

Sistem Prediksi Produktifitas Pertanian Padi Menggunakan Data Mining

Siti Maesaroh¹⁾, Kusrini²⁾

^{1,2)} Magister Teknik Informatika, Universitas Amikom Yogyakarta
Jl. Ring Road Utara Condong Catur, Sleman, Yogyakarta

Terima Naskah : 9 November 2017
Terima Revisi : 13 November 2017

ABSTRAK

Penelitian ini melakukan analisis terhadap peningkatan produktivitas padi. Beberapa faktor memiliki peran penting dalam peningkatan dan penurunan produktivitas. Faktor tersebut adalah cuaca, kesedian pupuk, kondisi tanah. Faktor tersebut tidak dapat diprediksi, terutama perubahan cuaca. Beberapa petani gagal memprediksi dan beradaptasi terhadap perubahan faktor yang mengakibatkan penurunan produktivitas. Pembangunan Sistem Prediksi menggunakan pendekatan Data Mining. Prediksi akan melakukan pengolahan data data pendukung dalam peningkatan produktifitas pertanian. Pembahasan berfokus pada tiga hal: faktor-faktor yang mempengaruhi produktivitas padi, klasifikasi untuk prediksi dan metode klasifikasi. Kesimpulan dari penelitian ini adalah perbandingan tiga metode klasifikasi untuk mendapatkan akurasi sistem prediksi. Sistem Prediksi telah diuji dengan menggunakan K Fold Cross Validation. Tiga metode memiliki kesalahan prediksi di bawah 30% menggunakan 100 data training. Kata kunci : Data Mining, Produktivitas Padi, Klasifikasi

ABSTRACT

The study is analyzing the rice productivity improvement. Some factors have important role of increasing and decreasing productivity. There are climate, fertilizer stock, sand condition. These factor nowadays very unpredictable, particularly climate change. Some farmers failed to predict and adapt for factors change. Productivity will decreased. The discussion conducted by Prediction System using Data Mining approach. This concept emphasizes prediction system for rice production are discussed. The discussion will focus on three things: factors that impact rice productivity, classification for prediction and classification method testing. The conclusion of this study is a comparison of three classification method in order to obtain the accuratte prediction system. Prediction System has been tested using K Fold Cross Validation. Three methods have prediction error below 30% using 100 data training.
Keywords : Data Mining, Rice Productivity, Classification

PENDAHULUAN

Pertanian adalah salah satu sumber ekonomi bangsa. Banyak penduduk Indonesia yang berprofesi sebagai petani. Untuk dapat meningkatkan kesejahteraan petani adalah dengan meningkatkan produktivitas lahan pertanian. Produktivitas pertanian pada dasarnya adalah bagaimana meningkatkan hasil dengan kualitas maupun kuantitas maksimal dan menggunakan pupuk, obat minimal dan memanfaatkan keterbatasan sumber daya alam (air, energi dan kondisi tanah). Keterbatasan pupuk dan obat menjadikan harga yang tinggi. Kondisi cuaca yang tidak menentu juga menjadikan ketersediaan air

cukup mengganggu pertumbuhan padi. Dan masih banyak faktor lain yang mengakibatkan penurunan produktivitas bahkan sampai ke gagal panen.

Perubahan dan perkembangan yang terjadi saat ini mengakibatkan peningkatan kompleksitas data dan informasi. Kenyataan ini menjadikan semakin sulit dan rumit dalam membangun sebuah prediksi ke depan. Secara umum bercocok tanam padi mencakup persemaian, pemindahan atau penanaman, pemeliharaan (termasuk pengairan, penyiangan, perlindungan tanaman, serta pemupukan) dan panen. Masih banyak faktor yang mempengaruhi. Dari sisi bibit sudah membutuhkan pengetahuan dan pertimbangan banyak hal [1] Ketersediaan air irigasi juga memberikan banyak

pengaruh, pengetahuan tentang ini juga membutuhkan pengelolaan khusus [2] Semua tahapan bercocok tanam padi adalah sebuah kesinambungan. Kompleksitas semakin tinggi dengan tingkat ketidakpastian yang sering muncul. Sangat dibutuhkan sebuah konsep pendukung keputusan yang mampu mengolah data dengan karakteristik tersebut menjadi sebuah pengetahuan yang akurat dan *valid*. Pengetahuan tersebut dapat menjadi acuan bagi petani dalam melakukan cocok tanam dengan baik dan pada akhirnya meningkatkan produktivitas.

