

VOL. 18 NO. 3 SEPTEMBER 2017

ISSN : 1411-3201

Jurnal Ilmiah

DASI

DATA MANAJEMEN DAN TEKNOLOGI INFORMASI



UNIVERSITAS
AMIKOM
YOGYAKARTA

JURNAL
ILMIAH
DASI

**DATA MANAJEMEN DAN
TEKNOLOGI INFORMASI**



**UNIVERSITAS
AMIKOM
YOGYAKARTA**

VOL. 18 NO. 3 SEPTEMBER 2017
JURNAL ILMIAH
Data Manajemen Dan Teknologi Informasi

Terbit empat kali setahun pada bulan Maret, Juni, September dan Desember berisi artikel hasil penelitian dan kajian analitis kritis di dalam bidang manajemen informatika dan teknologi informatika. ISSN 1411-3201, diterbitkan pertama kali pada tahun 2000.

KETUA PENYUNTING

Abidarin Rosidi

WAKIL KETUA PENYUNTING

Heri Sismoro

PENYUNTING PELAKSANA

Emha Taufiq Luthfi

Hanif Al Fatta

Hartatik

Hastari Utama

STAF AHLI (MITRA BESTARI)

Jazi Eko Istiyanto (FMIPA UGM)

H. Wasito (PAU-UGM)

Supriyoko (Universitas Sarjana Wiyata)

Ema Utami (AMIKOM)

Kusrini (AMIKOM)

Amir Fatah Sofyan (AMIKOM)

Ferry Wahyu Wibowo (AMIKOM)

Rum Andri KR (AMIKOM)

Arief Setyanto (AMIKOM)

Krisnawati (AMIKOM)

ARTISTIK

Robert Marco

TATA USAHA

Nila Feby Puspitasari

PENANGGUNG JAWAB :

Rektor UNIVERSITAS AMIKOM YOGYAKARTA, Prof. Dr. M. Suyanto, M.M.

ALAMAT PENYUNTING & TATA USAHA

UNIVERSITAS AMIKOM YOGYAKARTA, Jl. Ring Road Utara Condong Catur Yogyakarta, Telp. (0274) 884201 Fax. (0274) 884208, Email : jurnal@amikom.ac.id

BERLANGGANAN

Langganan dapat dilakukan dengan pemesanan untuk minimal 4 edisi (1 tahun)

pulau jawa Rp. 50.000 x 4 = Rp. 200.000,00 untuk luar jawa ditambah ongkos kirim.

VOL. 18 NO. 3 SEPTEMBER 2017

ISSN : 1411- 3201

JURNAL ILMIAH

DASI

DATA MANAJEMEN DAN TEKNOLOGI INFORMASI

UNIVERSITAS AMIKOM YOGYAKARTA

JURNAL ILMIAH

DASI

KATA PENGANTAR

Puji syukur kehadirat Tuhan Yang Maha Kuasa atas anugerahnya sehingga jurnal edisi kali ini berhasil disusun dan terbit. Beberapa tulisan yang telah melalui koreksi materi dari mitra bestari dan revisi redaksional dari penulis, pada edisi ini diterbitkan. Adapun jenis tulisan pada jurnal ini adalah hasil dari penelitian dan pemikiran konseptual. Redaksi mencoba selalu mengadakan pembenahan kualitas dari jurnal dalam beberapa aspek.

Beberapa pakar di bidangnya juga telah diajak untuk berkolaborasi mengawal penerbitan jurnal ini. Materi tulisan pada jurnal berasal dari dosen tetap dan tidak tetap Universitas AMIKOM Yogyakarta serta dari luar Universitas AMIKOM Yogyakarta.

Tak ada gading yang tak retak begitu pula kata pepatah yang selalu di kutip redaksi, kritik dan saran mohon di alamatkan ke kami baik melalui email, faksimile maupun disampaikan langsung ke redaksi. Atas kritik dan saran membangun yang pembaca berikan kami menghaturkan banyak terimakasih.

