

PENERAPAN K-NEAREST NEIGHBOR BERBASIS ALGORITMA GENETIKA UNTUK KLASIFIKASI MUTU PADI ORGANIK

Wali Ja'far Shudiq

Teknik Informatika, Sekolah Tinggi Teknologi Nurul Jadid
Karanganyar Paiton Probolinggo
email : walijafar@sttnj.ac.id

Abstrak

Penentuan kualitas mutu padi memegang peranan yang sangat penting terhadap terjaminnya ketersediaan padi, sebab di Indonesia beras merupakan bahan makanan pokok dan merupakan sumber kalori bagi sebagian besar penduduk, sehingga stabilitas pertumbuhan tanaman padi juga akan berpengaruh terhadap ketahanan pangan, untuk mengetahui jenis mutu padi yang berkualitas harus dilihat dari beberapa aspek yang ditelusuri dari karakteristiknya seperti, mutu, warna, variatas, panjang, bentuk, rasa, teknik, musim, ama, dan PH. Untuk itu dalam makalah ini akan menjelaskan bagaimana menghasilkan mutu padi yang akurat dan maka dilakukan sebuah penelitian dalam mencari pola dari mutu padi dengan menggunakan metode k-nearest neighbor dan algoritma genetika dengan proses k-fold cross validation dan confusion matrik bertujuan agar pengklasifikasian tidak lagi sekedar hanya menggunakan perkiraan semata tapi menggunakan data dari pengalaman expert yang sudah di ekstrak dan menjadi acuan utama dalam menentukan klasifikasi mutu dari sebuah padi. Proses yang dilakukan menghitung dari data testig dan training dengan k-fold validation dan menghasilkan akurasi yang sangat signifikan, yang semula dari proses pengolahan data dengan K-NN menghasilkan 81.60% kemudian lebih meningkat ketika pengolahan data kombinasi K-NN dan Algoritma Genetika yaitu 96,40% untuk dataset 4952.

Kata kunci: Mutu padi, K-Nearest Neighbor, Algoritma Genetika, Confusion Matrik

1. PENDAHULUAN

Sistem perekonomian di negara Indonesia sangat meningkat sehingga pertanian memiliki peran yang sangat penting, mengikuti permintaan konsumen yang sangat besar sehingga menjadi penghasil bahan baku makanan yang bervariasi, seperti penghasil bahan baku makanan dan penghasil non pertanian, Indonesia juga dalam persaingan global yang makin liberal, karena sebagai sumber devisa, sebagai sumber investasi, dan sebagai sumber pemasok tenaga kerja yang setiap tahun bisa meningkat.

Ketahanan pangan merupakan salah satu faktor penentu dalam stabilitas nasional suatu negara, baik di bidang ekonomi, keamanan, politik dan sosial. Oleh sebab itu, ketahanan pangan merupakan program utama dalam pembangunan pertanian saat ini dan masa mendatang.

Ketahanan pangan sendiri menurut literatur memiliki 5 unsur yang harus dipenuhi :

1. Berorientasi pada rumah tangga dan individu,
2. Dimensi waktu setiap saat pangan tersedia dan dapat diakses,
3. Menekankan pada akses pangan rumah tangga dan individu, baik fisik, ekonomi dan sosial,
4. Berorientasi pada pemenuhan gizi,
5. Ditujukan untuk hidup sehat dan produktif.

Salah satu target yang akan dicapai kementerian pertanian dalam rangka mewujudkan ketahanan pangan adalah dengan melakukan swasembada beras.

Kualitas mutu padi adalah yang diutamakan oleh peneliti sehingga jenis klasifikasi dalam data mining yang digunakan untuk mengklasifikasi mutu padi menggunakan metode atau algoritma yang ada di *data mining*, sementara klasifikasi data mining memiliki banyak algoritma, seperti algoritma *C4.5*, *Naive Bayes*, *Neural Network*, *K-Nearest Neighbor*, *ID3*, *CART*, *Random Forest*, *Linear Discriminant Analysis*, namun peneliti memilih klasifikasi menggunakan algoritma *K-Nearest Neighbor* untuk mengetahui kualitas padi dalam proses data mining.