Dari penelitian yang sudah dilakukan dapat dilihat fokus penelitian terbagi menjadi:

- a. Pengelolaan penanaman padi mengacu kepada kondisi lingkungan seperti cuaca, ketersediaan air dan kondisi tanah. Penelitian dilakukan di bidang pertanian dan lingkungan.
- b. Penelitian *decision support system* dan *data mining* untuk melakukan prediksi hujan terkait dengan pertanian.
- c. Penelitian di bidang teknologi informasi terkait penelitian yaitu di bidang citra digital tentang *Normalized Difference Vegetation Index* (NDVI).

Di sisi lain, pertanian saat ini membutuhkan pengelolaan lebih seksama, mengingat banyak faktor yang mempengaruhi keberhasilan sebuah proses bercocok tanam. Faktor tersebut adalah kondisi lingkungan/ cuaca, ketersediaan benih/bibit, ketersediaan obat dan pupuk dan kondisi tanah. Keempat faktor tersebut menjadi acuan dalam menentukan waktu tanam untuk mendapatkan hasil yang maksimal.

Kalender tanam yang ada saat ini dan dikeluarkan oleh Departemen Pertanian hanya fokus pada faktor cuaca sehingga akurasi dalam memberikan waktu tanam yang cocok kurang maksimal. Perlu untuk melakukan penguatan kalender tanam tersebut dengan memasukkan ketiga faktor lainnya. Dengan melihat kondisi tersebut akan dilakukan penelitian untuk membuat sebuah prediksi kalender tanam, dengan menambah acuan ketiga faktor di atas.

Sistem prediksi yang dibangun menggunakan metode *naive bayes*. Pemilihan metode berdasarkan sifat sifat dari *naive bayes* sebagai *classifier* yaitu [3]

- a. Mampu mengisolir data yang tidak perlu atau tidak relevan.

- b. Metode yang memiliki inti dari perhitungan probabilitas
- c. Memiliki kecepatan proses dan akurasi tinggi untuk *database* dengan ukuran besar.
- d. Mampu menangani atribut yang banyak dan tidak saling bergantung (*interdependency*).

Produksi Padi

Produksi padi merupakan salah satu hasil bercocok tanam yang dilakukan dengan penanaman bibit padi dan perawatan serta pemupukan secara teratur sehingga menghasilkan suatu produksi padi yang dapat dimanfaatkan. Padi tersebut kemudian diproses menjadi beras, yang mana beras itu sendiri akan diolah menjadi nasi. Nasi merupakan sumber kalori utama yang banyak mengandung unsur karbohidrat yang sangat tinggi sehingga sangat bermanfaat dan menjadikan sebagai bahan pangan utama.[4]

Data Mining

Konsep data mining adalah sebuah proses melakukan pengolahan data histori kejadian kejadian terdahulu dan dipergunakan sebagai dasar untuk membangun sebuah pengetahuan. Dalam kasus penelitian ini pengetahuan yang dibangkitkan adalah sebuah prediksi dari sebuah kasus. [5]

Dalam membangun sebuah data mining diperlukan Kumpulan data Training yang menjadi dasar acuan prediksi, kemudian metoda *classification* ditentukan sesuai dengan karakteristik data training. Data training dapat dikategorikan menjadi Katagori dan Kontinu. Katagori memiliki data dengan nilai diskrit atau level seperti Kurang, Cukup dan Sedang. Untuk Kontinu data training berbentuk nilai atau angka kontinyu.

Untuk mennangai data training dengan bentuk Katagori dapat dipergunakan Meroda Decision Tree atau Naive Bayes. Sedangkan untuk data training dengan bentuk kontinyu dapat dipepgurnkan K Nearest Neighbor. Namun ketentuan ini tidak mutlak, data training bebrbentuk kontnyu dapat juga mempergunakan Naive Bayes, tetapi memerlukan proses katagorisasi data trainingnya. Demikian pula sebaliknya.