Redaksi

DAFTAR ISI

HALAMAN JUDUL.....	i
KATA PENGANTAR	ii
DAFTAR ISI.....	iii
Prediksi Jumlah Pendaftaran Calon Mahasiswa Baru Dengan Metode Regresi Linier.....	1-5
Harliana ¹⁾ , Andri Syafrianto ²⁾ (¹⁾ Ilmu Komputer STIKOM Poltek Cirebon, ²⁾ STMIK EL-RAHMA Yogyakarta)	
Pengembangan Sistem Terintegrasi Berbasis Supply Chain Management Menggunakan Barcode Scanner PDA Pada PT XYZ Semarang.....	6-10
Ryan Putranda Kristianto (Informatika Universitas AMIKOM Yogyakarta)	
Analisis Performa Algoritma Klasifikasi Pada Pengelompokan Benih Gandum.....	11-15
Ika Nur Fajri (Sistem Informasi Universitas AMIKOM Yogyakarta)	
Implementasi Token Based Authentifikasi Dan Authorisasi Pada Mekanisme Single Sign On.....	16-23
Norhikmah ¹⁾ , Acihmah Sidauruk ²⁾ (^{1) 2)} Ilmu Komputer Universitas AMIKOM Yogyakarta)	
Simple Additive Weighting (SAW) Untuk Pemilihan Karyawan (Studi Kasus: Rumah Makan Saung Bu Mansur Banjarnegara).....	24-29
Aditya Putut Mahendra ¹⁾ , Yuli Astuti ²⁾ (¹⁾ Sistem Informasi Universitas AMIKOM Yogyakarta, ²⁾ Manajemen Informatika Universitas AMIKOM Yogyakarta)	
Perancangan Metode Sinkronisasi Informasi Akademik Amikom Social.....	30-35
Rizqi Sukma Kharisma ¹⁾ , Arif Dwi Laksito ²⁾ (^{1) 2)} Informatika Universitas AMIKOM Yogyakarta)	
Sistem Penunjang Keputusan Untuk Investor Pada Entrepreneur Campus.....	36-42
Windha Mega PD ¹⁾ , Dina Maulina ²⁾ , Adji Sukmana ³⁾ , Agus Muhammad Z F ⁴⁾ (¹⁾ Informatika Universitas AMIKOM Yogyakarta, ²⁾ Manajemen Informatika Universitas AMIKOM Yogyakarta, ³⁾ Informatika Universitas AMIKOM Yogyakarta, ⁴⁾ Sistem Informasi Universitas AMIKOM Yogyakarta)	
Pembuatan Media Presentasi Animasi Cerita Rakyat Untuk Anak Usia Dini Dengan Konsep Pemilihan Alternatif Alur Cerita.....	43-48
Agus Purwanto ¹⁾ , Yudi Sutanto ²⁾ (^{1) 2)} Informatika Universitas AMIKOM Yogyakarta)	
Multimedia Interaktif Pengenalan Gamelan Jawa “E-Gamel” Menggunakan Teknologi Augmented Reality.....	49-54
Endah Handayani ¹⁾ , Bhanu Sri Nugraha ²⁾ (¹⁾ Informatika Universitas AMIKOM Yogyakarta, ²⁾ Sistem Informasi Universitas AMIKOM Yogyakarta)	

Tingkat Kepastian Certainty Factor Hasil Diagnosis Sistem Pakar Gangguan Tanaman Padi.....	55-62
Suryo Sumpeno ¹⁾ , Emilya Uy Artha ²⁾ , Ardhin Primadewi ³⁾ (¹⁾²⁾³⁾ Teknik Informatika UNIVERSITAS MUHAMMADIYAH Magelang)	
Analisa Trafik Dan Quality Of Service (QoS) Untuk Optimalisasi Manajemen Bandwidth (Studi Kasus : Universitas Amikom Yogyakarta).....	63-70
Nila Feby Puspitasari ¹⁾ , Akhmad Dahlan ²⁾ (¹⁾ Teknik Informatika Universitas AMIKOM Yogyakarta, ²⁾ Manajemen Informatika Universitas AMIKOM Yogyakarta)	

