

KLASIFIKASI MUTU PADI ORGANIK MENGGUNAKAN C4.5 DI DINAS PERTANIAN BONDOWOSO

Mochammad Faid¹

Program Studi Teknik Informatika dan Teknik Elektronika, STT Nurul Jadid
JL. KH Zaini Mun'im, Kecamatan Paiton, Karanganyar, Paiton, Probolinggo, Jawa Timur 67291
*Email: ahmad.faid08@gmail.com

Abstrak

*Pertumbuhan yang sangat pesat dari akumulasi data mutu padi organik dari seorang expert telah menciptakan kondisi kaya akan data tapi minim informasi. Data mining merupakan penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data dalam jumlah besar yang diharapkan dapat mengatasi kondisi tersebut. Pada preproses dilakukan perubahan data dari bentuk *.CSV yang bisa dikenali pada excel diubah formatnya menjadi *.arff agar bisa dikenali oleh library weka. Setelah menjadi bentuk *.arff dilakukan pemasukkan data pada software data mining yang menggunakan library weka langkah selanjutnya adalah penentuan kandidat dengan menggunakan membandingkan akurasi dari berbagai macam kondisi diantaranya tanpa pruning (Pemangkasan tree) dan tanpa missing value (Penanganan Data kosong), tanpa pruning dan menggunakan proses missing value, dengan pruning dan tanpa missing value dan kondisi yang terakhir dengan pruning dan proses missing value. Membandingkan algoritma C4.5 dengan algoritma klasifikasi yang lainnya, yang bertujuan untuk mengetahui kinerja C4.5 dibandingkan dengan yang lain. Dari penelitian ini didapatkan bahwa tingkat akurasi juga dipengaruhi oleh berbagai macam kondisi yang ada seperti pruning dan missing value dengan adanya software ini diharapkan mampu membantu permasalahan di dinas pertanian terkait dengan kelengkapan data dan hasil prediksi data dan hasil prediksi dengan akurasi yang cukup baik.*

Kata kunci: mutu padi organik, klasifikasi, dan C4.5

1. PENDAHULUAN

Beras merupakan salah satu sereal paling penting di dunia untuk konsumsi manusia. Di negara-negara Asia yang penduduknya padat, khususnya Bangladesh, Myanmar, Kamboja, Cina, Indonesia, Korea, Laos, Filipina, Sri Lanka, Thailand, dan Vietnam, beras juga merupakan salah satu makanan pokok. Sebanyak 75% makanan harian masyarakat di negara-negara Asia tersebut berasal dari beras. Lebih dari 50% penduduk dunia tergantung pada beras sebagai sumber kalori utama (Resita Wahyu Dianti, 2010).

Meningkatkan produksi padi dengan meningkatkan dosis pupuk anorganik dan pestisida bisa tidak meningkatkan hasil padi karena kondisi tanah yang buruk. Kondisi ini perlu perbaikan karena tanah adalah sumber kehidupan bagi tanah mikro-organisme dan aktivitas biologis. Untuk mengatasi masalah tersebut, petani seharusnya dididiki dan dilatih untuk dapat menganalisis dan mempelajari dan praktik budidaya menekankan pada manajemen potensi lokal (Entun Santosa, 2012).

Bondowoso merupakan salah satu kabupaten di Jawa Timur yang juga membudidayakan padi organik dengan tujuan meningkatkan pendapatan kabupaten Bondowoso. Selain itu padi organik juga memiliki gizi yang cukup tinggi untuk dikonsumsi oleh masyarakat Bondowoso. Untuk saat ini padi organik hanya bisa dinikmati oleh masyarakat Bondowoso menengah keatas yang orientasinya bukan hanya masalah kenyang tapi juga tak kalah penting adalah sehat. Dan juga tidak menuntut kemungkinan masyarakat menengah kebawah juga bisa merasakan manfaat dari padi organik ini kedepannya.

Pertumbuhan yang sangat pesat dari akumulasi data mutu padi organik dari seorang expert telah menciptakan kondisi kaya akan data tapi minim informasi. Data Mining merupakan penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data dalam jumlah besar yang diharapkan dapat mengatasi kondisi tersebut.