Pemilihan metoda klasifikasi yang tepat merupakan salah satu modal keberhasilan data mining yang dibangun.

Metoda Klasifikasi

1. K Nearest Neighbor

Metoda klasifikasi ini menggunakan konsep perhitungan jarak terdekat dengan sebuah titik. Proses prediksi yang dilakukan adalah menentukan ukuran jarak dalam angka terhadap sebuah titik yang menjadi identikan dari grup atau kelas. Penentuan jarak menggunakan ruang Euclidian.

2. Decision Tree (Algoritma C45)

Metoda klasifikasi ini adalah membangun sebuah pohon keputusan dimulai dari *node root* sampai ke anaknya. Penurunan pohon keputusan menggunakan konsep pemecahan sampel data. Pemisahan menggunakan perhitungan entropi atau formula logaritma. Semua node kecuali terbawah (*leaf node*) didasarkan pada himpunan dari kondisi atribut. Node terbawah atau *leaf* adalah atribut target. Proses prediksi dilakukan dengan memberikan masukan ke atribut secara berurutan dari *Root* sampai *leaf* melalui *path*/jalur sesuai kondisi.

3. Naïve Bayes

Metoda *naïve bayes* didasarkan pada perhitungan peluang atau probabilitas setiap kejadian. Metoda ini dapat menangani sejumlah variabel masukan yang independen atau tidak memiliki ketergantungan. Prediksi dilakukan dengan menghitung peluang sebuah kejadian berdasarkan kumpulan kejadian kejadian yang sudah terekam. [3][6]

Perbandingan ketiga metode tersebut ditunjukkan pada tabel 1:

Tabel 1 Perbandingan Metode

Karakteristik	<i>K Nearest Neighbor</i>	<i>Decision Tree</i>	<i>Naïve Bayes</i>
Data Training	Kontinyu	Kategori	Kategori
Konsep	Jarak <i>Euclidian Space</i>	Entropi	Probabilistik
Hasil Akhir	Rumus Jarak Terdekat	Urutan <i>Rule (if then else)</i>	Perhitungan Peluang

Sistem Prediksi

Sistem prediksi adalah hasil yang diperoleh dari Data Mining setelah dilakukan proses *training* atau klasifikasi. Melalui metoda K NN akan diperoleh klasifikasi berdasarkan jarak terdekat di ruang *Euclidian*. Dengan *Decision Tree* akan diperoleh sebuah *Rule* dan dengan *Naive Bayes* akan diperoleh perhitungan probabilistik atau peluang.

K-NN akan melakukan proses pengelompokan terhadap kondisi atau data baru. Pengelompokan atau klasifikasi tersebut adalah proses prediksi. *Decision Tree* akan memberikan jawaban berdasarkan tahapan *Rule* yang dihasilkan untuk data baru. Hasil jawaban *Rule* tersebut adalah prediksi. Sedangkan Metoda *Naive Bayes* akan menghitung peluang data baru

METODE

Metode Pencarian Data

Produktivitas padi adalah sebuah ukuran dari keberhasilan sebuah proses produksi pertanian padi. Ukuran yang dilakukan adalah dalam satu luas atau bidang sawah membutuhkan modal bibit, pupuk, pengobatan dan tenaga kerja dapat menghasilkan padi dalam ukuran berat (kg). Selanjutnya nilai nilai tersebut dilakukan konversi terhadap nilai rupiah.

Dalam penelitian ini dilakukan pengamatan di wilayah Kabupaten Tasikmalaya. Data penelitian dikumpulkan dalam 2 tahun terakhir. Pertanian yang dimaksud adalah penanaman padi. Data didapatkan dari Kantor Dinas Pertanian Kabupaten Tasikmalaya dan Kantor Biro Pusat Statistik Kabupaten Tasikmalaya.