ANALISIS PERFORMA ALGORITMA KLASIFIKASI PADA PENGELOMPOKAN BENIH GANDUM

Ika Nur Fajri

Sistem Informasi Universitas AMIKOM Yogyakarta

Email : ika.nur.fajri@amikom.ac.id

Abstraksi

Gandum merupakan bahan makanan pokok selain beras, permintaan terhadap gandum dunia sampai tahun 2020 diperkirakan meningkat sebesar 1.6% per tahun. Pengolahan data benih gandum sudah banyak dilakukan salah satunya yaitu dengan menggunakan teknik klasifikasi data mining. Dengan menggunakan klasifikasi, data-data yang sebelumnya telah terkumpul dapat digunakan sebagai pengetahuan baru. Melalui beberapa teknik klasifikasi, dapat diperoleh tingkat akurasi yang baik dan tinggi. Dalam penelitian ini akan dilakukan perbandingan performa algoritma Naive Bayes, K-Nearest Neighbour, dan Decision Tree C4.5 untuk klasifikasi benih gandum.

Hasil dari penelitian ini menunjukkan algoritma C4.5 dan K-NN dengan implementasi k=100 memiliki performa terbaik berdasarkan nilai accuracy, kappa statistic dan recall. Nilai akurasi yang diperoleh sebesar 95,24%, kappa statistic sebesar 0,929 dan recall 95,24. Sedangkan performa terendah yaitu algoritma naive bayes dengan accuracy 90%, kappa statistic 0,85 dan recall 90.

Kata kunci : Benih Gandum, Data Mining, Klasifikasi, Naive Bayes, K-NN, C4.5

Abstract

Wheat is a staple food other than rice. Demand for world wheat by 2020 is expected to increase by 1.6% per year. Processing of wheat seed data has been done one of them is using data mining classification technique. Using the classification, previously collected data can be used as new knowledge. Through some classification techniques, can be obtained a good level of accuracy and high. In this study we will compare the performance of Naive Bayes algorithm, K-Nearest Neighbour, and Decision Tree C4.5 for the classification of wheat seeds.

The results of this study show that the algorithm C4.5 and K-NN with the implementation of k = 100 has the best performance based on accuracy, kappa statistic and recall. Accuracy value obtained is 95.24%, kappa statistic of 0.929 and recall 95.24. While the lowest performance is naive bayes algorithm with accuracy 90%, kappa statistic 0.85 and recall 90.

Keywords : Wheat Seed, Data Mining, Classification, Naive Bayes, K-NN, C4.5

PENDAHULUAN

Gandum (*Triticum aestivum L*) merupakan salah satu bahan makanan pokok manusia selain beras. Tanaman ini lebih diminati dibanding sesama sereal yang lain karena memiliki kandungan gizi yang cukup tinggi diantaranya Karbohidrat 60%-80%, protein 6%-17%, lemak 1,5%-2,0%, mineral 1,5%-2,0% dan sejumlah vitamin. Permintaan terhadap gandum dunia sampai tahun 2020 diperkirakan meningkat sebesar 1.6% per tahun. Di negara-negara berkembang peningkatan permintaan gandum diperkirakan mencapai sekitar 2% per tahun. Keragaman penggunaan, kandungan nutrisi dan kualitas penyimpanannya yang tinggi menjadikan gandum sebagai bahan makanan pokok lebih dari sepertiga populasi dunia.

Gandum sudah lama ada di Indonesia dan tumbuh di daerah dataran tinggi bersuhu sejuk [1]. Kebutuhan gandum di Indonesia relatif besar dan selama ini seluruhnya dipenuhi melalui impor. Data impor gandum dan olahan gandum dari Dinas perindustrian menunjukkan nilai impor yang terus meningkat dari tahun ke tahun pada tahun 2008

Indonesia mengimport gandum sebesar 4.514.852 ton, tahun 2009 meningkat menjadi 4.666.418 ton pada 2010 mencapai 4.824.049 ton. Untuk periode Januari-Juni 2011, impor gandum sudah mencapai 2,8 juta ton [2]. Untuk memenuhi kebutuhan tersebut diperlukan peningkatan produksi gandum dua kali dari rata-rata produksi gandum dunia saat ini. Laju peningkatan produksi gandum pada saat ini masih terlalu rendah untuk dapat memenuhi kebutuhan gandum di masa depan.