Jika mutu padi organik diketahui maka pihak dinas pertanian Bondowoso dapat melakukan penentuan harga yang sesuai dengan mutu dan juga meningkatkan kepercayaan mitra yang berkerja sama dengan pihak dinas pertanian. Oleh karena itu maka di pandang perlu untuk melakukan

sebuah penelitian dalam mencari pola dari mutu padi organik dengan menggunakan metode C4.5 dengan tujuan agar pengklasifikasian tidak lagi sekedar hanya menggunakan perkiraan semata tapi menggunakan data pengalaman dari seorang expert yang sudah di ekstrak dan menjadi acuan utama dalam menentukan klasifikasi mutu dari sebuah padi organik.

2. METODOLOGI

Dalam sub bab ini akan membahas tentang analisa permasalahan, analisa kebutuhan sistem arsitektur sistem, beserta spesifikasi dan kemampuan sistem.

2.1 ANALISIS PERMASALAHAN

Berdasarkan penelitian yang dilakukan di Dinas Pertanian Bondowoso, ditemukan ada beberapa masalah yang dihadapi dalam sistem untuk klasifikasi mutu padi organik seperti berikut ini :

1. Di Dinas pertanian Bondowoso memiliki data terkait padi organik, namun data tersebut masih berbentuk manual sehingga sulit untuk melihat pola mutu padi organik secara menyeluruh, untuk kepentingan pengembangan selanjutnya maka dibutuhkan sebuah software yang bisa mengklasifikasi data mutu padi organik di Dinas pertanian bondowoso, hal ini diperlukan agar data yang ada bisa lebih bermanfaat lagi bukan hanya sekedar tumpukan data yang minim informasi.
2. Algoritma yang digunakan untuk membuat software tersebut adalah C45, karena yang ada di Dinas pertanian bondowoso memiliki banyak numerik value sehingga C4.5 sangat tepat untuk dijadikan solusi dari keadaan yang ada, sebab c4.5 bisa mengatasi data yang berupa numerik, pada kenyataannya data yang bernilai numerik sulit untuk diklasifikasikan.
3. Setelah metode klasifikasi ditemukan maka masalah selanjutnya adalah mencari library data mining yang bisa digunakan untuk membuat software tersebut dan dalam penelitian ini librari yang digunakan adalah weka, weka adalah library data mining yang bisa digunakan diberbagai bahasa pemograman contohnya python dan java.
4. Permasalah lain yang ada di dinas pertanian Bondowoso adalah Banyak -nya data yang missing value dan hal ini sudah bisa diatasi oleh algoritma C4.5 dengan melihat pola dari data maka C4.5 akan mengisi data yang missing value tersebut sesuai dengan pola yang ada pada data sebelumnya.

2.2 ANALISIS KEBUTUHAN SISTEM

Dalam penelitian yang dibutuhkan untuk membangun sistem ini agar dapat mencapai tujuan yang diinginkan adalah sebagai berikut ini :

2.2.1 Teknik Pengumpulan Data

Dalam penelitian ini teknik pengumpulan data dilakukan dengan menggunakan wawancara. Wawancara pada penelitian sampel besar biasanya hanya dilakukan sebagai studi pendahuluan karena tidak mungkin menggunakan wawancara pada 1000 responden, sedangkan pada sampel kecil teknik wawancara dapat diterapkan sebagai teknik pengumpul data (umumnya penelitian kualitatif).

Teknik wawancara yang digunakan dalam penelitian ini adalah Wawancara terstruktur , Setelah wawancara selesai baru pengumpulan data dilakukan oleh bapak Mulyono selaku ketua kelompok tani dalam naungan Dinas pertanian Bondowoso

2.2.2 Data Padi Organik

Organik sebagaimana digunakan pada kebanyakan tanaman sawah yang umumnya, adapun penjelasan tentang organik antara lain :

- Tidak ada pestisida dan pupuk dari bahan kimia sintesis atau buatan yang telah digunakan.
- Kesuburan Tanah dipelihara melalui proses alami seperti penanaman tumbuhan penutup dan penggunaan pupuk kandang yang dikomposkan dan juga limbah tumbuhan.
- Tanaman dirotasikan disawah untuk menghindari penanaman tanaman yang sama dari tahun ke tahun disawah yang sama.