Oleh karena itu maka di pandang perlu untuk melakukan sebuah penelitian dalam mencari pola dari mutu padi organik dengan menggunakan metode *K-Nearest Neighbor* dengan tujuan agar pengklasifikasian tidak lagi sekedar hanya menggunakan perkiraan semata tapi menggunakan data

pengalaman dari seorang *expert* yang sudah di *extrak* dan menjadi acuan utama dalam menentukan klasifikasi mutu dari sebuah padi.

2. METODOLOGI

Adapun metode penelitian yang digunakan dalam penelitian ini adalah penelitian *Experiment*. Penelitian eksperimen merupakan sebuah penyelidikan hubungan kausal menggunakan tes dikendalikan oleh peneliti. Dalam eksperimen biasanya terdiri dari:

1. Mendefinisikan hipotesis teoritis
2. Memilih sampel dari populasi yang diketahui
3. Mengalokasikan sampel untuk kondisi percobaan yang berbeda
4. Memperkenalkan perubahan yang direncanakan untuk satu atau lebih variabel
5. Mengukur sejumlah kecil variabel
6. Mengontrol semua variabel

Pada penelitian ini, data mutu padi diolah menggunakan metode *data mining* sehingga diperoleh satu metode yang paling akurat dan dapat digunakan sebagai *rules* dalam prediksi kualitas mutu padi. Dalam penelitian ini akan dilakukan beberapa langkah-langkah atau tahapan penelitian seperti yang digambarkan pada gambar 3.1.



Gambar 3.1 : Tahapan Penelitian

2.1 ANALISIS PERMASALAHAN

Permasalahan yang ditemukan dalam lapangan, ada beberapa kesulitan yang dihadapi dalam sistem untuk klasifikasi mutu padi organik seperti :

1. Data masih berbentuk manual jadi masih perlu di input satu agar nanti bisa dijadikan dataset sebagai acuan untuk mengklasifikasi mutu pada padi..
2. Algoritma yang digunakan adalah *K-Nearest Neighbor* atau yang lebih di kenal dengan sebutan *K-NN*, karena yang ada di Dinas pertanian Probolinggo memiliki data numerik value sehingga *K-NN* sangat tepat untuk dijadikan solusi dari keadaan yang ada, sebab *K-NN* bisa mengatasi data campuran yang berupa polinomial dan numerik, pada kenyataannya data yang bernilai numerik sulit untuk diklasifikasikan.
3. Setelah melakukan konversi data, adalah melakukan penanganan data yang hilang atau *missing values*, Untuk melengkapi data yang hilang dilakukan teknik *replace missing values* berbasis *average values*.
4. Melakukan pemecahan 4952 dataset menggunakan teknik *k-fold validation* ke dalam 10 bagian set data dengan sebuah metode validasi yang menggunakan semua data yang ada menjadi *training set* dan *test set*. Dalam *10-fold cross validation* yang berarti bahwa 9 *subset* menjadi *training set* dan 1 *subset* sebagai *testing set*, dengan 10 kali pengulangan pengukuran