Analisis Data Mining

Analisis data mining berupa kegiatan mengamati data yang didapatkan serta pola produksi padi. Dari pola produksi padi dapat ditentukan faktor - faktor yang mempengaruhi produktivitas tanaman padi. [7]

Faktor yang mempengaruhi adalah:

1. Kondisi Cuaca atau Iklim berupa curah hujan (musim kering, musim hujan dan lainnya). [8][9]
2. Kondisi Tanah (Tingkat Kekersan tanah, pH dan kandungan nutrisi tanah)[10][11]
3. Ketersediaan Bibit
4. Ketersediaan Pupuk dan Obat

Keempat faktor tersebut selanjutnya akan dilakukan penilaian dan membangun data *training*.

a. Atribut

Atribut adalah data data yang dikelompokkan berdasarkan faktor yang mempengaruhi dan berkorelasi dengan keberhasilan produksi padi. Terdapat dua macam atribut yaitu Data Atribut yang diturunkan dari faktor faktor yang berpengaruh dan Target Atribut yaitu hasil akhir dari produksi padi atau produktivitas padi. Dalam penelitian ini data atribut adalah cuaca, tanah, benih dan obat, sedangkan target atribut adalah produktivitas padi [12]

b. Data *Training*

Data *training* adalah kumpulan data hasil *survey* terkait dengan produksi padi. *Survey* dilakukan di area pertanian padi di wilayah Kabupaten Tasikmalaya. Data yang diambil adalah kondisi empat faktor di atas dan korelasinya dengan target atribut, dalam hal ini keberhasilan produksi padi. [13] [14] [15] Data hasil *survey* lapangan dilakukan analisis dan dirangkum di tabel 2

Tabel 2 Hasil Survey Pertanian

Atribut	Instances	Keterangan
Cuaca	Kurang Cukup Baik	0-100 mm 100 – 300 mm 300 – 500 mm
Tanah	Kurang Cukup Tinggi	pH <5,5 5,5 <pH < 6,5 pH > 6,5
Bibit	Tidak Tersedia Tersedia	Bibit tersedia sesuai waktu, jenis dan harga
Pupuk	Tidak Tersedia Tersedia	Pupuk tersedia sesuai waktu, jenis dan harga
Hasil Panen	Berhasil Tidak Berhasil	GKP > 7 ton/Ha GKP < 7 ton/Ha

Data *training* yang diperoleh dapat diringkas dalam tabel 3 (ditampilkan 3 instances dari 100):

Tabel 3 Data *Training*

CUACA	TANAH	BIBIT	PUPUK	HASIL
Cukup	Kurang	Tersedia	Tersedia	Berhasil
Kurang	Kurang	Tidak	Tersedia	Tidak
Cukup	Cukup	Tersedia	Tersedia	Berhasil

Tabel 4 Data Dalam Bentuk Angka

CUACA	PH	BIBIT	PUPUK	HASIL
75	4	Tersedia	Tersedia	7,2
90	4,5	Tidak	Tersedia	5,4
200	5,5	Tersedia	Tersedia	7,6

Metoda Klasifikasi

Proses klasifikasi dilakukan menggunakan penurunan formula sesuai metoda yang dipergunakan. [16]

1. K Nearest Neighbor

Metoda ini melakukan perhitungan jarak terdekat terhadap sebuah titik (*centroid*). Perhitungan dilakukan menggunakan ruang *Euclidian* 2 dimensi dengan variabel cuaca (curah hujan dalam mm) dan kondisi tanah (diukur nilai pH atau derajat asam basa tanah). Ketersediaan bibit dan pupuk diproses terpisah.

2. Decision Tree (Algoritma C45)

Metoda ini menurunkan pohon keputusan berdasarkan data *training* yang ada (empat variabel/atribut dan satu target variabel). Perhitungan akan menentukan atribut yang akan menjadi *Root Node* dan *Child Node*.

3. Naïve Bayes

Metoda *naïve bayes* menurunkan perhitungan berdasarkan nilai peluang.