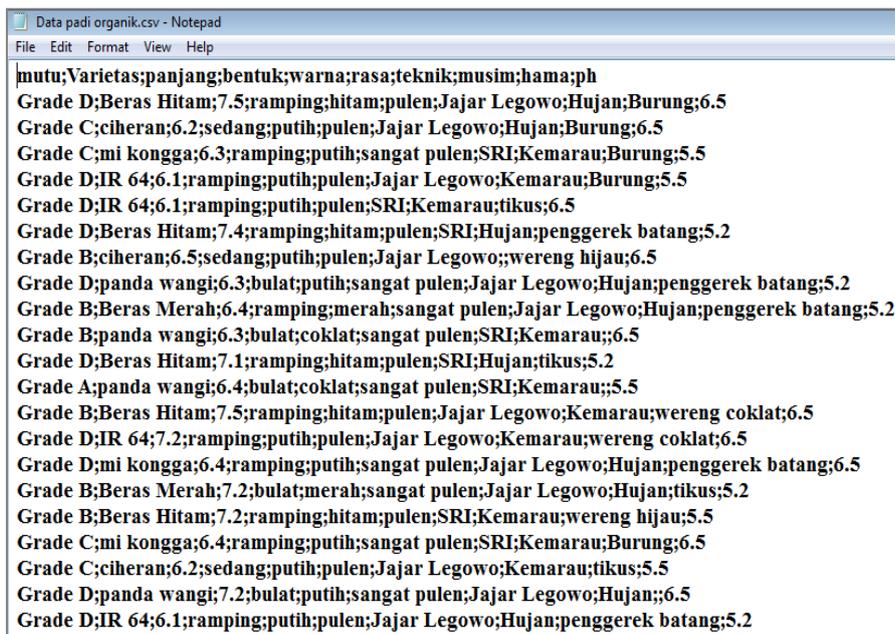
	A	B	C	D	E	F	G	H	I	J
1	mutu	Varietas	panjang	bentuk	warna	rasa	teknik	musim	hama	ph
2	Grade D	Beras Hita	7.5	ramping	hitam	pulen	Jajar Legowo	Hujan	Burung	6.5
3	Grade C	ciheran	6.2	sedang	putih	pulen	Jajar Legowo	Hujan	Burung	6.5
4	Grade C	mi kongga	6.3	ramping	putih	sangat pulen	SRI	Kemarau	Burung	5.5
5	Grade D	IR 64	6.1	ramping	putih	pulen	Jajar Legowo	Kemarau	Burung	5.5
6	Grade D	IR 64	6.1	ramping	putih	pulen	SRI	Kemarau	tikus	6.5
7	Grade D	Beras Hita	7.4	ramping	hitam	pulen	SRI	Hujan	penggerek	5.2
8	Grade B	ciheran	6.5	sedang	putih	pulen	Jajar Legowo		wereng hijau	6.5
9	Grade D	panda wangi	6.3	bulat	putih	sangat pulen	Jajar Legowo	Hujan	penggerek	5.2
10	Grade B	Beras Merah	6.4	ramping	merah	sangat pulen	Jajar Legowo	Hujan	penggerek	5.2
11	Grade B	panda wangi	6.3	bulat	coklat	sangat pulen	SRI	Kemarau		6.5
12	Grade D	Beras Hita	7.1	ramping	hitam	pulen	SRI	Hujan	tikus	5.2
13	Grade A	panda wangi	6.4	bulat	coklat	sangat pulen	SRI	Kemarau		5.5
14	Grade B	Beras Hita	7.5	ramping	hitam	pulen	Jajar Legowo	Kemarau	wereng coklat	6.5
15	Grade D	IR 64	7.2	ramping	putih	pulen	Jajar Legowo	Kemarau	wereng coklat	6.5
16	Grade D	mi kongga	6.4	ramping	putih	sangat pulen	Jajar Legowo	Hujan	penggerek	6.5

Gambar 1. Data File .xlsx open excel

- Data dalam penelitian ini didapat dengan cara wawancara pada seorang expert padi organik yaitu bapak mulyono ketua kelompok tani yang bernaung di dinas pertanian bondowoso.