Pengolahan data benih gandum sudah banyak dilakukan salah satunya yaitu dengan menggunakan teknik klasifikasi data mining. Dengan menggunakan klasifikasi, data-data yang sebelumnya telah terkumpul dapat digunakan sebagai pengetahuan baru. Melalui beberapa teknik klasifikasi, dapat diperoleh tingkat akurasi yang baik dan tinggi. Beberapa algoritma klasifikasi telah dibandingkan untuk mengetahui algoritma mana yang paling baik dan akurat diantara beberapa algoritma yang digunakan. Melalui penelitian Andriana [3] tentang identifikasi benih gandum dengan jaringan syaraf tiruan dengan metode

optimasi klasifikasi menunjukkan hasil yang baik dan cepat. Analisis komparatif evaluasi performa algoritma klasifikasi pada readmisi pasien diabetes, algoritma yang dianalisa meliputi *Naïve Bayes*, K-NN dan *Decision tree* telah dilakukan oleh Yusa dkk [4]. Hasilnya menunjukkan algoritma terbaik yaitu *Naïve Bayes* dengan tingkat akurasi paling tinggi. Perbandingan ketiga algoritma ini juga dilakukan oleh Sartika [5] dengan studi kasus pengambilan keputusan pemilihan pola pakaian, hasilnya menunjukkan algoritma decision tree J48 memiliki hasil terbaik dengan akurasi tertinggi yaitu 75,6%. Berbeda dengan Yusa dan Sartika, hasil yang diperoleh dari penelitian Fakhurrufqi [6], dengan membandingkan algoritma klasifikasi *Nearest Neighbour*, C4.5 dan LVQ, *Nearest Neighbour* menghasilkan akurasi tertinggi dibandingkan dengan algoritma yang lain. Dalam penelitian ini, algoritma klasifikasi yang dibandingkan meliputi algoritma *Naïve Bayes*, K-NN dan *Decision Tree* C4.5 untuk klasifikasi benih gandum.

DASAR TEORI

Data Mining

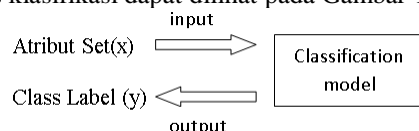
Data Mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis [7].

Data mining berisi pencarian *trend* atau pola yang diinginkan dalam *database* besar untuk membantu pengambilan keputusan diwaktu yang akan datang. Pola-pola ini dikenali oleh perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasan yang kemudian dapat dipelajari dengan lebih teliti, yang mungkin saja menggunakan perangkat pendukung keputusan yang lainnya.

Klasifikasi

Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi ada dua pekerjaan utama yang dilakukan, yaitu (1) pembangunan model sebagai prototipe untuk disimpan sebagai memori dan (2) penggunaan model tersebut untuk melakukan pengenalan/ klasifikasi/ prediksi pada suatu objek data lain agar diketahui di kelas mana objek data tersebut dalam model yang sudah disimpannya. [8]

Proses klasifikasi membentuk suatu model yang mampu membedakan data kedalam kelas-kelas yang berbeda berdasarkanaturan atau fungsi tertentu. Model itu sendiri bisa berupa aturan"jika-maka", berupa pohon keputusan, atau formula matematis. Proses klasifikasi dapat dilihat pada Gambar 1.



Gambar 1 Proses Klasifikasi

Naive Bayes

Algoritma *Naive Bayes* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. *Bayesian classification* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. *Bayesian classification* didasarkan pada teorema bayes yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network*. *Bayesian classification* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam *database* dengan data yang benar [9]. Persamaan *Naive Bayes* dapat dilihat pada persamaan (1).

