

ALGORITMA C4.5 UNTUK MEMPREDIKSI PRODUKTIVITAS TANAMAN PADI DI KARAWANG

C4.5 Algorithm to Predict Productivity of Rice Plants in Karawang

Sofi Defiyanti¹, Muhammad Syafi'i², Budi Arif Darmawan³

^{1,3}Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang

²Fakultas Pertanian Universitas Singaperbangsa Karawang

Jl. HS. Ronggowaluyo Telukjambe Timur Karawang - Jawa Barat - Indonesia

Telp. (0852) 22365323, Fax. (0852) 22365323

E-mail: Sofi.defiyanti@staff.unsika.ac.id, muhhammad.syafii@staff.unsika.ac.id, budi.arif@staff.unsika.ac.id

(Makalah diterima 19 Januari 2019 – Disetujui 06 Desember 2019)

ABSTRAK

Karawang merupakan salah satu daerah penghasil padi terbesar di provinsi Jawa Barat, maka tak salah Karawang dijuluki sebagai kota lumbung padi. Tetapi saat ini Karawang lebih dikenal dengan daerah industri. Ini berefek kepada hasil pertanian yang ada di Karawang. Data dari tahun 2010-2015 didapat hasil bahwa sebesar 67% produktivitas padi di Karawang berada dalam kategori rendah sisanya berada dalam kategori sedang dan tinggi, terlihat bahwa banyak lahan pertanian di Karawang yang sudah beralih fungsi yang dari awalnya persawahan menjadi penggunaan lahan lain. Hal ini harusnya dapat mendorong pemerintah Kabupaten Karawang untuk lebih perhatian terhadap masalah ini. Prediksi yang dilakukan dengan menggunakan teknik Data Mining dengan metode CRISP-DM dan algoritma C4.5 telah dilakukan dengan akurasi sebesar 66,1%. Penelitian ini memperlihatkan bahwa selain nilai produksi dan luas tanam untuk menentukan nilai produktivitas tanaman padi, ternyata produktivitas tanaman padi juga dipengaruhi oleh beberapa faktor seperti Luas Baku sawah, Rata - Rata Curah Hujan, dan Luas Panen.

Kata kunci: algoritma, padi, prediksi, produktivitas, Karawang

ABSTRACT

Karawang is one of the largest rice-producing regions in the province of West Java, so Karawang is not mistaken as the city of rice barns. But increasingly here Karawang is better known as an industrial area. This has an effect on agricultural products in Karawang. Data from 2010-2015 showed that 67% of paddy productivity in Karawang was in a low category while the rest was in the medium and high category, it was seen that many agricultural lands in Karawang had shifted functions from rice fields to other land uses. This should encourage the Karawang regency government to pay more attention to this problem. Predictions made using Data Mining techniques using the CRISP-DM method and C4.5 algorithm have been carried out with an accuracy of 66.1%. This study shows that in addition to the value of production and planting area to determine the value of productivity of rice, it turns out that the productivity of rice plants is also influenced by several factors such as the area of rice fields, average rainfall, and harvest area.

Key words: algorithm, rice, prediction, productivity, Karawang

PENDAHULUAN

Padi (*Oryza Sativa*) merupakan salah satu tanaman yang penting di Indonesia. Dari tanaman padi dihasilkan beras yang merupakan bahan makanan pokok bagi rakyat Indonesia. Padi dapat tumbuh dengan baik di daerah panas dengan curah hujan yang tinggi. Produksi padi di Indonesia sebesar 9,67% dari total produksi padi di dunia. karenanya Indonesia menempati posisi ke-3 negara penghasil beras terbesar di dunia setelah China sekitar 30% dan India sekitar 21% dari produksi padi di dunia (FAO, 2018).

Indonesia adalah negara dengan jumlah kebutuhan padi yang terus meningkat seiring dengan peningkatan jumlah penduduknya. Sampai saat ini hampir seluruh masyarakat Indonesia masih menjadikan padi atau beras sebagai makanan pokok. Sehingga produksi padi perlu ditingkatkan untuk memenuhi kebutuhan penduduk yang terus bertambah (Ikhwani, 2014). Kekurangan beras atau pangan dapat memicu kekacauan politik, sosial, dan ekonomi. BPS 2010 mencatat bahwa sejak tahun 2007 sampai 2010 produksi beras di Indonesia telah mengalami surplus antara 0,84 sampai 4,31 juta ton (Ikhwani, 2014). Secara politis, pemerintah menjadikan beras sebagai komoditas strategis dalam pembangunan ekonomi dan swasembada beras menjadi target pertumbuhan. Untuk itu telah banyak program-program pemerintah yang berkaitan dengan upaya peningkatan produksi beras (padi).

