

Penerapan Algoritma C 5.0 Dalam Analisa Data Potensi Pertanian dan Perternakan (Studi Kasus: Unit Pembinaan Perlindungan Tanaman (UPPT) Biru-Biru)

Nurnia Zamasi*, Nelly Astuti Hasibuan, Suginam

Program Studi Teknik Informatika Universitas Budi Darma, Medan, Indonesia

Email: nurnia11223@gmail.com

Abstrak—Permasalahan yang dihadapi di UPPT di setiap wilayah berupa lahan produktif beralih fungsi menjadi area perumahan, industri, dan sentra bisnis dapat mengakibatkan kurangnya luas lahan untuk bertani dan berternak dan keadaan cuaca yang tidak baik. Salah satu cara untuk menganalisa data tersebut adalah dengan menerapkan atau menggunakan teknik data mining. Dimana data mining merupakan proses analisa data yang besar menjadi informasi untuk menentukan pola dari kesimpulan data yang di analisa. Penerapan metode C. 50 yang nantinya akan menggali data yang sudah ada dan akan membentuk suatu pohon keputusan. Dengan menganalisa data potensi pertanian dan perternakan pihak UPPT mengadakan penyuluhan di wilayah mana jenis usaha bertani dan berternak sehingga tidak berdampak pada pendapatan masyarakat dan juga mengurangi kerugian dalam berusaha bertani dan berternak.

Kata Kunci : Algoritma C.50; Data Mining; Data Potensi

Abstract—The problems faced at UPPT in each area in the form of productive land being converted into residential, industrial, and business centers can result in insufficient land area for farming and livestock and bad weather conditions. One way to analyze this data is to apply or use data mining techniques. Where data mining is the process of analyzing large data into information to determine patterns from the conclusions of the analyzed data. The application of the C 50 method which will later dig up existing data and will form a decision tree. By analyzing data on agricultural and livestock potential, the UPPT holds outreach in areas where the type of business is farming and livestock so that it does not have an impact on community income and also reduces losses in trying to farm and raise livestock.

Keywords: Algorithm C.50; Data Mining; Potential Data

1. PENDAHULUAN

Unit pembinaan pemeliharaan tanaman (UPPT) merupakan suatu instansi yang melakukan pendataan potensi pertanian dan perternakan terutama pendataan data luas komodity tanaman hortikultura, data luas komodity tanaman pangan, Data luas tanaman perkebunan dan data jumlah ternak tiap tahunnya. Tetapi adanya Masalah terhadap pertanian dan perternakan yaitu lahan produktif beralih fungsi menjadi area perumahan, industri, dan sentra bisnis dapat mengakibatkan kurangnya luas lahan untuk bertani dan berternak dan keadaan cuaca yang tidak baik.

Sehingga dapat mempengaruhi potensi pertanian dan peternakan yang nantinya akan berdampak pada pendapatan masyarakat binaan UPPT diwilayahnya masing-masing dan juga dapat mengalami kerugian bagi masyarakat yang bertani ataupun berternak. Untuk mengetahui jenis usaha pertanian atau perternakan di wilayah mana yang lebih berpotensi yaitu dengan mengetahui luas lahannya berapa dan jenis usahanya serta produksi setiap jenis usaha tiap tahunnya di setiap wilayah. Untuk mengatasi permasalahan yang ada di unit pembinaan perlindungan tanaman salah satu teknik digunakan yaitu meningkatkan data potensi pertanian dan perternakan dengan data mining dengan menerapkan algoritma C5.0.

Data mining adalah suatu penyaringan data secara implisit dimana sebelumnya tidak diketahui terdapatnya informasi yang potensial, data mining menganalisis data menggunakan *tools* untuk menemukan pola dan aturan dalam himpunan data, banyak algoritma kalsifikasi dalam metode data mining salah satunya adalah algoritma C5.0.

Pada penelitian terdahulu oleh Ridawati manik, Pristiwanto dan Kennedy tampubolon yang berjudul Prediksi Kolektibilitas Kredit Anggota Dewan Algoritma C5.0 (Studi Kasus: CU Damai Sejahtera Medan) di STMIK Budi Darma, Medan mengatakan "*Algoritma C5.0* berhasil membuat suatu keputusan dalam mengidentifikasi suatu permasalahan dan melihat hubungan antara faktor-faktor yang mempengaruhi suatu masalah. Peranan algoritma C5.0 dalam membangun pohon keputusan bisa dijadikan acuan sebagai alat bantu alternatif pemecahan masalah. Dari hasil klasifikasi berupa pohon keputusan tersebut dapat digunakan untuk hasil penilaian tingkat kinerja hasil klasifikasi prediksi kolektibilitas kredit anggota dewan [1].