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad (1)$$

Dengan :

X = Kriteria suatu kasus berdasarkan masukan

C_i = kelas solusi pola ke-i, dimana i adalah jumlah label kelas

$P(C_i|X)$ = Probabilitas kemunculan label kelas C_i dengan kriteria masukan X

$P(X|C_i)$ = Probabilitas kriteria masukan X dengan label kelas C_i

$P(C_i)$ = Probabilitas label kelas C_i

K-Nearest Neighbor

Algoritma *Nearest Neighbor* (kadang disebut K *Nearest Neighbor* atau K-NN) merupakan algoritma yang melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data yang lain. Dekat atau jauhnya lokasi (jarak) biasanya dihitung berdasarkan jarak *Euclidean* [10] dengan rumus seperti persamaan (2).

$$d(X_i, X_j) = \sqrt{\sum_{l=1}^N (diff(X_{il}, X_{jl}))^2} \quad (2)$$

Dengan :

X_{il} = data testing ke-i pada variabel ke-l

X_{jl} = data training ke-i pada variabel ke-l

$d(X_i, X_j)$ = jarak

N = dimensi data variabel bebas

$diff(X_{il}, X_{jl})$ = *difference* atau ketidaksamaan

C4.5

Algoritma C4.5 untuk membangun pohon keputusan menurut Larose [11] yaitu :

1. Pilih atribut sebagai root
2. Buat cabang untuk masing-masing nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atributatribut yang ada. Untuk menghitung gain digunakan persamaan (3) :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (3)$$

Dengan :

$\{S_1, S_2, S_3, \dots, S_n\}$ = partisi S, sesuai dengan nilai atribut A

A = Atribut

n = jumlah partisi atribut A

$|S_i|$ = jumlah kasus pada partisi S_i

$|S|$ = jumlah kasus dalam S

Sedangkan perhitungan nilai entropy menggunakan persamaan (4) :

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i \quad (4)$$

Dengan :

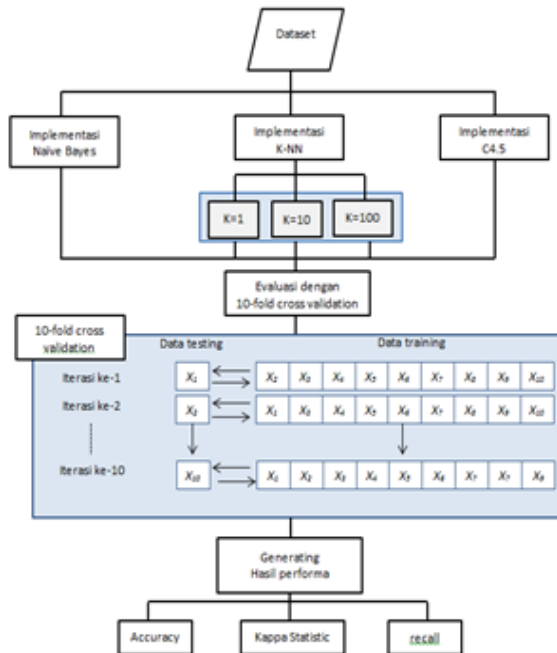
S = Himpunan kasus

n = jumlah kasus pada partisi S

p_i = proporsi S_i terhadap S

HASIL DAN PEMBAHASAN

Penelitian ini dilakukan dengan mengimplementasikan dataset benih gandum dengan beberapa metode diantaranya metode *naïve bayes*, K-NN dengan k=1, 100 dan k=1000 serta menggunakan metode C4.5. Hasil implementasi selanjutnya akan dievaluasi menggunakan 10-fold *cross validation*. Teknik validasi ini dimulai dari membagi subset dataset menjadi 10 subset (x_1, x_2, \dots, x_{10}). Kemudian ketika x_1 menjadi data validasi atau data testing dan x_2, x_3, \dots, x_{10} akan menjadi data training kemudian data tersebut diiterasi secara berurutan sampai dengan data testing x_{10} dengan data training x_1, x_2, \dots, x_9 [4].