2.2.3 Membuat Data ke Format .Arff

Kelas mutu yang ada di Dinas pertanian Bondowoso ada 4 kelas, diantaranya Grade A, Grade B, Grade C, dan Grade D. mutu yang terbaik adalah mutu dengan Grade A, dengan begini kualitas mutu, padi organik bisa dilihat dengan melihat level alfabet mulai dari A sampai D. Sebenarnya file CSV ini bisa diterima oleh weka sebagai input. Tapi ada 2 standar yang menggunakan (,) dan (;) dan weka menggunakan (,) sehingga ketika disimpan ke .CSV dengan menggunakan (;) tidak akan terbaca oleh karena itu harus secara manual diubah (;) menjadi (,) dengan cara membuka notepad dan menggunakan fungsi replace.



Gambar 2. Data File.csv dengan notepad

Gambar2 adalah bentuk file csv jika file tersebut di open dengan menggunakan notepad format inilah yang nantinya dirubah secara manual kemudian disave dengan nama yang sama tapi ekstensinya disimpan dalam format arff

```

data.arff - Notepad
File Edit Format View Help
@RELATION mutu

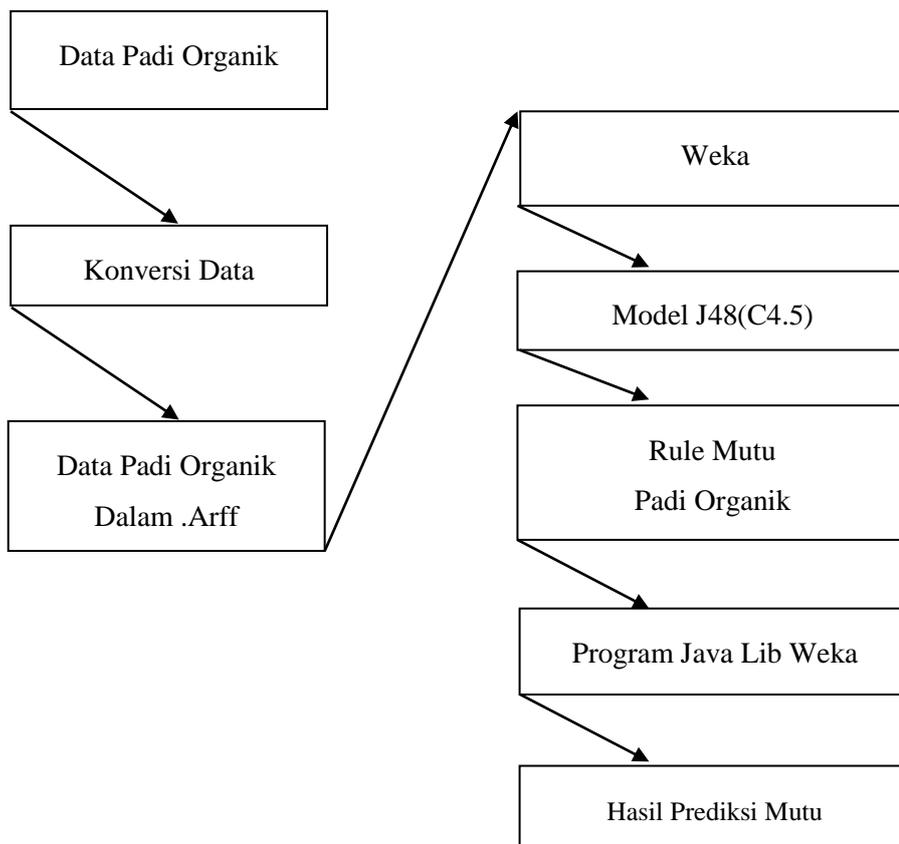
@ATTRIBUTE classMutu {Grade-A,Grade-B,Grade-C,Grade-D}
@ATTRIBUTE Variates {panda-wangi,ciheran,mi-kongga,IR-64,Beras-Merah,Beras-Hitam}
@ATTRIBUTE Panjang REAL
@ATTRIBUTE Bentuk {ramping,sedang,bulat}
@ATTRIBUTE warna {putih,merah,hitam,coklat}
@ATTRIBUTE rasa {sangat-pulen,pulen}
@ATTRIBUTE teknik {Jajar-Legowo,SRI}
@ATTRIBUTE Musim {Hujan,Kemarau}
@ATTRIBUTE penyakit {Burung,tikus,penggerek-batang,wereng-coklat,wereng-hijau}
@ATTRIBUTE PH REAL

@DATA
Grade-D,Beras-Hitam,7.5,ramping,hitam,pulen,Jajar-Legowo,Hujan,Burung,6.5
Grade-C,ciheran,6.2,sedang,putih,pulen,Jajar-Legowo,Hujan,Burung,6.5
Grade-C,mi-kongga,6.3,ramping,putih,sangat-pulen,SRI,Kemarau,Burung,5.5
    
```