Produksi padi pada tahun 2015 meningkat hampir empat kali lipat dibandingkan tahun 1980. Gambar 1 menampilkan perkembangan luas panen dan produktivitas dari tanaman padi dalam waktu 35 tahun tidak mengalami peningkatan yang signifikan jika dibandingkan produksi padi. Produktivitas tanaman padi ke depannya akan tidak stabil dan mengalami stagnan. Rendah dan tidak stabilnya pertumbuhan produksi padi disebabkan oleh lambatnya pertumbuhan luas area panen, rehabilitasi jaringan irigasi, serta gejala melambannya pertumbuhan produktivitas tanaman padi masih belum berhasil dipecahkan (Kusnadi *et al.*, 2011).

Produktivitas padi erat kaitannya dengan penggunaan faktor produksi karena produktivitas menyangkut seberapa besar jumlah output yang dihasilkan untuk setiap unit input tertentu. Produksi padi di Indonesia dipengaruhi oleh beberapa faktor seperti luas lahan, benih kompos, urea dan tenaga kerja berpengalaman (Gultom *et al.*, 2014). Produksi padi juga dipengaruhi oleh beberapa faktor luar (eksternal) dan dalam (internal). Faktor luar yang mempengaruhi produksi padi yaitu keadaan iklim dan geografis. Keadaan iklim meliputi curah hujan, suhu, cahaya matahari, air, dan musim (Ina, 2007). Keadaan geografis meliputi ketinggian tempat dan luas lahan. Peran sumber daya air juga berperan strategis dalam mempengaruhi produktivitas dan spektrum komoditas yang diusahakan (Rohmani *et al.*, 2015). Curah hujan merupakan unsur iklim yang paling berperan dalam menentukan produksi padi (Apriyana dan Lindawati, 2015).

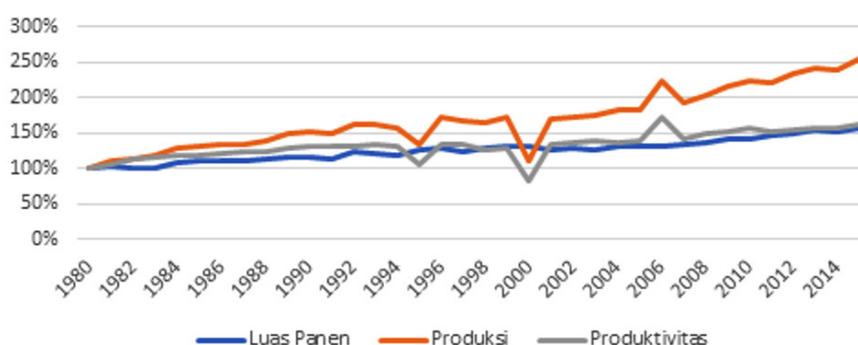
Sebuah prediksi sangat diperlukan untuk mengetahui gambaran di masa depan mengenai hasil produktivitas tanaman padi apakah akan meningkat atau menurun. Hal ini dapat membantu para pihak berkepentingan untuk merencanakan program-program yang dibuat agar tepat sasaran dan efektif.

Algoritma C4.5 menjadi pilihan pertama yang sering dipergunakan dalam pengembangan data mining karena kecepatannya dalam mengklasifikasikan pohon keputusan. Algoritma C4.5 digunakan karena dapat menangani nilai atribut yang hilang serta dapat menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat di antara algoritma yang lainnya (Dhika, 2015; Karyono, 2016).

MATERI DAN METODE

Algoritma C4.5

Algoritma C4.5 adalah salah satu metode untuk membuat *decision tree* berdasarkan data latih yang telah disediakan. Algoritma C4.5 merupakan pengembangan



Gambar 1. Perkembangan Produksi, Luas panen, dan produktivitas tanaman padi sawah tahun 1980-2015 (sumber : <http://www.pertanian.go.id/>)

dari algoritma ID3. Beberapa pengembangan yang dilakukan pada pada C4.5 berupa penanganan *missing value*, *continue data*, dan *pruning* (Patel dan Upadhyay, 2012).

Terdapat beberapa tahapan dalam membuat sebuah *decision tree* dalam algoritma C4.5 (Rokach *et al.*, 2015) yaitu:

1. Mempersiapkan *data training*. *Data training* merupakan data yang diambil dari *data history* atau disebut data masa lalu dan sudah dikelompokkan dalam kelas-kelas tertentu. Contohnya dalam penelitian ini adalah data mengenai produktivitas yang dijelaskan pada bagian 3.2.
2. Menghitung akar dari pohon. Akar akan diambil dari atribut yang akan terpilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi akan menjadi akar pertama. Sebelum menghitung nilai gain dari atribut, hitung dahulu entropi. Untuk menghitung nilai entropi digunakan rumus pada persamaan 1:

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

Keterangan :