2. METODOLOGI PENELITIAN

2.1 Data Mining

Menurut Eko Prasetyo pada buku *Data mining* Konsep dan Aplikasi Menggunakan Matlab mengatakan "Teknik *Data mining* adalah bagaimana mencari data yang tersedia untuk menciptakan sebuah model, lalumemanfaatkan model tersebut untuk mengenali pola data lain yang tidak tersediadidalam basis data yang tersimpan[2]. *Data mining* memiliki 4 (Empat) akar bidang ilmu, yaitu[3].

1. Statistik

Bidang ini merupakan akar paling tua, tanpa ada statistika *Data mining* mungkin tidak ada. Dengan menggunakan statistik data yang diolah bisa diringkas atau dapat dikenal dengan *Exploratory Data Analysis* (EDA). EDA bertujuan sebagai proses identifikasi hubungan yang tersistematis antar variabel/fitur ketika tidak memiliki informasi alami. Teknik EDA klasik yang digunakan dalam *Data mining* di antaranya :

- a. Metode komputasional : *statistic* deskriptif (distribusi, parameter *statistic* klasik (*mean*, *median*, rata – rata, varian dan sebagainya), korelasi, table frekuensi, teknik eksplorasi multivariat (analisis cluster, analisis faktor, analisis komponen utama dan klasifikasi, analisis kanoni, analisis diskriminan, *classification tree*, analisis korespondensi), model *linear (nonlinear)* lanjutan (*regresi linear (nonlinear)*, *time series (forecasting)* dan sebagainya).
 - b. Visualisasi data : mengarah pada representasi informasi dalam bentuk visual dan dapat dipandang sebagai satu yang paling berguna. Pada saat yang sama, visualisasi data merupakan metode eksplorasi data yang atraktif. Teknik visualisasi yang paling umum yang dikenal adalah histogram semua jenis kolom, silinder, kerucut, piramida, lingkaran, batang dan sebagainya), kotak, scatter, kontur matrik, ikon dan sebagainya.
2. Kecerdasan buatan atau *artificial intelligence (AI)*
Artificial Intelligence berbeda dengan statistik. Teorinya dibangun berdasarkan teknik heuristik sehingga *Artificial Intelligence* memiliki peran pada pengolahan informasi sesuai dengan model penalaran manusia. Salah satu cabang dari AI, yaitu pembelajaran mesin atau *machine learning*, merupakan ilmu penting direpresentasikan dalam pembangunan *Data mining*, menggunakan teknik di mana system computer berjalan dengan pelatihan.
 3. Pengenalan Pola
Data mining juga merupakan bagian dari pengenalan pola, hanya saja pengolahan data berasal dari basis data. Data yang digunakan berasal dari basis data kemudian dilakukan pengolahan dimana proses pengolahan bukan pada bentuk relasi, tetapi pada bentuk normal. *Data mining* memiliki ciri tersendiri yakni melakukan pencarian pola asosiasi serta melakukan pencarian pola sekuensial.
 4. Sistem Basis Data
Sistem basis data pada *Data mining* memiliki fungsi sebagai penyedi informasi data yang akan diproses untuk penggaliannya dengan memanfaatkan metode – metode yang sudah ada.

2.2 Algoritma C 5.0

Algoritma C 5.0 merupakan algoritma berbasis *decision tree* yang merupakan penyempurnaan dari algoritma ID3 dan C4.5 yang dibentuk oleh Ross Quinlan pada tahun 1987. Algoritma C5.0 dapat menangani atribut kontinyu dan diskrit. Pemilihan atribut dalam algoritma ini akan diproses menggunakan *information gain*. Atribut dengan nilai *Gain* tertinggi akan dipilih sebagai akar bagi node selanjutnya [1].

2.3 Pertanian

Pengertian pertanian secara luas yaitu pemanfaatan sumber daya hayati yang dilakukan oleh manusia dengan cara menanam tanaman produktif yang dapat menghasilkan dan dipergunakan untuk kehidupan. Menurut Kamus Besar Bahasa Indonesia (KBBI) pertanian merupakan kegiatan bertani yang berkaitan dengan tanam-menanam [4].