Gambar 3. Data File .arff open Notepad

2.2.4 Rancangan algoritma

Rancangan algoritma yang digunakan untuk mendapatkan hasil yang sesuai dengan tujuan penelitian ini terlihat pada gambar berikut :



Gambar 4. Blok Diagram metode J48 (C4.5)

Tahap pertama yang dilakukan adalah mengumpulkan data padi organik di Dinas Pertanian Bondowoso dalam excel yang dikonversi kedalam bentuk Arff, Tahap kedua adalah melakukan proses C4.5 pada input Data padi Organik yang sudah terkonversi ke bentuk Arff. Tahap ketiga

adalah melakukan pembuatan rule yang dibantu dengan software weka sehingga proses pembuatan rulenya bisa lebih cepat dan efisien. Tahap keempat adalah membuat aplikasi dengan menggunakan java dengan menggunakan library yang ada di weka sehingga proses pembuatan program atau aplikasi data mining bisa lebih *cepat* dari pada membuat secara manual dengan mengkodekan seperti biasa. Tahap kelima adalah melakukan uji coba pada program java yang sudah terkoneksi pada weka, uji coba ini dilakukan untuk menghasilkan prediksi mutu padi organik sesuai dengan inputan data yang ada.

3. HASIL DAN PEMBAHASAN

Tujuan dari penelitian ini menguji keakuratan analisa mutu padi organik dengan menggunakan algoritma C4.5, Data yang dianalisa adalah data grade mutu padi organik yang disetujui oleh pihak Dinas pertanian bondowoso.

3.1. Membandingkan Algoritma

untuk mengetahui kualitas sebuah algoritma maka perlu kiranya melakukan sebuah perbandingan dengan algoritma yang sejenis, sehingga bisa melihat algoritma yang memiliki kinerja baik dalam menyelesaikan masalah, pada tabel 1 adalah hasil uji coba dengan data yang sama namun dengan algoritma yang berbeda.

Tabel 1 Perbandingan Tingkat akurasi

Algoritma	Tingkat Akurasi
NBTree	82.4111%
C4.5 / J48	83.0372%
Decision Stump	64.378%
FTTree	82.2092%
LADTree	72.8595%

Berdasarkan uji coba yang telah dilakukan, diperoleh bahwa *Decision Tree J48* memberikan hasil akurasi sedikit lebih baik dari FTTree dengan selisih akurasi 0.828% sedangkan pada LADTree 10.1777

3.2. Pengujian Model

Proses pemodelan diawali dengan melakukan *feature selection*. tahap *feature selection* dilakukan dengan dua tahap yaitu *ranking selection*. Atribut yang di-*filter* diperoleh dari proses *ranking selection* dimana 3 buah atribut paling bawah yang tidak berpengaruh akan diabaikan dalam proses selanjutnya, proses *feature selection* yang dilakukan pada *tools* weka dengan 10 *cross-validation* diperoleh 3 atribut terbawah yang tidak terlalu berpengaruh terhadap akurasi perhitungan, dalam kasus ini, 3 atribut yang diabaikan yaitu : musim, rasa dan panjang.