Gambar 4 Gambaran umum Penelitian

Dataset

Penelitian ini menggunakan dataset publik yaitu *seeds dataset* berupa data nominal *real* yang diambil dari UCI *machine learning repository*. Dataset ini sudah terbukti dan banyak digunakan peneliti untuk pemodelan dan aplikasi klasifikasi

benih gandum. Jumlah *records* data yang digunakan pada proses klasifikasi ini sebanyak 210 *records* dengan atribut sebanyak 7 yaitu :

1. Area A,
2. Perimeter P,
3. Compactness $C = 4 * \pi * A / P^2$,
4. Length of kernel
5. Width of kernel
6. Asymmetry coefficient
7. Length of kernel groove

Data yang diperoleh dari UCI *machine Learning* untuk benih gandum sudah tidak ada redundansi data, *noisy* maupun *missing value* sehingga tidak perlu dilakukan *preprocessing* data terlebih dahulu.

Hasil Penelitian

Penelitian yang dilakukan menggunakan metode *supervised learning* yaitu pembelajaran terarah berdasarkan data historis terkait data benih gandum. Penelitian dimulai dengan mengimport dataset ke *machine learning Rapid miner*. Setelah dataset diimport selanjutnya dilakukan uji coba dengan mengimplementasikan beberapa model algoritma klasifikasi seperti *Naïve Bayes*, KNN, *Decision Tree C4.5*. Setelah diimplementasikan, selanjutnya akan dilakukan pengukuran nilai *accuracy*, *kappa statistic* dan *recall*. Hasil implementasi dapat dilihat pada tabel 1.

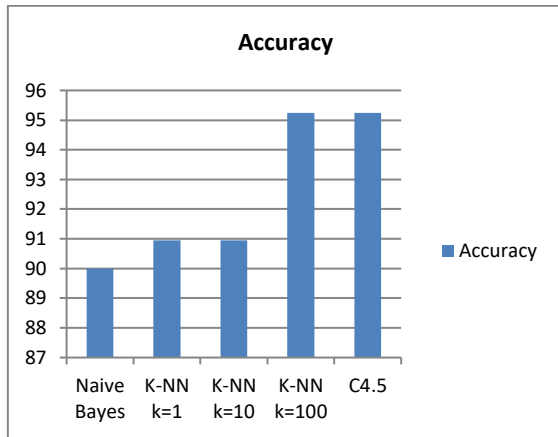
Tabel 1 Perbandingan Algoritma Klasifikasi

Model Algoritma	Parameter Setting	Accuracy	Kappa Statistik	Recall
Naïve Bayes	-	90	0,85	90
K-NN	K=1	90,95	0,864	90,95
	K=10	90,95	0,864	90,95
	K=100	95,24	0,929	95,24
C4.5	-	95,24	0,929	95,24

Setelah mengimplementasikan beberapa algoritma klasifikasi terhadap data set, tahapan selanjutnya dari penelitian ini yaitu membandingkan algoritma mana yang memiliki performa terbaik berdasarkan hasil pengukuran nilai *accuracy*, *kappa statistic* dan *recall*.

Nilai Accuracy

Accuracy dalam klasifikasi adalah persentase ketepatan *record* data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi [10]. Semakin tinggi nilai akurasi, maka dapat dikatakan semakin efektif model algoritma klasifikasi tersebut [12]. Nilai *accuracy* dari ketiga metode dapat dilihat pada gambar 2.