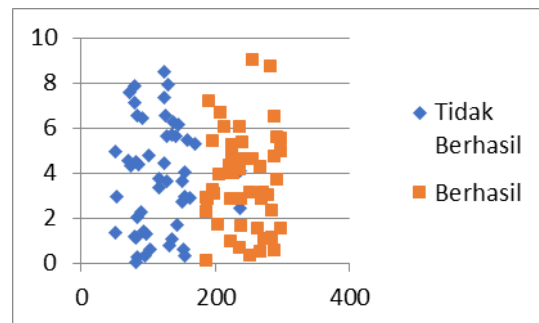
Dari ketiga metoda yang dipergunakan di atas, selanjutnya dilakukan pengujian tingkat kesalahan prediksi dengan menggunakan data data *training* yang ada. Data *training* dijadikan masukan ke *Rule (decision tree)*, *Grup/Cluster (KNN)*, atau *Formula Probabilitas (Naïve Bayes)*. Hasil prediksi

HASIL DAN PEMBAHASAN

Hasil penurunan sistem prediksi dengan ketiga metoda untuk kasus produksi padi adalah sebagai berikut:

1. K Nearest Neighbor

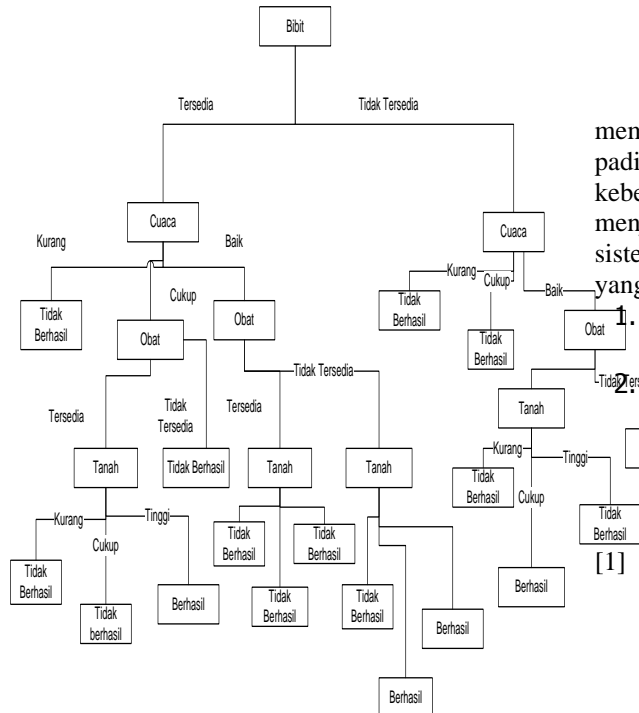
Proses klasifikasi dengan KNN ditunjukkan pada gambar 1.



Gambar 1 Hasil Metode KNN

2. Decision Tree

Dari data training *desicion tree* dapat diturunkan dengan *rule/* kaidah prediksi ditunjukkan pada gambar 2.



Gambar 2 Hasil Metode Decision Tree

Dari hasil penurunan perhitungan untuk ketiga metoda data mining diperoleh system yang mampu melakukan prediksi tingkat keberhasilan produksi padi.

Pengujian akurasi prediksi dilakukan dengan memanfaatkan data training sebagai data uji. Pengujian dilakukan menggunakan K Fold Cross Validation, dengan K=5., Dari pegujian diperoleh tingkat kesalahan prediksi untuk masing masing metoda ditunjukkan di tabel 5:

Tabel 5 Hasil Uji

Metoda	Prosentase Error
K Nearest Neigbor	26 %
Decision Tree	35 %
Naïve Bayes	13 %

Pengujian dilakukan dengan 100 data training. Dari data tersebut dibagi menjadi 5 bagian dengan jumlah sama (masing masing 20 data). Selanjutnya

dilakukan pengujian menggunakan Cross Validation. Rata rata presentasi kesalahan tiap bagian menjadi ukuran tingkat akurasi prediksi masing masing metoda.

SIMPULAN

Tiga metoda prediksi dilakukan untuk membangun sistem prediksi produksi tanaman padi. Metoda melibatkan empat faktor penentu keberhasilan produksi padi. Empat faktor tersebut menjadi atribut sistem prediksi pembangunan sistem prediksi menggunakan 100 data *survey*. Hasil yang diperoleh adalah:

1. Prediksi tanaman padi dapat dibangun menggunakan data mining.
2. Metoda yang dipergunakan memberikan akurasi yang berbeda namun masih diangka yang dapat dipercaya.