Tabel 2 Attribute selection 10 fold

Average	Average	rank	Attribute
0.104	0.002	1	Penyakit
0.088	0.002	2	Warna
0.073	0.001	3	PH
0.044	0	4	Varietas
0.016	0.001	5	Bentuk
0.014	0.001	6	Teknik
0.005	0	7	Musim
0.004	0	8	Rasa
0.002	0.002	9	Panjang

Nilai *accuracy*, *precision* dan *recall* dari data training dapat dihitung dengan menggunakan weka 3.6.9 , Setelah diuji coba dengan menggunakan *cross-validation* , didapatkan hasil pengukuran terhadap data training yaitu *accuracy*=80.4725 % pada hal sebelum atribut dikurangi tingkat akurasi adalah sebesar 83.8384 %

Tabel 3 Detail Akurasi Per-Atribute

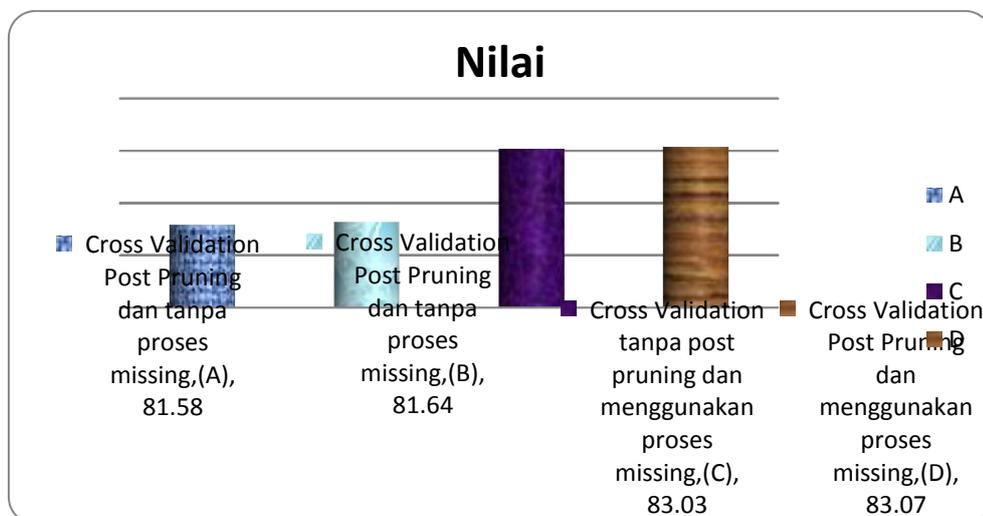
No	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC	Class
1	0.7	0.002	0.795	0.7	0.745	0.869	Grade-A
2	0.838	0.124	0.727	0.838	0.779	0.878	Grade-B
3	0.791	0.102	0.905	0.791	0.844	0.864	Grade-C
4	0.972	0.038	0.828	0.972	0.894	0.975	Grade-D
AVG	0.832	0.097	0.841	0.832	0.832	0.885	

Tabel 4 Confusion Matrix Tanpa atribut rasa, musim dan panjang

		Grade Prediksi			
		Grade A	Grade B	Grade C	Grade D
Actual Grade	Grade A	12	11	24	3
	Grade B	3	1059	328	14
	Grade C	18	378	2184	142
	Grade D	1	23	22	730

3.3. Perbandingan akurasi pruning dan missing value

Pada gambar 5 terdapat grafik yang membandingkan tingkat akurasi data mining , pada tabung berwarna biru tingkat akurasinya sebesar 81.58 % dimana hasil treenya tidak pruning dan tidak replaceMissing, pada tabung berwarna merah memiliki tingkat akurasi sebesar 81.64 % , tabung berwarna kuning sekitar 83.03 % sedangkan tabung berwarna hijau dengan akurasi sebesar 83.05 % , hal ini membuktikan bahwa pruning dan missing value dapat mempengaruhi akurasi dari sebuah data mining, semakin lengkap sebuah data mining maka semakin meningkatkan akurasi dan pruning juga sama semakin banyak tree yang dipruning maka semakin meningkatkan akurasinya.