- S = Himpunan kasus
- n = Jumlah partisi S
- pi = Proporsi Si terhadap S

Kemudian hitung nilai gain menggunakan rumus pada persamaan 2.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Keterangan :

- S = Himpunan kasus
- A = Fitur
- n = Jumlah partisi atribut A
- |Si| = Proporsi Si terhadap S
- |S| = Jumlah kasus dalam S

$$Gain Ratio(A) = Gain(A)/Entropy(A) \quad (3)$$

3. Buat cabang untuk tiap-tiap nilai *Gain ratio* dari persamaan 3 yang dihasilkan dengan memilih nilai *Gain ratio* terbesar.
4. Mengulang langkah ke-2 dan langkah ke-3 hingga semua *record* terpartisi.
5. Proses partisi pohon keputusan akan berhenti saat :
 - a. Semua *record* dalam simpul N mendapat kelas yang sama.
 - b. Tidak ada atribut di dalam *record* yang di partisi lagi.
 - c. Tidak ada *record* di dalam cabang yang kosong.

Metode Penelitian

Penelitian ini menggunakan metodologi penelitian kuantitatif dengan memanfaatkan metode yang ada dalam *data mining* yaitu metode CRISP-DM (*Cross Industry Standard Process for Data Mining*) (Wirth, 2000) dengan langkah-langkah seperti pada gambar 2.

a. Pemahaman Bisnis (*Business Understanding*)

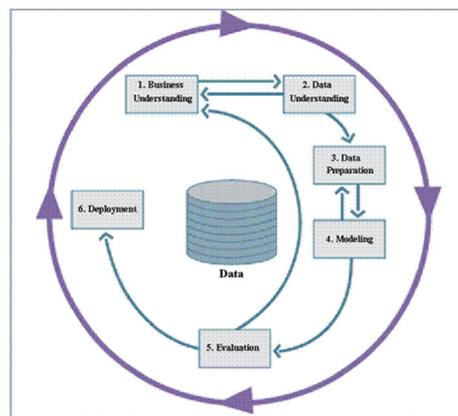
Pada tahap ini akan ditetapkan tujuan dan kebutuhan dari sudut pandang bisnis, kemudian menerjemahkan pengetahuan ini ke dalam pendefinisian masalah dalam *data mining*. Selanjutnya akan ditentukan rencana dan strategi untuk mencapai tujuan tersebut yang akan dibahas pada bagian 3.1

b. Pemahaman Data (*Data Understanding*)

Tahap ini akan dilakukan pengumpulan data yang berpengaruh terhadap produktivitas padi di Karawang, data tersebut didapat dari berbagai sumber seperti dinas pertanian, BMKG, BPS dan lain-lain untuk mendukung penelitian ini.

c. Pengolahan Data (*Data Preparation*)

Pengolahan data dilakukan setelah pemahaman data. Pada tahap ini dilaksanakan secara intensif. Yaitu



Gambar 2. Bagan Alur CRISP-DM

dengan memilih kasus atau atribut-atribut yang akan berpengaruh terhadap penelitian ini dan dianalisis, melakukan perubahan pada beberapa atribut yang sesuai untuk masuk ke tahap selanjutnya.

Penelitian ini menggunakan atribut luas baku sawah, luas tanam, luas panen, rata-rata curah hujan, produksi, OPT dan irigasi. Sedangkan atribut target adalah atribut produktivitas yang akan di rubah yang awalnya bernilai numerik menjadi nominal.

Pengolahan data pada penelitian ini tidak menangani nilai yang hilang secara spesifik. Pada tahap ini penanganan nilai yang hilang hanya mengubah tanda (-) menjadi 0.

d. Pemodelan (*Modeling*)

Memilih teknik pemodelan yang akan dipergunakan yaitu dengan algoritma C4.5 dengan validasi yang dipergunakan adalah *cross-validation*. *Cross validation* dipergunakan karena pada metode ini semua data dipergunakan untuk proses training dan validasi. Dalam penggunaan *cross validation* data harus dipisahkan menjadi k set data. Dalam penelitian ini, menggunakan 10 set (*10-fold cross validation*) data menggunakan keseluruhan data. Sembilan data dipergunakan sebagai training set dan data terakhir dipergunakan untuk set validasi. Prosesnya dilakukan secara berulang sebanyak 10 kali dengan data baru untuk training dan validasi. Kemudian akurasi dihitung menggunakan akurasi rata-rata dari kesepuluh iterasi. Hal ini dilakukan untuk menganalisis kekuatan generalisasi dari model klasifikasi (Kamal *et al.*, 2016). Tools yang dipergunakan dalam penelitian menggunakan WEKA 3.8 sebagai alat bantu yang bisa dipergunakan untuk membantu menyelesaikan permasalahan data mining.

e. Evaluasi (*Evaluation*)

Evaluasi merupakan fase interpretasi terhadap hasil pemodelan *data mining*. Pada fase evaluasi ini akan dihitung hasil prediksi algoritma C4.5 dalam memprediksi produktivitas tanaman padi. Nilai evaluasi ini diukur berdasarkan tingkat *accuracy*, *Kappa Statistic*, *Mean Absolute Error*, dan *Roc Area*.

f. Penyebaran (*Deployment*)

Tahap terakhir dari metodologi CRISP-DM adalah tahap *Deployment*. Pada tahap ini akan dilakukan pembuatan laporan hasil kegiatan yang sudah dilakukan. Laporan akhir mengenai pengetahuan yang didapat tentang hasil analisis yang telah dilakukan dan yang telah dievaluasi.