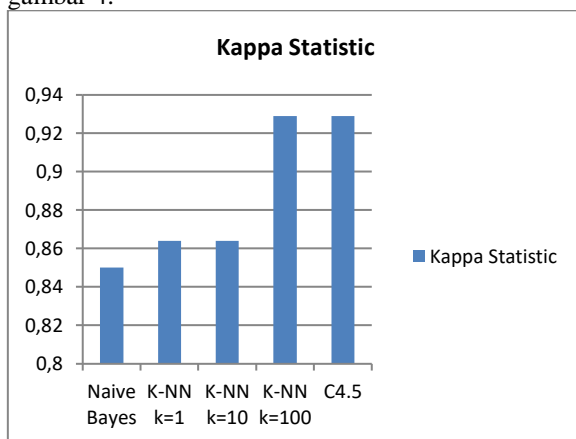


Gambar 3 Nilai accuracy

Berdasarkan Gambar 3, dapat dilihat bahwa diantara ketiga metode yang digunakan dalam penelitian ini, metode K-NN dengan implementasi $k=100$ dan C4.5 memiliki akurasi yang sama. Kedua metode ini memiliki akurasi yang paling tinggi yaitu 95,24% jika dibandingkan dengan *naive bayes* maupun K-NN dengan implementasi $k=1$ maupun $k=10$. Akurasi terendah ditemukan pada metode *Naive Bayes* dengan nilai 90%.

Nilai Kappa Statistik

Kappa Statistic adalah analisis statistik yang didasarkan pada kecocokan penafsiran atau derajat kesepakatan (*degree of agreement*) untuk data kualitatif. Pada dasarnya *kappa statistic* menunjukkan analisis diantara kelas-kelas yang berbeda. Semakin tinggi *kappa statistic*, maka akan dipertimbangkan sebagai performa yang mempunyai kriteria atau kinerja yang bagus [12]. Nilai *kappa statistic* dari ketiga metode dapat dilihat pada gambar 4.



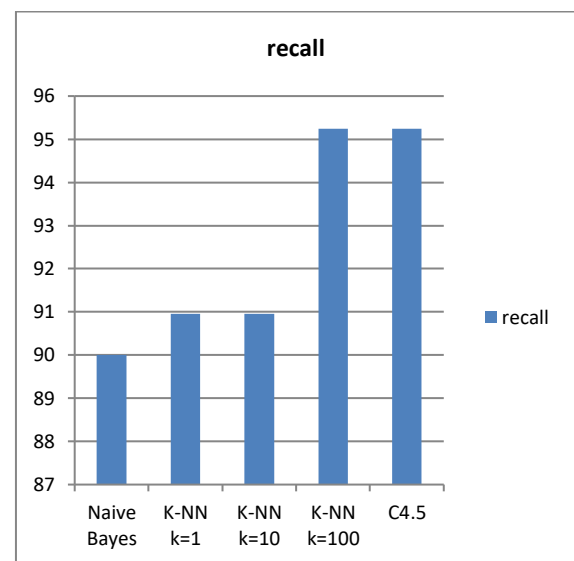
Gambar 4 Nilai kappa statistic

Berdasarkan hasil uji *kappa statistic* seperti yang terlihat pada gambar 4, dapat dilihat nilai *kappa statistic* juga menunjukkan bahwa penggunaan metode K-NN dengan implementasi $k=100$ memiliki hasil yang sama dengan metode C4.5. Nilai kapabilitas terbaik secara statistik yang

diperoleh dari dua metode tersebut menunjukkan nilai tertinggi jika dibandingkan dengan metode *naive bayes* maupun K-NN dengan implementasi $k=1$ maupun $k=10$. Nilai hasil uji *kappa statistic* tertinggi yang dicapai dengan metode C4.5 dan K-NN dengan $k=100$ yaitu 0,929. Hasil nilai *kappa statistic* terendah yang diperoleh setelah pengujian menunjukkan hasil yang sama dengan perhitungan akurasi yaitu dengan metode *naive bayes* sebesar 0,85.

Nilai Recall

Recall adalah perbandingan jumlah dokumen relevan yang terambil sesuai dengan *query* yang diberikan dengan total kumpulan dokumen yang relevan dengan *query* [13]. Istilah *recall* dibidang sistem temu kembali informasi (*information retrieval*) berkaitan dengan kemampuan menemukan kembali informasi yang sudah tersimpan. Nilai *recall* dari penelitian ini dapat dilihat pada gambar 5.