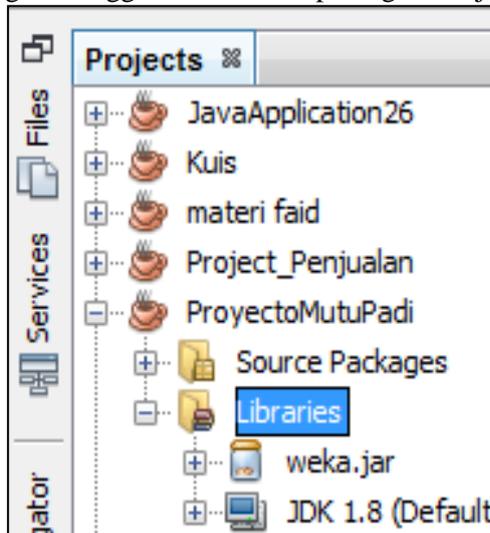


Gambar 5. Perbandingan Akurasi Pruning dan missing value

3.4. Penerapan Library Weka pada Java Netbeans

Implementasi program klasifikasi dokumen menggunakan tools yang berbeda untuk masing masing metode. Klasifikasi data padi organik dengan metode C4.5 / J48 menggunakan lib weka digunakan untuk melakukan klasifikasi.

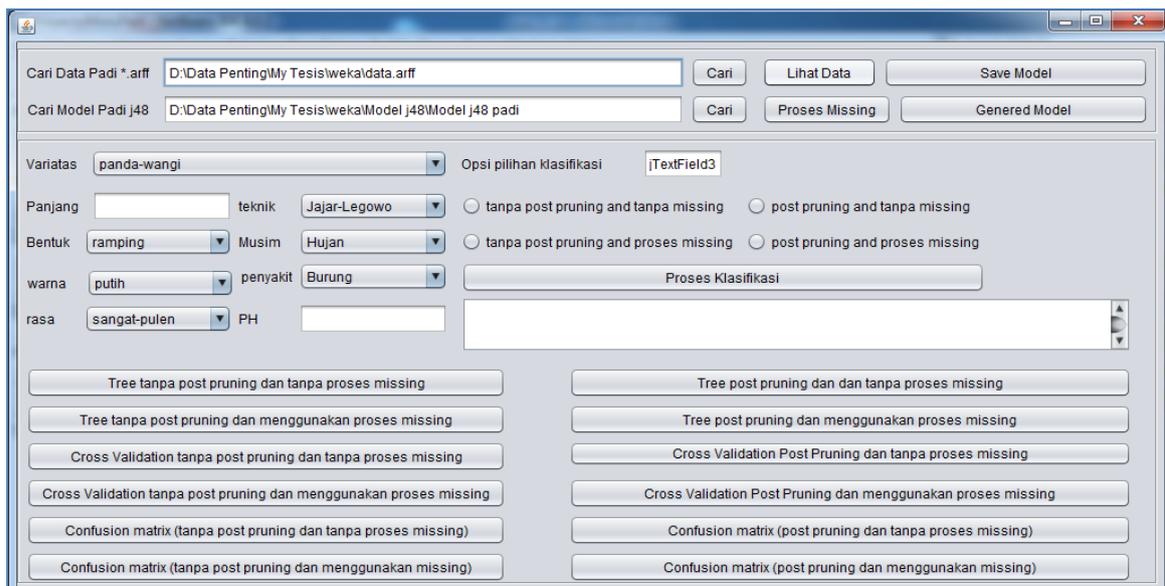
Library weka menggunakan berkas *attribute-relation* file format (Arff) sebagai masukan untuk melakukan klasifikasi dokumen, ontologi yang direpresentasikan dalam bahasa pemrograman java akan memudahkan dalam proses klasifikasi dokumen karena implemmentasi klasifikasi dokumen dikembangkan dengan menggunakan bahasa pemrograman java.