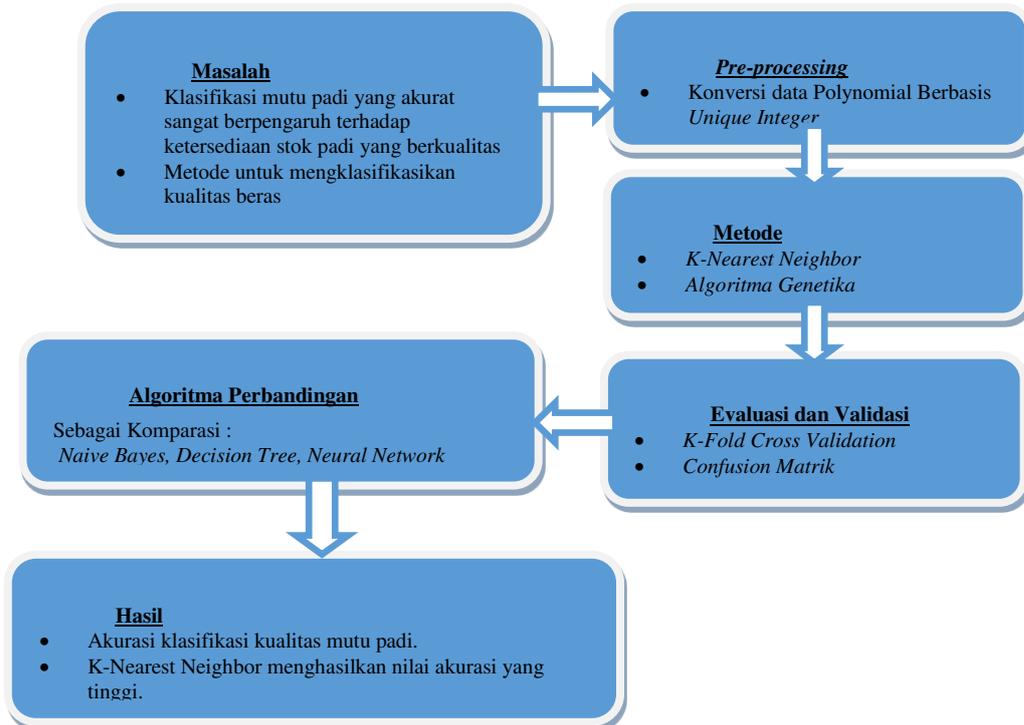
2.2 ANALISIS KEBUTUHAN SISTEM

Kebutuhan yang diinginkan dalam membangun sistem ini agar dapat mencapai tujuan yang sesuai dengan keinginan yaitu dengan beberapa tahapan metode sebagai berikut :

1. Penerapan metode *Algoritma K-NN*
2. Penerapan metode *K-NN* dan *Algoritma Genetika*
3. Validasi menggunakan *K-Fold cross validation*
4. Evaluasi menggunakan *confusion matrix*

2.1. Kerangka Pemikiran

Dalam menyelesaikan penelitian, penulis melakukan kerangka penelitian agar terstruktur.



Gambar 2.5 :Kerangka Pemikiran Penelitian

2.2. Eksperimen dan Pengujian Metode

Ekperimen awal yang di uji coba oleh penulis yaitu melakukan uji coba terhadap dataset dengan menggunakan beberapa algoritma. Dari hasil beberapa uji coba tersebut akan menghasilkan metode yang terbaik, kemudian dilakukan optimasi dengan Algoritma Genitika untuk menghasilkan tingkat akurasi yang lebih memuaskan.

Metode yang digunakan untuk mengukur hasil akurasi mutu padi ini adalah menggunakan algoritma C4.5, Naïve Bayes, Neural Network dan *K-Nearest Neighbor* dengan mengatur k untuk mendapatkan akurasi tertinggi dari algoritma ini.

Tahapan analisis data yang dilakukan menggunakan metode *K-Nearest Neighbor (K-NN)* adalah sebagai berikut:

- a. Menentukan data uji.
- b. Menghitung jarak (kedekatan lokasi) antara data uji dengan data yang lain, parameter jarak yang digunakan adalah jarak Euclidean
- c. Menentukan nilai K
- d. Hitung jumlah data mutu yang ada dari nilai K-tetangga tersebut
- e. Menghitung klasifikasi

3. HASIL DAN PEMBAHASAN

Analisa hasil produksi padi menggunakan penerapan dengan metode *K-NN* dan *Algoritma Genetika* ini diterapkan ke dalam 4 tahapan metode. Dari hasil uji coba beberapa tahapan berfungsi untuk mengetahui tingkat akurasi hasil yang baik juga didapat-kan bahwa analisa produksi padi menggunakan pendekatan ini menghasilkan tingkat akurasi yang tinggi dengan rata-rata akurasi 90% - 95%. Tipe fungsi keanggotaan yang diuji adalah mencari nilai *k* tertinggi dari tetangga terdekat dan hasil dari komparasi dengan *Algoritma genetika*.

2.3. Preprocessing

Dalam tahapan ini dataset yang berupa polynomial akan dikonversi menjadi numerik. Karena dalam proses perhitungan data polynomial sangat sulit untuk dihitung secara manual, maka dari itu proses konversi sangat di butuhkan, berikut adalah proses konversi.