HASIL DAN PEMBAHASAN

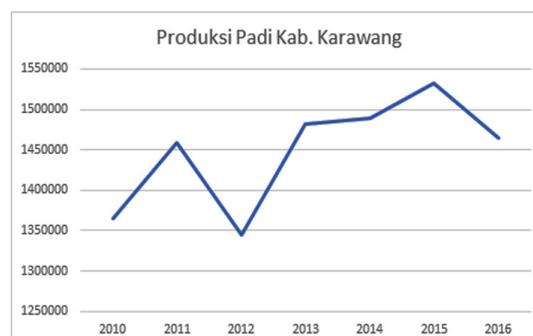
Pemahaman Bisnis (*Business Understanding*)

Tahun 2010-2015 BPS mencatat bahwa wilayah di provinsi Jawa Barat yang memiliki produksi padi tertinggi adalah Kabupaten Indramayu di posisi pertama dan Kabupaten Karawang di posisi kedua dengan tercatat rata-rata produksi padi di Kabupaten Indramayu sebesar 1.373.594 ton per tahun dan untuk Kabupaten Karawang sebesar 1.130.722 ton per tahun.

Hasil produksi padi di Karawang yang tercatat oleh dinas pertanian Karawang mengalami kenaikan dan penurunan yang terjadi setiap tahunnya seperti pada gambar 3. Pada tahun 2012 produksi padi mengalami penurunan sebesar 8% tetapi pada tahun setelahnya produksi padi di Kabupaten Karawang mengalami peningkatan.

Selain dikenal dengan sebutan lumbung padi, saat ini Karawang juga dikenal dengan sebutan kota industri, hal ini seperti menjadi dua mata uang yang saling berseberangan. Karena beberapa lahan sawah produktif lama-lama akan berubah fungsi menjadi lahan penunjang industri, seperti perumahan, sekolah, dan lain-lain. Sehingga hal ini dapat menyebabkan hilangnya lahan sawah produktif dan mempengaruhi hasil produksi padi di Kabupaten Karawang.

Ketakutan ini menyebabkan penelitian ini dilakukan yaitu ingin mengetahui faktor-faktor yang dapat mempengaruhi produktivitas padi selain dari luas lahan dan produksi padi. Produksi padi dipengaruhi banyak faktor di antaranya adalah musim, organisme pengganggu tanaman dan alat produksi yang ada di wilayah tersebut.



Gambar 3. Produksi Padi Kab. Karawang (*sumber: Dinas Pertanian Kab. Karawang*)

Tujuan *data mining* dalam penelitian ini adalah pemanfaatan algoritma C4.5 dalam memprediksi produktivitas padi di Kabupaten Karawang dan mengetahui faktor-faktor yang mempengaruhi produktivitas padi.

Pemahaman Data (*Data Understanding*)

Pemahaman data dilakukan untuk mengumpulkan data yang diperlukan dengan melakukan observasi dan wawancara ke dinas Pertanian Kabupaten Karawang, ke beberapa website yang mendukung dari penelitian ini yaitu: website bps.go.id, jabar.bps.go.id, Karawang.bps.go.id, bmkgo.go.id dan katam.litbang.pertanian.go.id. data yang didapat secara lengkap hanya berasal dari tahun 2010-2015 sehingga pada penelitian ini hanya menggunakan data pada tahun 2010-2015 saja.

Kegiatan selanjutnya pada tahap *data understanding* adalah eksplorasi data yang telah didapatkan. Eksplorasi data dapat dilakukan salah satunya dengan mencari nilai terendah, tertinggi dan *mean* terlihat pada tabel 1. Eksplorasi data berfungsi untuk mengetahui karakteristik data yang akan digunakan.

Persiapan Data (*Data Preparation*)

Pada tahap ini, akan dilakukan pemilihan atribut dan pembersihan data. Data yang akan digunakan adalah data tahun 2010 sampai dengan 2015. Data yang telah didapatkan pada tahap data understanding tidak semua digunakan, melainkan hanya data yang berhubungan dengan produktivitas padi sawah. Atribut yang akan dipilih adalah luas baku sawah, luas tanam, luas panen, rata-rata curah hujan, produksi padi sawah, OPT dan irigasi.