DAFTAR PUSTAKA

[1] Astrina Yulianti, Introduction Technology of Cropping Calendar-Information System (CC-IS) for Rice Farming as A Climate Change Adaptation in Indonesia, International Journal of Advanced Science Engineering Information Technologi, Vol 6- 2016

[2] Andrew W. Robertson, Vincent Morona, and Yunus Swarinoto. Seasonal predictability of daily rainfall statistics over Indramayu district, Indonesia International Journal Of Climatology *Int. J. Climatol.* 29: 1449–1462 (2009)

[3] Jiawei Han and Micheline Kamber.2006. Data Mining: Concepts and Techniques, Second Edition. Morgan Kaufmann Publishers is an imprint of Elsevier

[4] Vivi Diannita Sari, Bangun Muljo Sukojo, Analisa Estimasi Produksi Padi Berdasarkan Fase Tumbuh Dan Model Peramalan *Autoregressive Integrated Moving Average (Arima)* Menggunakan Citra Satelit Landsat (Studi Kasus: Kabupaten Bojonegoro), GEOID Vol. 10, No. 02, Februari 2015 (194-203)

- [5] Florin Gorunescu. 2011. Data Mining Concepts, Models and Techniques. Springer-Verlag Berlin Heidelberg
- [6] Ondrej Habala, Ladislav Hluchý, Viet Tran, Peter Krammer, Martin Šeleng., 2012. Using Advanced Data Mining And Integration In Environmental Prediction Scenarios, Computer Science 13 (1)
- [7] ADE CANDRA PRABANDARI Analisis Faktor-faktor yang Mempengaruhi Produksi Padi Sawah pada Daerah Tengah dan Hilir Aliran Sungai Ayung (Studi Kasus Subak Mambal, Kabupaten Badung dan Subak Pagutan, Kota Denpasar), E-Jurnal Agribisnis dan Agrowisata 2013
- [8] Elly Rasmikayati dan Endah Djuwendah, Dampak Perubahan Iklim Terhadap Perilaku Dan Pendapatan Petani (*The Impact of Climate Change to Farmers' Behavior and Revenue*), J. MANUSIA DAN LINGKUNGAN, Vol. 22, No. 3, November 2015: 372-379
- [9] Fadhlullah Ramadhani, SISTEM TEKNOLOGI INFORMASI KALENDER TANAM TERPADU, Informatika Pertanian, Vol. 22 No.2, Desember 2013: 103 – 112
- [10] Didik Srianto, Edy Mulyanto. 2016. Perbandingan K-Nearest Neighbor Dan Naive Bayes Untuk Klasifikasi Tanah Layak Tanam Pohon Jati. Jurnal Techno.COM, Vol. 15, No. 3
- [11] Vaneesbeer Singh. 2017. Analysis of soil and prediction of crop yield (Rice) using Machine Learning approach, International Journal of Advanced Research in Computer Science.
- [12] Witten, I. H. (Ian H.), Eibe Frank. 2006. Data mining : Practical Machine Learning Tools And Techniques. Morgan Kaufmann series – 2nd ed.
- [13] Elza Surmaini dan Haris Syahbuddin. 2016. Kriteria Awal Musim Tanam: Tinjauan Prediksi Waktu Tanam Padi Di Indonesia. Jurnal Penelitian dan Pengembangan Pertanian, Vol 55
- [14] Krzysztof Siwek. 2016. Data Mining Methods For Prediction Of Air Pollution, Int. J. Appl. Math. Comput. Sci. Vol. 26, No. 2, 467–478
- [15] Sudha Mohankumar and Valarmathi Balasubramanian. 2016. Identifying Effective Features and Classifiers for Short Term Rainfall Forecast. Journal of Computing and Information Technology, Vol. 24, No. 2, June 181–194
- [16] Suhaila Zainudin. 2016. Comparative Analysis of Data Mining Techniques for Malaysian Rainfall Prediction. International Journal of Advanced Science Engineering Information Technology, Vol 6