Hasil pengolahan data dengan model pohon keputusan sesuai dengan *software RapidMiner*, dapat dilihat pada Gambar 5. sebagai berikut :



Gambar 5. Decision Tree Pada *RapidMiner*

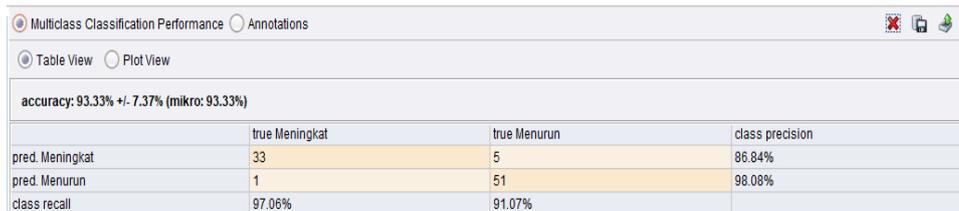
Berdasarkan Gambar 5 tersebut merupakan pohon keputusan yang dihasilkan pada *Rapidminer* dengan aturan atau *rule* yang dapat dilihat pada text view pada Gambar 6 sebagai berikut :

```
Tree
Bulan Tanam = Bulan 10-11: Meningkatkan {Meningkat=15, Menurun=0}
Bulan Tanam = Bulan 12-1: Menurun {Meningkat=0, Menurun=37}
Bulan Tanam = Bulan 2-6
| Hama = Burung: Menurun {Meningkat=0, Menurun=6}
| Hama = Tikus
| | Bulir Padi = Sedikit
| | | Warna Daun = Tidak Normal
| | | | Cara Tanam = Tradisional : Meningkatkan {Meningkat=6, Menurun=4}
| Hama = Ulat: Menurun {Meningkat=0, Menurun=5}
| Hama = Wereng
| | Bulir Padi = Banyak
| | | Warna Daun = Normal
| | | | Cara Tanam = Modern: Menurun {Meningkat=1, Menurun=3}
| | | Bulir Padi = Sedikit: Menurun {Meningkat=0, Menurun=1}
Bulan Tanam = Bulan 7-9: Meningkatkan {Meningkat=12, Menurun=0}
```

Gambar 6. Rule Decision Tree Pada *RapidMiner*

### 3.3. Akurasi Data

Hasil penerapan Algoritma C4.5 menggunakan *software RapidMiner* dengan operator *X Validation* diperoleh nilai akurasi yaitu sebesar 93,33%. dimana untuk *Class Precision* pada prediksi label Meningkat sebesar 86,84% dan *Class Precision* pada prediksi label Menurun sebesar 98,08%. Berikut ini adalah hasil akurasi yang diperoleh.



	true Meningkat	true Menurun	class precision
pred. Meningkat	33	5	86.84%
pred. Menurun	1	51	98.08%
class recall	97.06%	91.07%	

Gambar 7. Nilai Akurasi Algoritma C4.5

### 3.4. Pembahasan

Berdasarkan pengolahan yang telah dilakukan menghasilkan hasil aturan atau *rule* yang sama. Artinya hasil proses yang dilakukan peneliti pada perhitungan Algoritma C4.5 dan *Rapidminer* diperoleh hasil yang sama dan sesuai. Sehingga pengujian dengan *RapidMiner* dapat dikatakan berhasil dan dapat menemukan pohon keputusan pada kasus prediksi faktor menurunnya hasil panen padi didesa Siborna Kec.Panei.

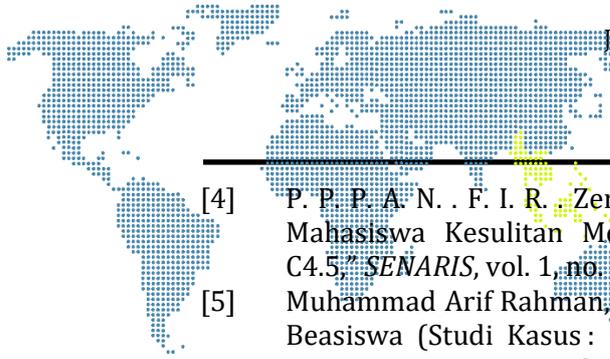
## 4. SIMPULAN

Berdasarkan hasil penelitian yang dilakukan, penulis menarik beberapa kesimpulan diantaranya :

- Data Mining dengan algoritma C4.5 dapat diterapkan untuk mengklasifikasi prediksi factor menurunnya hasil panen padi di Desa Siborna Kec.Panei. Hasil yang diperoleh dari pengujian metode algoritma C4.5 kedalam *RapidMiner* memiliki nilai validiasi yang sama.
- Hasil Pengujian Menggunakan Algoritma C4.5 menggunakan aplikasi *RapidMiner* memiliki nilai *accuracy* sebesar 93,33%. Dapat disimpulkan bahwa prediksi faktor menurunnya hasil panen padi didasari oleh keterhubungan Atribut Bulan Tanam, Hama, Bulir Padi, Warna Daun dan Cara Tanam.

## DAFTAR PUSTAKA

- [1] S. Ramdhaniati, A. Yulyatin, M. Dianawati, and I. Ishaq, "Efektivitas Pola Distribusi dan Minat Petani Terhadap Benih UPBS BPTP Jawa Barat," *BPTP JABAR*, vol. 7, no. 07, pp. 1–8, 2017.
- [2] S. Wahyuni, K. Saputra, and M. I. Perangin-Angin, "Implementasi *RapidMiner* Dalam Menganalisa Data Mahasiswa Drop Out," *Abdi Ilmu*, vol. 10, no. 2, pp. 4516–4524, 2017, doi: 10.1002/pmic.201000579.
- [3] S. Hawani, D. Hartama, A. P. Windarto, and S. Solikhun, "Penerapan C4.5 Untuk Menentukan Calon Suami Terbaik Dalam Pernikahan Pada Kantor KUA Siantar Martoba Pematangsiantar," *JURASIK*, vol. 1, pp. 64–68, 2016.



- [4] P. P. P. A. N. . F. I. R. Zer, D. Hartama, and S. R. Andani, "Analisa Faktor Dominan Mahasiswa Kesulitan Memahami Bahasa Pemrograman Menggunakan Metode C4.5," *SENARIS*, vol. 1, no. September, p. 492, 2019, doi: 10.30645/senaris.v1i0.55.
- [5] Muhammad Arif Rahman, "Algoritma C45 Untuk Menentukan Mahasiswa Penerima Beasiswa (Studi Kasus : Pps Iain Raden Intan Bandar Lampung)," *J. Teknol. Inf. Magister Darmajaya*, vol. 1, no. 02, pp. 118–128, 2015.
- [6] S. Saifullah, M. Zarlis, Z. Zakaria, and R. W. Sembiring, "Analisa Terhadap Perbandingan Algoritma Decision Tree Dengan Algoritma Random Tree Untuk Pre-Processing Data," *J-SAKTI (Jurnal Sains Komput. dan Inform.,* vol. 1, no. 2, p. 180, 2017, doi: 10.30645/j-sakti.v1i2.41.
- [7] A. Purwanto, M. Asbari, and P. B. Santoso, "Pengaruh Kompetensi, Motivasi, Kepemimpinan, Komitmen dan Budaya Kerja Sistem Manajemen Integrasi ISO 9001, ISO 14000 Dan ISO 45001 Pada Industri Otomotif," *PRODUKTIVITAS*, vol. 6, pp. 158–166, 2019.