Tabel 1. Karakteristik Data

Atribut	Min	Max	Mean	Satuan	Keterangan
Luas baku sawah	583	6480	97724,175	Ha	Luas baku sawah terendah berada di Kecamatan Ciampel pada tahun 2013 dan 2014. Sedangkan luas baku sawah tertinggi berada di Kecamatan Tempuran pada tahun 2013, 2014, dan 2015.
Luas tanam	0	5658	31167,17	Ha	Luas tanam terendah berada di Kecamatan Batujaya, Tirtajaya, Pedes, Cilebar, Cibuyaya dan tempuran pada tahun 2015. Sedangkan luas tanam tertinggi berada di Kecamatan Tirtajaya pada tahun 2010.
Luas panen	0	5634	29080,29	Ha	Luas panen terendah berada di Kecamatan Cikampek dan Ciampel pada tahun 2014. Sedangkan luas panen tertinggi berada di Kecamatan Tempuran pada tahun 2010.
Produksi	0	245878	225363,4	Ton	Produksi terendah terjadi pada tahun 2016. Sedangkan produksi tertinggi berada di Kecamatan Tempuran pada tahun 2012.
Rata-rata Curah Hujan	0	1483	5357,546	Mm	Curah hujan terendah terjadi pada tahun 2011. Sedangkan curah hujan tertinggi berada di Kecamatan Batujaya pada tahun 2014.
Produktivitas	0,008	716,667	262,5214	Ton/ Ha	Produktivitas terendah berada di Kecamatan Banyusari pada tahun 2012. Sedangkan produktivitas Tertinggi berada di Kecamatan Karawang Barat pada tahun 2013.
OPT	1	538	1619,383	-	Serangan OPT Terendah terjadi di beberapa Kecamatan di antaranya Telukjambe Timur dan Teluk jambe Barat pada tahun 2011, Purwasari pada tahun 2013, serta Kecamatan Majalaya dan Tempura pada tahun 2016. Sedangkan serangan OPT tertinggi terjadi pada tahun 2012 di Kecamatan Cilebar.
Pengairan/ irigasi	477	6480	3271	Ha	Pengairan/irigasi merupakan luasnya persawahan yang dilakukan pengairan baik pengairan irigasi ataupun tadah hujan/non-irigasi.

Hasil verifikasi pada tahap *data understanding* menunjukkan adanya data dengan kualitas yang tidak baik. Kualitas data yang kurang baik dapat mempengaruhi hasil yang akan diperoleh pada tahap pemodelan sehingga diperlukan adanya perbaikan kualitas data. Salah satu kualitas data yang tidak baik terdapat pada atribut curah hujan. Pada data curah hujan terdapat kolom kosong dan kolom yang diisi dengan tanda (-). Karena itu untuk kolom berisi data yang berisi tanda (-) akan diganti menjadi 0 karena berdasarkan hasil wawancara, data dengan tanda (-) memiliki arti bahwa tidak ada kejadian.

Transformasi data yang dilakukan adalah dengan mencari total nilai dalam setiap kecamatan untuk atribut curah hujan, produksi, luas tanam, luas panen dan produktivitas. Salah satu contoh total nilai tiap kecamatan dapat dilihat pada pada Gambar 4.

Tahap *construct data* dilakukan untuk atribut OPT, dalam pengamatan atribut sebelumnya terdapat 6 (enam) OPT yang diamati yaitu pengerek batang, tikus, wereng batang coklat, siput murbai, hama putih palsu dan blasit. *Construct data* dilakukan untuk membuat atribut OPT menjadi satu kesatuan yaitu dengan menggabungkan keseluruhan OPT yang diamati untuk mengetahui derajat serangan. Atribut OPT memuat tentang persentase derajat serangan dengan menggunakan penilaian kerusakan tanaman dalam persen. Pada serangan mutlak angka persen intensitas kerusakan dihitung dengan menggunakan persamaan 4.

$$i = \left(\frac{a}{b}\right) * 100\% \quad (4)$$

Keterangan :

i : Intensitas kerusakan (%)