Gambar 5 Nilai recall

Berdasarkan gambar 5, dapat dilihat bahwa nilai yang diperoleh melalui uji *recall*, memiliki hasil yang sama dengan penggunaan uji akurasi. Melalui uji *recall*, metode C4.5 dan K-NN dengan $k=1$ juga memiliki nilai tertinggi seperti halnya uji *kappa statistic* maupun uji akurasi. Nilai yang dihasilkan melalui kedua metode tersebut sama seperti uji akurasi yaitu 95,24. Nilai uji *recall* terendah ditemukan pada metode *naive bayes* dengan nilai 90.

Berdasarkan pengukuran nilai *accuracy*, *kappa statistic* maupun *recall* yang sudah dilakukan, menunjukkan hasil yang sama terkait performa tertinggi maupun terendah. Performa tertinggi ditemukan pada metode C4.5 dan K-NN dengan implementasi $k=100$ sedangkan performa terendah ditemukan pada metode *naive bayes* dimana ketiga

metode tersebut diimplementasikan pada dataset dengan jumlah atribut sebanyak 7 dan 210 records.

5. KESIMPULAN

Berdasarkan hasil dan pembahasan yang sudah di paparkan sebelumnya, diperoleh kesimpulan bahwa diantara ketiga algoritma yang digunakan untuk benih gandum dengan jumlah atribut 7 dan 210 records, algoritma C4.5 dan K-NN dengan implementasi k=100 memiliki performa terbaik berdasarkan nilai *accuracy*, *kappa statistic* dan *recall*. Nilai *accuracy* yang diperoleh sebesar 95,24%, *kappa statistic* sebesar 0,929 dan *recall* 95,24. Sedangkan performa terendah yaitu algoritma *naive bayes* dengan *accuracy* 90%, *kappa statistic* 0,85 dan *recall* 90.

6. DAFTAR PUSTAKA

- [1] Soeranto H, 2012, *Riset dan Pengembangan Tanaman Sorghum dan Gandum untuk Ketahanan Pangan*, Badan Atom Nasional.
- [2] BPS, 2011, *Import Gandum*, Badan Pusat Statistik, Jakarta.
- [3] http://www.academia.edu/30318954/Identifikasi_Benih_Gandum_Dengan_Jaringan_Saraf Arry Andriana.
- [4] Yusa, M., Utami, E., dan Luthfi, E.T., 2016, *Analisis Komparatif Evaluasi Performa Algoritma Klasifikasi pada Readmisi Pasien Diabetes*, Thesis, STMIK AMIKOM, Yogyakarta.
- [5] Sartika, D., Sensuse, D.I., 2017, Perbandingan algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian, *Jatsi Vol.1 No.2 Maret 2017*, ISSN:1978-1520 hal 151-161.
- [6] Fakhurrifqi, M., Wardoyo, R., 2013, Perbandingan Algoritma Nearest Neighbor, C4.5 dan LVQ untuk klasifikasi kemampuan mahasiswa, *IJCCS, Vol.7 No.2 July 2013*, ISSN:1978-1520 hal 145-154.
- [7] Hermawati, Fajar A, 2013, *Data Mining*, Penerbit Andi, Yogyakarta.
- [8] Prasetyo, E., 2012, *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta, ANDI Yogyakarta.
- [9] Kusriani & Emha, T.L., 2009, *Algoritma Data Mining*, Yogyakarta, Andi Offset.
- [10] Kamber, M., & Han, J., 2006, *Data mining; Concepts and Techniques Second Edition*, San Francisco, Morgan Kaufmann Publishers.
- [11] Larose, Daniel, T., 2005, *Discovering Knowledge in Data: an Introduction to Data Mining*, John Wiley and Sons, USA.
- [12] Mittal, P., & Gill, N. S., 2014, A Comparative Analysis Of Classification Techniques On Medical Data Sets. *IJRET: International Journal of Research in Engineering and Technology, Volume: 03 Number: 06*, pp. 454-460.
- [13] Kurniawan, D., 2010, Evaluasi sistem temu kembali informasi model ruang vector dengan pendekatan user judgement, *J. Sains MIPA, Desember 2010, Vol. 16, No. 3*, Hal.: 155 - 162 .