3. HASIL DAN PEMBAHASAN

Analisa berfungsi untuk mengetahui dan mengamati permasalahan yang sedang terjadi, kemudian merumuskan masalah dan mencari cara solusi untuk memecahkan masalah yang ada. Untuk mengetahui data potensi pertanian dan perternakan setiap tahunnya di wilayah mana yang pertanian dan perternakannya lebih berpotensi.

Masalah terhadap pertanian dan perternakan yaitu lahan produktif beralih fungsi menjadi area perumahan, industri, dan sentra bisnis dapat mengakibatkan kurangnya luas lahan untuk bertani dan berternak dan juga cuaca yang tidak baik. Sehingga dapat mempengaruhi potensi pertanian dan perternakan yang nantinya akan berdampak pada pendapatan masyarakat binaan UPPT di wilayahnya masing-masing.

Tahapan untuk mengetahui jenis usaha pertanian atau perternakan di wilayah mana yang berpotensi adalah dengan mengumpulkan data, memeriksa kejelasan ataupun kelengkapan data, melakukan proses identifikasi dan proses klasifikasi terhadap semua pernyataan yang ada pada instrumen pengumpulan data berdasarkan variabel yang sedang diteliti, melakukan proses pengujian kualitas data, proses membuat deskripsi data dengan menyajikannya dalam bentuk tabel agar memahami karakteristik data sampel dari suatu penelitian dan tahap akhir melakukan pengujian terhadap proposisi apakah ditolak atau diterima dan memiliki makna atau tidak. Atas dasar hipotesis inilah nantinya keputusan akan dibuat.

Salah satu cara untuk menganalisa data tersebut adalah dengan menerapkan atau menggunakan teknik data mining. Dimana data mining merupakan proses analisa data yang besar menjadi informasi untuk menentukan pola dari kesimpulan data yang di analisa. Penerapan algoritma C. 50 yang nantinya akan menggali data yang sudah ada dan akan membentuk suatu pohon keputusan, dimana pohon keputusan inilah nantinya yang akan membantu unit pembinaan perlindungan tanaman di setiap wilayah agar mengetahui pengelompokan dan pengolahan data potensi pertanian dan perternakan agar usaha untuk pertanian dan perternakan data potensi tiap tahunnya meningkat sehingga tidak mengalami kerugian.

Tabel 1. Daftar Data Potensi Pertanian Dan Perternakan

No	Kecamatan	Jenis Usaha	Luas Lahan (Ha/ekor)	Produksi	Hasil
1	Lubuk Pakam	Padi Sawah	14	500	Baik
2	Lubuk Pakam	Domba	18	100	Tidak
3	Lubk Pakam	Domba	10	150	baik
4	Lubuk Pakam	Padi Sawah	12	350	baik
5	Lubuk Pakam	Padi Sawah	12	400	baik
6	Percut Sei Tuan	Padi Sawah	15	100	Tidak
7	Percut Sei Tuan	Padi Sawah	19	400	baik
8	Percut Sei Tuan	Domba	17	200	baik
9	Percut Sei Tuan	Domba	19	230	baik
10	Percut Sei Tuan	Padi Sawah	18	275	baik
11	Batang Kuis	Padi Sawah	12	275	baik
12	Batang Kuis	Padi Sawah	22	350	Baik
13	Batang Kuis	Padi Sawah	29	345	Baik
14	Batang Kuis	Padi Sawah	19	175	Tidak
15	Tanjung Morawa	Domba	24	98	Tidak
16	Lubuk Pakam	Domba	27	100	Tidak
17	Lubuk Pakam	Domba	10	103	Tidak
18	Lubuk Pakam	Domba	12	115	Tidak
19	Tanjung Morawa	Domba	12	93	Tidak
20	Tanjung Morawa	Domba	18	89	Tidak
21	Percut Sei Tuan	Padi Sawah	15	110	Tidak
22	Percut Sei Tuan	Padi Sawah	14	210	Baik
23	Percut Sei Tuan	Padi Sawah	19	130	Tidak
24	Percut Sei Tuan	Padi Sawah	21	125	Tidak
25	Tanjung Morawa	Padi Sawah	20	320	Baik
26	Tanjung Morawa	Padi Sawah	13	250	Baik
27	Lubuk Pakam	Padi Sawah	18	112	Tidak
28	Lubuk Pakam	Padi Sawah	17	287	Baik
29	Lubuk Pakam	Padi Sawah	19	99	Tidak
30	Lubuk Pakam	Domba	15	87	Tidak