a : Luas serangan yang terserang hama

b : Luas lahan yang diamati

KECAMATAN	Curah Hujan 2010											Jumlah	
	0.29	-1	-0.24	-0.13	0.38	1	-0.53	1	1	-0.07	-0.99		-0.74
KARAWANG BARAT	0.29	-1	-0.24	-0.13	0.38	1	-0.53	1	1	-0.07	-0.99	-0.74	-0.03
KARAWANG TIMUR	0.29	-1	-0.24	-0.13	0.38	1	-0.53	1	1	-0.07	-0.99	-0.74	-0.03
MAJALAYA	-0.66	-1	-0.78	1	-0.56	-0.67	0.84	-0.15	0.84	0.73	-1	0.99	-0.42
KLARI	-1	-0.62	0.37	-0.08	0.18	-0.86	0.27	0.19	0.48	-0.86	0.96	-0.88	-1.85
TELUKJAMBE BARAT	-0.3	-0.17	0.2	0.58	0.25	0.85	0.6	0.81	-0.04	0.48	-0.67	0.88	3.47
TELUKJAMBE TIMUR	-0.3	-0.17	0.2	0.58	0.25	0.85	0.6	0.81	0.06	0.56	-0.67	1	3.77
CIAMPEL	-0.85	-0.62	0.37	0.49	-0.98	0.04	0.27	0.19	0.9	-1	-0.71	0.79	-1.11
PANGKALAN	0.97	0.5	-0.62	-1	0.22	0.81	0.04	-0.61	0.93	1	0.98	0.07	3.29
TEGALWARU	0.97	0.5	-0.62	-1	0.22	0.81	0.04	-0.61	0.93	1	0.98	0.07	3.29
RENGASDENGKLOK	0.96	-0.73	-0.96	0.41	-0.84	-0.4	-0.59	-0.84	0.96	-0.97	0.73	-0.16	-2.43
JAYAKARTA	0.12	0.93	0.99	-0.96	-0.58	0.87	-0.47	0.44	0.58	0.9	-0.95	0.81	2.68
KUTAWALUYA	0.96	-0.73	-0.96	0.41	-0.84	-0.4	-0.59	-0.84	0.96	-0.97	0.73	-0.16	-2.43
BATUJAYA	0.12	0.93	0.99	-0.96	-0.58	0.87	-0.47	0.44	0.58	0.9	-0.95	0.81	2.68
TIRTAJAYA	0.12	0.93	0.99	-0.96	-0.58	0.87	-0.58	0.44	0.58	0.9	-0.95	0.81	2.57
PAKISJAYA	0.94	0.27	0.87	-0.47	-0.26	-0.24	0.61	0.53	-0.65	0.05	-0.66	0.92	1.91

Gambar 4. Construct Data

Tabel 2. Klasifikasi Produktivitas (Pertanian, 2019)

Kalsifikasi	Range
Rendah	<5 ton/ha
Sedang	5-6 ton/ha
Tinggi	>6 ton/ha

Untuk menentukan derajat serangan pada penelitian ini, maka atribut OPT akan dijumlahkan keseluruhan datanya dan dibagi dengan jumlah luas tanam dikalikan 100% maka didapat persen derajat serangan.

Format *dataset* akhir berupa tabel atau query (*flat file*) untuk pemodelan *data mining*. Format tabel yang berupa *excel* akan diganti menjadi format *csv* agar *dataset* siap untuk diproses menggunakan Weka 3.8.

Class target yang dapat diterima oleh algoritma C4.5 mengharuskan *output* yang dapat ditangani adalah bertipe data nominal, sehingga *class target* pada penelitian ini adalah *class* produktivitas harus dilakukan format data dengan mengubah dari bertipe data numerik menjadi bertipe data nominal dengan ketentuan seperti pada tabel 2 berikut:

Pemodelan (Modelling)

Pemodelan dilakukan dengan algoritman C4.5 dengan nilai *confidence factor* 0,5 untuk menyelesaikan masalah pada penelitian ini dengan alat bantu *software* WEKA 3.8. Pemodelan yang dilakukan dengan menggunakan *test option* 10-folds *cross validation*. Hasil yang didapat setelah dilakukan proses pemodelan terlihat pada gambar 5 dan gambar 6. Gambar 5 merupakan hasil pemodelan algoritma C4.5 dengan nilai *confidence factor* 0,5 yang terdiri dari 180 data, 8 atribut yang dipakai 7 atribut *input* dan 1 atribut *output* dengan waktu komputasi adalah 0,06 detik. Sedangkan gambar 6 merupakan hasil pohon yang terbentuk dari algoritma C4.5 dengan 18 jumlah daun (*number of leaves*) dan 35 ukuran pohon (*Size of The Tree*).

Evaluasi (Evaluation)

Evaluasi merupakan fase interpretasi terhadap hasil pemodelan *data mining*. Pada fase evaluasi ini akan dilakukan perhitungan mengenai hasil prediksi algoritma C4.5 dalam memprediksi produktivitas. Perbandingan ini diukur berdasarkan tingkat Akurasi, *kappa statistik*, *Mean absolute error* dan *Roc Area*, yang terlihat pada tabel 3. nilai akurasi didapat dari 180 data, dengan 119 data diprediksi benar dan 61 data diprediksi salah.