Gambar 6. Library Weka Dalam Netbeans IDE 8.0.2

3.5. Penerapan klasifikasi sesuai dengan input attribute

Penerapan klasifikasi sesuai dengan inputan data padi organik dengan menggunakan model j48 yang pertama yang harus kita lakukan adalah mengetahui letak model j48 kemudian memanggil lib wekaWrapper.java untuk menentukan hasil klasifikasinya



Gambar 7. Penerapan Klasifikasi

4. KESIMPULAN

Mengenai kesimpulan dari hasil penelitian adalah sebagai berikut :

1. Masalah *missing value* pada data padi organik bisa diatasi dengan *replace missing value* tingkat akurasi pada data sebesar 81.582% tapi setelah dilakukan proses *replace missing value* tingkat akurasi meningkat menjadi 81.6438 %
2. Pemangkasan pohon keputusan atau *pruning* juga mempunyai pengaruh terhadap tingkat akurasi data mining, pemangkasan pohon tanpa *missing value* memiliki tingkat akurasi data mining, pemangkasan pohon tanpa *missing value* memiliki tingkat akurasi sebesar 83.0372 %, sedangkan jika pemangkasan *pruning* pohon dengan *replace missing value* akurasinya meningkat menjadi 83.0775 %.
3. Menentukan mutu padi organik, menggunakan metode *levenshtein string metric*, Metode *levenshtein string metric* melakukan pengecekan setiap karakter string pada setiap atribut data padi organik. Metode C4.5/j48 menghasilkan akurasi sebesar 83.0775% namun ketika attribute yang kurang berpengaruh diabaikan tingkat akurasi menurun sebesar 4,06%

DAFTAR PUSTAKA

- A.Nithya, Dr V. Sundaram. 2011. Classification rules for indian race diseases, IJCSI, India
- Han, J., & Kamber, M (2006). Data mining concept and tehniques, san fransisco morgan Kauffman
- Zulfian Azmi,Muhammad Dahria, 2013, Decision Tree Berbasis Algoritma Untuk Pengambilan Keputusan,saintikom
- Wahyudin, 2009, Metode Iterative Dichotomizer 3 (ID3) Untuk Penyeleksian Penerimaan Mahasiswa Baru,JPTIK
- Alfa saleh, 2014, Penerapan Data Mining Dengan Metode Klasifikasi Naïve Bayes Untuk Memprediksi Kelulusan Mahasiswa Dalam Mengikuti English Proficiency Test, Medan.
- syaeful mujab, 2012, pencarian model terb aik antara algoritma c4.5 dan c4.5 berb asis particle swarm op timization untuk prediksi promosi deposito, semarang.
- Kiarash Afsharpour Rezaeieh, Alireza Eivazi, 2011 Evaluation of morphological characteristics in five Persian maize (*Zea maysL.*) under drought stress, ajar, Turkey.
- s. Dewi Indrasari, 2009, peningkatan nilai tambah beras melalui mutu fisik, cita rasa , dan gizi
- Ihsan A. Kareem, Mehdi G. Duaimi, 2014 , International Jurnal Computer Science and mobile Computing , ijcsmc.
- Yosoa Putra Raharja, 2013, rancangan bangun sistem rekomendasi beasiswa menggunakan algoritma klasifikasi C4.5,Semarang
- Berry and Linoff, 2004, Data Mining Techniques for Marketing,. Sales and CRM,wiley Chapter1.
- Azwar Rizal Alfarisi, Handayani Tjandrasa, dan Isye Arieshanti ,Perbandingan Performa antara Imputasi Metode Konvensional dan Imputasi dengan Algoritma Mutual Nearest Neighbor.
- Johanes Furnkranz , 1994, A comparison of pruning methods for relational concept learning, AAAi, Austria.
- L Prema RAJESWARI, Kannan ARPUTHARAJ. 2008, J. Communications, Network and System Sciences.ijcnsI.
- Resita Wahyu Dianti. 2010. kajian karakteristik fisikokimia dan sensori beras organik mentik susu dan ir64; pecah kulit dan giling selama penyimpanan. universitas sebelas maret surakarta.
- Entun Santosa. 2012, Rice Organik Farming is a Programme for Strengtenning Food Security in Sustainable Rural Development. IJBAS.
- Larose, D. T. 2005. *Discovering Knowledge in Data*. Canada. Wiley-Interscience.
- Zurada, J. (2010). Could Decision Trees Improve the Classification Accuracy Interpretability of Loan Granting Decisions. *HICSS '10 Proceedings of the 2010 43rd Hawaii International Conference on System Sciences*, (hal. 1- 9). Koloa.
- Han, J., & Kamber, M. (2006). *Data Mining Concept and Tehniques*. San Fransisco: Morgan Kauffman.