Sebelum menerapkan data pada algoritma C.50 yang digunakan perlukiranya dilakukan tahapan preprosesing yang berguna untuk menyesuaikan data terhadap metode yang digunakan. Adapun preprosesing data dapat dilihat pada tabel 4.2 – 4.3 dibawah ini:

Tabel 2. PreprosesingLuasLahan

No	Luas Lahan	Keterangan
1	0 – 15	Kecil
2	16 – 30	Sedang
3	>30	Luas

Keterangan Preprosesing Suhu :

1. Jika luas lahan diantara 0-15 Ha maka dikategorikan kecil
2. Jika luas lahan diantara 16-30 Ha maka dikategorikan sedang
3. Jika luas lahan diantara >30 Ha maka dikategorikan luas

Tabel 3. Preprosesing Produksi

No	Usaha	Produksi	Keterangan
1	PadiSawah	0 – 200	Sedikit
2		201 – 450	Sedang
3		>450	Banyak
4	Domba	0 – 100	Sedikit
5		101 – 200	Sedang
6		>200	Banyak

Keterangan Preprosesing Suhu :

1. Jika jenis usaha padi sawah dengan produksi diantara 0-200 maka dikategorikan sedikit.
2. Jika jenis usaha padi sawah dengan produksi diantara 201-450 maka dikategorikan sedikit.
3. Jika jenis usaha padi sawah dengan produksi diantara >450 maka dikategorikan banyak.
4. Jika jenis usaha domba dengan produksi diantara 0-100 maka dikategorikan sedikit.

5. Jika usaha padi sawah dengan produksi diantara 101 - 200 maka dikategorikan sedang.

6. Jika usaha padi sawah dengan produksi diantara >200 maka dikategorikan sedikit.

Langkah menghitung gain dan *entropy* yang dimana *gain* dan *entropy* digunakan untuk pembentukan pohon keputusan sebagai berikut :

Tabel 4. Perhitungan Nilai Node 1

Node	Atribut	Nilai	Jumlah Kasus / Data Set	Baik	Tidak Baik	Entropy	Gain
1			30	13	17	0.987137774	
	Kecamatan						0.09400486
		Lubuk Pakam	12	4	8	0.918295834	
		Percut Sei Tuan	9	5	4	0.99107606	
		Batang Kuis	4	3	1	0.811278124	
		Tanjung Morawa	5	1	4	0.721928095	
	Jenis Usaha						0.04510854
		Padi Sawah	19	10	9	0.998000884	
		Domba	11	3	8	0.845350937	
	Luas Lahan						0.0248777
		Kecil	13	7	6	0.995727452	
		Sedang	17	6	11	0.936667382	
	Produksi						0.5688174
		Sedikit	13	0	13	0	
		Sedang	15	11	4	0.836640742	
		Banyak	2	2	0	0	

Penjelasan :

Jumlah kasus sebanyak 30

Kategori yang baik 13

Kategori yang tidak baik 17

Cara perhitungan Entropy = $((-13/30) * (13/30) + (17/30) * (17/30))$.

Cara perhitungan Gain = $0,987137774 - ((12/30) * 0,918295834) - ((9/30) * 0,99107606) - ((4/30) * 0,811278124) - ((5/30) * 0,721928095)$.

Tabel 5. Perhitungan Nilai Node 1.1

Node	Atribut	Nilai	Jumlah Kasus / Data Set	Baik	Tidak Baik	Entropy	Gain
1.1	Produksi -- >Sedang		15	11	4	0.836640742	
	Kecamatan						0.30330741
		Lubuk Pakam	6	3	3	1	
		Percut Sei Tuan	4	4	0	0	
		Batang Kuis	3	3	0	0	
		Tanjung Morawa	2	1	1	1	
	Jenis Usaha						0.06834589
		Padi Sawah	11	9	2	0.684038436	
		Domba	4	2	2	1	
	Luas Lahan						0.00116948
		Kecil	8	6	2	0.811278124	
		Sedang	7	5	2	0.863120569	