Tabel 4.1 : Atribut dan Value Dataset

Atribut	Value					
Mutu	Grade-A	Grade-B	Grade-C	Grade-D		
Varietas	Beras-hitam	Beras-merah	Mi-kongga	Ciheran	Pandanwangi	IR-64
Panjang	Antara 6.0 hingga 7.8					
Bentuk	Ramping	Sedang	Bulat			
Warna	Putih	Merah	Coklat	Hitam		
Rasa	Pulen	Sangat-pulen				
Teknik	SRI	Jajar-legowo				
Musim	Hujan	Kemarau				
Hama	Tikus	Wereng-coklat	Wereng-hijau	Burung	Penggerek-batang	
PH	Antara 5.2 hingga 6.5					

Setelah menentukan value dan atributnya masing-masing kemudian tahapan konversi adalah dengan metode *unique integer* dengan memberikan nilai numerik secara berurutan dari tiap kolom yang ada pada tabel 4.1. dimana jika atribut pada tabel mutu adalah Grade A, Grade B, Grade C, Grade D. Maka hasil konversinya adalah angka 1,2,3 dan 4. Hasil dari konversi bisa dilihat pada tabel berikut:

Tabel 4.2 : Hasil konversi dataset dengan *unique integer*

Atribut	Value					
Mutu	1	2	3	4		
Varietas	1	2	3	4	5	6
Panjang	Antara 6.0 hingga 7.8					
Bentuk	1	2	3			
Warna	1	2		4		
Rasa	1	2				
Teknik	1	2				
Musim	1	2				
Hama	1	2	3	4	5	
PH	Antara 5.2 hingga 6.5					

2.4. Klasifikasi dengan metode KNN + GA

Hasil dari penghitungan dengan K-NN itu sudah mencapai 80% lebih, namun hasil itu bisa dimaksimalkan lgi, ya itu dengan menambah menambahkan satu metode lgi agar bisa menaikkan hasil grade akurasi dari komparasi tersebut, peneliti menambahkan satu lagi metode yaitu menggunakan metode *Algoritma Genetika(GA)*, dimana itu sendiri menghiung dengan cara ktomoson dan fitness yang ada pada dataset. Setelah menggabungkan dua metode mendapatkan hasil berikut :

Tabel. 4.3 : Hasil Rekapitulasi Metode K-NN+GA

Proses	$k-1$	$k-3$	$k-5$	$k-7$
Validation 1	89.40%	82.80%	78.80%	77.80%
Validation 2	89.40%	82.80%	78.80%	77.80%
Validation 3	93.70%	87.01%	83.60%	82.30%
Validation 4	95.40%	90.60%	86.80%	82.60%
Validation 5	5.60%	91.50%	87.70%	86.00%
Validation 6	95.60%	91.99%	88.80%	84.90%
Validation 7	95.50%	91.70%	87.89%	86.20%
Validation 8	96.40%	92.80%	88.80%	85.80%
Validation 9	95.80%	75.46%	90.10%	87.80%
Validation 10	96.40%	92.90%	89.70%	85.70%

Dari hasil pengujian Algoritma K-NN + GA menghasilkan akurasi yang sangat baik yang semula hanya di atas 81,60% kini bisa menghasilkan 96,40% antara validasi ke-8 dan validasi ke-10 yg menghasilkan akurasi yang sama.

Berikut ini adalah hasil rekapitulasi pengujian model *k-nearest neighbor*, *naive bayes*, *decision tree*, *neural network* dan *K-NN+GA*, selengkapnya pada tabel 4.3.

4. KESIMPULAN

Mengenai kesimpulan dari hasil penelitian adalah sebagai berikut :

Dari hasil analisa dan komputasi beberapa metode, dapat disimpulkan bahwa *K-NN* dan metode *Algoritma Genetika* yang digunakan untuk meningkatkan akurasi klasifikasi mutu padi menghasilkan beberapa kesimpulan berikut:

1. Atribut yang berpengaruh terhadap penentuan mutu beras adalah varietas, bentuk, warna, teknik, musim, hama dan PH. Sedangkan feature yang dianggap tidak berpengaruh adalah rasa dan panjang.
2. Untuk data sebanyak 4952, maka akurasi terbaiknya adalah 96.40%.