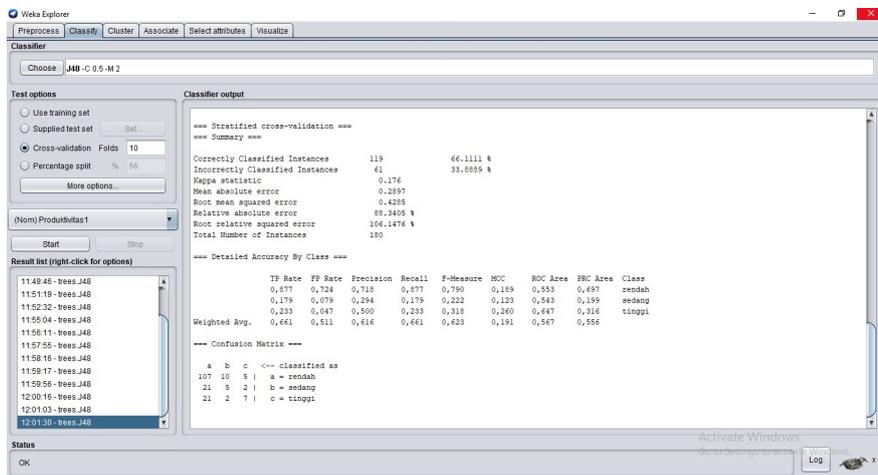
Penyebaran (Deployment)

Tahap terakhir dari metodologi CRISP-DM adalah tahap *Deployment*. Pada tahap ini akan dilakukan pembuatan laporan hasil kegiatan yang sudah dilakukan.

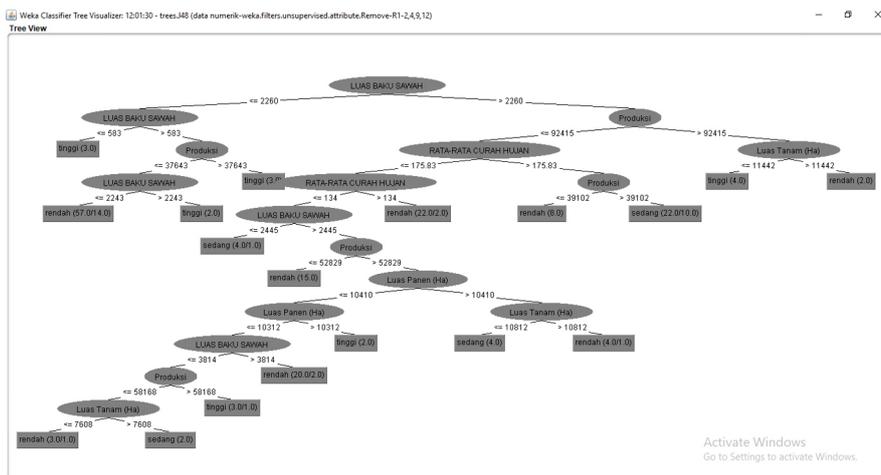
Laporan akhir mengenai pengetahuan yang didapat tentang hasil analisis yang telah dilakukan dan yang telah dievaluasi.

Hasil penelitian yang sudah dilakukan didapat bahwa untuk memprediksi produktivitas tanaman padi di Karawang menggunakan algoritma C4.5 dengan memanfaatkan atribut irigasi, OPT, Luas tanam, Luas Panen, Produksi, luas baku sawah dan rata-rata curah hujan memperlihatkan hasil yang tidak terlalu bagus, karena nilai akurasi yang dihasilkan hanya sebesar 66,1%

Pohon keputusan yang dihasilkan dari penelitian ini terlihat pada gambar 6, memperlihatkan bahwa untuk memprediksi produktivitas tanaman padi terdapat beberapa faktor penentu yaitu atribut Luas Baku sawah, Produksi, Rata-rata Curah Hujan, Luas Tanam, dan Luas Panen.



Gambar 5. Hasil Pemodelan Algoritma C4.5



Gambar 6. Pohon C4.5

Tabel 3. Evaluasi

	Nilai
Akurasi	66,11%
Kappa Statistic	0,17
Mean Absolute Error	0,28
ROC Area	0,56

Dalam penelitian ini terdapat dua buah tujuan yang ingin dicapai yaitu tujuan *data mining* dan tujuan bisnis yang telah ditentukan ditahap pertama. Dari tujuan data mining yang ingin dicapai yaitu menerapkan algoritma C4.5 dalam memprediksi produktivitas padi dari penelitian ini diketahui bahwa algoritma C4.5 dapat dipergunakan untuk melakukan prediksi produktivitas padi. Tetapi hasil yang didapatkan dalam penelitian ini tidak terlalu bagus karena nilai akurasi hanya sebesar 66,1%. Hal ini kemungkinan disebabkan karena atribut yang dipergunakan tidak terlalu sesuai untuk memprediksi produktivitas padi, maka perlu di cari kembali atribut-atribut yang berpengaruh terhadap produktivitas padi.