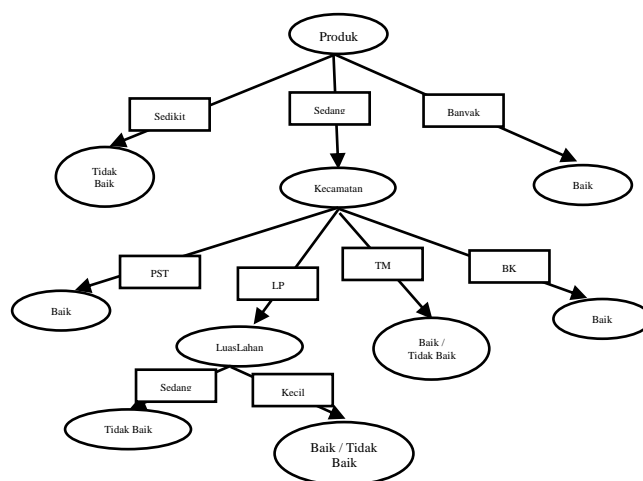
Nilai entropy dari masing-masing atribut telah didapatkan dan gain atribut juga sudah dihitung, selanjutnya hasil dari perhitungan tersebut dimasukkan ke tabel. Tabel akhir yang terbentuk dari perhitungan gain dan entropy.

Tabel 6. Nilai Perhitungan Node 1.1.1

Node	Atribut	Nilai	Jumlah Kasus / Data Set	Baik	Tidak Baik	Entropy	Gain
1.1.1	Produksi -->Sedang Kecamatan -- >LubukPakam		6	3	3	1	

Node	Atribut	Nilai	Jumlah Kasus / Data Set	Baik	Tidak Baik	Entropy	Gain
	Jenis Usaha						0.15087708
		Padi Sawah	3	1	2	0.918295834	
		Domba	3	2	2	0.779950001	
	Luas Lahan						0.1908745
		Kecil	5	2	3	0.970950594	
		Sedang	1	0	1	0	

Dari perhitungan pada tabel 4.9 dapat diketahui bahwa atribut dengan gain tertinggi adalah nilai yaitu 0,1908745. Adapun pohon keputusan yang terbentuk dari node 1.1.1 adalah seperti pada gambar 1.



Gambar 1. Pohon Keputusan Hasil

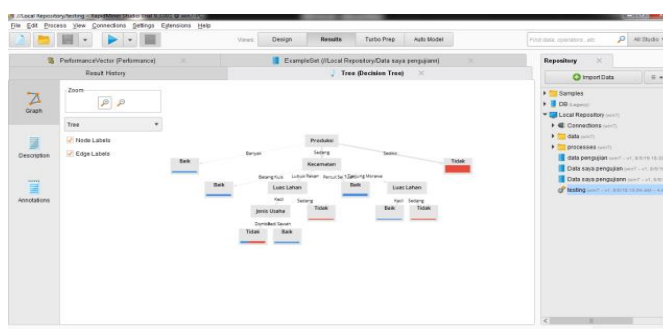
Dari hasil perhitungan node 1.1.1 tidak ada lagi yang akan menjadi atribut akar, maka perhitungan dihentikan. Rule hasil dari prediksi berdasarkan pada pohon keputusan terakhir yang terbentuk sesuai dengan perhitungan *Entropy* dan *Gain*. Adapun aturan atau rule yang terbentuk dari perhitungan *Gain* dan *Entropy* setiap variabel sampai menghasilkan node 1.1.1 adalah sebagai berikut:

1. IF Produksi = sedikit then hasil = Tidak baik atau tidak berpotensi
2. IF Produksi = sedang and kecamatan = PST then hasil = Baik atau berpotensi
3. IF Produksi = sedang and kecamatan = LP, luas lahan, sedang then hasil = Tidak baik atau tidak berpotensi
4. IF Produksi = sedang and kecamatan = LP, luas lahan kecil then hasil = Baik / tidak baik
5. IF Produksi = sedang and kecamatan = TM then hasil = Baik atau tidak baik
6. IF Produksi = sedang and kecamatan = BK then hasil = Baik atau berpotensi
7. IF Produksi = banyak then hasil = Baik atau berpotensi

3.2 Hasil Pengujian

Setelah melakukan perancangan pada aplikasi Rapidminer 9.3 maka pada tahap ini maka dilakukan pengujian melalui aplikasi *Rapidminer 9.3* dengan cara menekan tombol *run or resume the current process* atau dengan menekan *shortcut* tombol *F11*.