DAFTAR PUSTAKA

- (2015, Juli) Badan Pemeriksa Keuangan. [Online]. www.jdih.bpk.go.id
- (2015, September) Cia World factbook. [Online]. www.cia.gov
- (2015, September) Badan Pusat Statistik. [Online]. www.bps.go.id
- Harian Kompas Rabu, 16 September 2015
- (2015, September) Badan Pusat Statistik Jawa Timur. [Online] www.jatim.bps.go.id
- Nobertus Krisandi, “*Algoritma K-Nearest Neighbor dalam Klasifikasi data Hasil Produksi Kelapa Sawit pada PT. MINAMAS Kecamatan Parindu*”, 2013.
- NuRatih Suminar, “*Klasifikasi Kualitas Beras Berdasarkan Ciri Fisik Berbasis Pengoalahan Citra Digital dengan Metode K-NN*”, 2012.
- Nursalim, Suprapedi, dan H. Himawan, “*Klasifikasi Bidang Kerja Lulusan Menggunakan Algoritma K-Nearest Neighbor*”, 2014.
- Nur Rakhmad Setiawan, “*Perbandingan Klasifikasi Gejala Depresi Menggunakan K-Nearest Neighbor dan Naive Bayesian Classification*”, 2011.
- Azhar Arandika, Drs. Mardji, M.T, Imam Cholissodin S.Si., M.Kom, “*Implementasi Algoritma K-Nearest Neighbor (K-NN) Untuk Klasifikasi Data Wine*”, 2010.
- Henny Leidiyana, “*Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit*

- Kepemilikan *Kendaraan Bermotor*”, 2013.
- Iko Pramudiono, Publisher IlmuKomputer.com, “*Pengantar Data Mining, Ilmu Komputer*, (Online), www.ikc.depsos.go.id diakses 05 Oktober 2015
- Eko Prasetyo, “*Data Mining Mengolah Data Menjadi Informasi menggunakan Matlab*”, Yogyakarta: Andi Offset, 2014.
- Fajar Astutik Hermawati, “*Data Mining*”, Yogyakarta: Andi Offset, 2013.
- Kusrini, & Luthfi, Emha, “*Algoritma Data Mining*”, Yogyakarta: Andi, 2009.
- Santosa, B. Data Mining, “*Teknik Pemanfaatan Data untuk Keperluan Bisnis, Teori dan Aplikasi*”, Yogyakarta : Graha Ilmu, 2007.
- Han, J., & Kamber, M, “*Data mining concept and tehniques, san fransisco morgan kauffman*”, 2006.
- L Prema RAJESWARI, Kannan ARPUTHARAJ, “*An Active Rule Approach for Network Intrusion Detectionwith Enhanced C4.5 Algorithm*”, 2008.
- Budiwan Wijakso,Lailil Muflikhah, Achmad Ridok, “*Klasifikasi Jurnal Ilmiah Berbahasa Inggris Berdasarkan Abstrak menggunakan ALGORITMA ID3*”, 2012.
- Alfa Saleh, “*Penerapan Data Mining Dengan Metode Klasifikasi Naïve Bayes Untuk Memprediksi Kelulusan Mahasiswa Dalam Mengikuti English Proficiency Test*”, Medan, 2014.
- Syaeful Mujab, “*Pencarian Model Terbaik Antara Algoritma K-NN Dan C4.5 Berbasis Particle Swarm Op Timization Untuk Prediksi Promosi Deposito*”, Semarang, 2012.
- Durdu Omer Faruk, “*A Hybrid Neural Network And ARIMA Model For Water Quality Time Series Prediction*”, 2010.
- A. Moore. A tutorial on kd-trees. Technical repor, Available from www.cs.cmu.edu/~awm/papers.html, diakses 09 Oktober 2015.
- Zainudin Zukhri, “*Algoritma Genitika Metode Komputasi Evolusioner untuk Menyelesaikan Masalah Optimasi*”, Yogyakarta : Andi Offset, 2014.