Tujuan bisnis yang dicapai dalam penelitian ini adalah mengetahui faktor-faktor yang dapat mempengaruhi nilai produktivitas padi selain nilai produksi dan luas tanam. Dari penelitian ini didapat bahwa faktor Luas Baku sawah, rata-rata Curah Hujan, dan Luas Panen adalah faktor-faktor penentu dalam mempengaruhi produktivitas padi.

KESIMPULAN

Memprediksi produktivitas tanaman padi menggunakan algoritma C4.5 dapat dilakukan, tetapi hasil dari prediksi dalam percobaan ini tidak terlalu bagus terlihat pada hasil evaluasi dengan melihat nilai akurasi hanya 66,1%, nilai *Kappa statistic* yang masuk ke dalam kategori jelek dengan rentang nilai di antara 0 - 0,12 dan nilai ROC Area 0,56 yang masuk ke dalam kategori gagal. Hal ini disebabkan karena atribut masukan yang diamati tidak terlalu berpengaruh terhadap produksi produktivitas tanaman padi. Selain nilai produksi dan luas tanam untuk menentukan nilai produktivitas tanaman padi, ternyata produktivitas tanaman padi juga dipengaruhi oleh faktor Luas Baku sawah, rata-rata Curah Hujan, dan Luas Panen.

UCAPAN TERIMA KASIH

Ucapan terima kasih kami berikan kepada Lembaga Penelitian dan Pengabdian Kepada Masyarakat Universitas Singaperbangsa Karawang yang sudah menyokong penelitian ini. Dan dinas-dinas terkait dalam penelitian ini.

DAFTAR PUSTAKA

Apriyana, Y. dan Lindawati, 2015. Aplikasi model prediksi curah hujan pada dua sentra produksi padi di Jawa Barat. *Informatika Pertanian* 24(2) : 149–156.

- Dhika, H. 2015. Kajian Komparasi Penerapan Algoritma C4.5, Naïve Bayes, dan Neural Network dalam Pemilihan Mitra Kerja Penyedia Jasa Transportasi : Studi Kasus CV. Viradi Global Pratama. In Seminar Nasional Inovasi dan Tren (SNIT). Hlm. 197–202.
- FAO. 2018. Rice Market Monitor (RMM). Food and Agriculture Organization of the United Nations 21 (1) : 1–35. Available from: <http://www.fao.org/3/I9243ES/i9243es.pdf>
- Gultom, L., R. Winandi, dan S. Jahroh. 2014. Analisis efisiensi teknis usahatani padi semi organik di kecamatan cigombong, bogor. *Informatika Pertanian* 23 (1) : 7–18.
- Ikhwan. 2014. Teknologi budidaya varietas unggul baru padi sawah pada dua musim tanam. *Informatika Pertanian* 23 (1) : 19–28.
- Ina, H. 2007. Bercocok Tanaman Padi. Azka Mulia Media.
- Kamal, N. D. M., N. Jalil, and H. Hashim. 2016. The Analysis of Shape-based, DWT and Zernike Moments Feature Extraction Techniques for Fasterner Recognition Using 10-Fold Cross Validation Multilayer Perceptrons. *International Journal of Electrical and Electronic Systems Research*, 9. Available from: <https://doi.org/10.1073/pnas.1100304108/-/DCSupplemental>. www.pnas.org/cgi/doi/10.1073/pnas.1100304108
- Karyono, G. 2016. Analisis teknik data mining “algoritma c4.5 dan k-nearest neighbor” untuk mendiagnosa penyakit diabetes mellitus. In Seminar nasional Teknologi Infomasi, Bisnis dan Desain. Hlm. 77–82.
- Kusnadi, N., N. Tinaprilla, S. H. Susilowati, dan A. Purwoto. 2011. Rice farming efficiency analysis in some rice producing areas in Indonesia. *Jurnal Agro Ekonomi* 29(1) : 25–48.
- Patel, N., and S. Upadhyay. 2012. Study of various decision tree pruning methods with their empirical comparison in WEKA. *International Journal of Computer Applications* 60 (12) : 20–25.
- Pertanian, B. P. dan P. 2019. Penyusunan Rekomendasi Pupuk.
- Rohmani, S. A., E. Rustiadi, M. Firdaus, dan T. Sudaryanto. 2015. Dampak modal sosial dalam pengelolaan irigasi terhadap kesejahteraan petani di kabupaten sukoharjo Jawa Tengah. *Informatika Pertanian* 24 (1) : 67–90.
- Rokach, L., O. Maimom and O. Maimon. 2015. Data Mining With Decision Trees - Theory and Applications. *Data Mining and Knowledge Discovery*. Available from: <https://doi.org/10.1142/9097>
- Wirth, R. 2000. CRISP-DM : Towards a Standard Process Model for Data Mining. In *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*. Hlm. 29–39. Available from: <https://doi.org/10.1.1.198.5133>