Hasil akhir atau *output* dari aplikasi *RapidMiner* adalah berupa pohon keputusan atau *decision tree/id3*. Dari pohon keputusan inilah akan dihasilkan *rules* atau aturan-aturan yang dapat membantu dalam mengambil keputusan. Pohon keputusan akhir dapat dilihat pada gambar 2 berikut:



Gambar 2. Tampilan Hasil Pohon Keputusan Pengujian Data

Pada tahap pengujian ini di *Run* pada proses yang telah dijalankan. Hasil dari pengujian *Association Rule (Table)* yang dihasilkan dari *Rapidminer 9.3* dapat dilihat pada gambar 3:

Tree

```

Produksi = Banyak: Baik (Baik=2, Tidak=0)
Produksi = Sedang
| Kecamatan = Batang Kuis: Baik (Baik=3, Tidak=0)
| Kecamatan = Lubuk Pakam
| | Luas Lahan = Kecil
| | | Jenis Usaha = Domba: Tidak (Baik=1, Tidak=2)
| | | Jenis Usaha = Padi Sawah: Baik (Baik=2, Tidak=0)
| | Luas Lahan = Sedang: Tidak (Baik=0, Tidak=1)
| Kecamatan = Percut Sei Tuan: Baik (Baik=4, Tidak=0)
| Kecamatan = Tanjung Morawa
| | Luas Lahan = Kecil: Baik (Baik=1, Tidak=0)
| | Luas Lahan = Sedang: Tidak (Baik=0, Tidak=1)
Produksi = Sedikit: Tidak (Baik=0, Tidak=13)

```

Gambar 3. Hasil *Association Rules (Table)*

Hasil pengujian Algoritma C 5.0 yang dihasilkan oleh aplikasi *Rapidminer* pada menu *Description* dapat dilihat pada gambar 4:

accuracy: 98.67%

	true Baik	true Tidak	class precision
pred Baik	12	0	100.00%
pred Tidak	1	17	94.44%
class recall	92.31%	100.00%	

Gambar 4. *Accuracy data*

4. KESIMPULAN

Berdasarkan hasil analisa dari penelitian yang dilakukan, maka kesimpulan dengan memanfaatkan datamining dapat mengetahui potensi pertanian dan peternakan di setiap wilayah mana yang lebih berpotensi pada setiap tahunnya. Algoritma C5.0 dapat memberikan informasi berupa rule analisa untuk menggambarkan proses yang terkait dengan data potensi pertanian dan peternakan. Aplikasi *RapidMiner Classification Decision Tree* digunakan sebagai aplikasi pendukung keputusan dan pengujian atas hasil yang didapatkan secara manual, yang menghasilkan sebuah pohon keputusan. Dari pohon keputusan inilah akan menghasilkan sebuah aturan-aturan yang dapat membantu pihak UPPT Biru-biru dalam menganalisa potensi pertanian dan peternakan di setiap wilayahnya serta mudah dipahami oleh pengguna aplikasi.

REFERENCES

- [1] Ridawati Manik, Pristiwanto, and Kennedy Tampubolon, "Prediksi Kolektibilitas Kredit Anggota Dengan Algoritma C5.0 (Studi Kasus: CU Damai Sejahtera Medan)," *Jurnal Riset Komputer (JURIKOM)*, vol. 5, no. 2, pp. 151-160, 2018.
- [2] Eko Prasetyo, *Data Mining Konsep dan Aplikasi Menggunakan Matlab*, 1st ed., Nikodemus WK, Ed. Yogyakarta: CV. Andi Offset, 2012.
- [3] Eko Prasetyo, *Data Mining Mengolah Data Menjadi Informasi Dengan Menggunakan MATLAB*. Yogyakarta: CV. Andi Offset, 2014.
- [4] Kamus Besar Bahasa Indonesia. [Online]. <https://kbbi.kata.web.id/pertanian/> [Accessed 18 Mei 2019].
- [5] [Online]. Available: <http://www.pengertianku.net/2016/04/pengertian-pertanian-secara-umum.html>. [Accessed 10 Mei 2019].
- [6] "Kamus Besar Bahasa Indonesia," [Online]. Available: <https://kbbi.kata.web.id/perternakan/>. [Accessed 18 Mei 2019].
- [7] B. S. D. Oetomo, *Perencanaan dan Pengembangan Sistem Informasi*, Yogyakarta: CV. Andi Offset, 2002.
- [8] A. Preantoko, *Jago Microsoft Excel 2010*, Jakarta: Kawan Pustaka, 2010.
- [9] D. Aprilla D. A. Baskoro, L. Ambarwati, and I. W.S. Wicaksana, *Belajar Data Mining Dengan Rapid Miner.*, 2013.