

**ANALISIS PERBANDINGAN ALGORITMA C4.5 DAN ALGORITMA NAÏVE BAYES
UNTUK PREDIKSI KELULUSAN MAHASISWA
(STUDI KASUS : PRODI TEKNIK INFORMATIKA UNIVERSITAS MUHAMMADIYAH JEMBER)**

¹M. Hairul Umam (1310651100)

²Victor Wahanggara, S.T, M.Kom

³Triawan Adi Cahyanto, M.Kom, ⁴Lutfi Ali Muharom, S.Si., M.Si

Jurusan Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember

Email : umam2002@gmail.com

ABSTRAK

Universitas Muhammadiyah Jember adalah jenjang pendidikan yang dianggap sebagai gerbang terakhir bagi pelajar untuk menimba ilmu sebelum akhirnya melibatkan diri dalam persaingan kerja. Jumlah mahasiswa yang lulus tepat waktu menjadi indikator efektifitas dari sebuah perguruan tinggi baik negeri dan swasta. Program Sarjana (S1) jurusan Teknik Informatika Universitas Muhammadiyah Jember adalah pendidikan akademik setelah pendidikan menengah, yang memiliki beban studi sekurang-kurangnya 144 SKS yang dijadwalkan untuk 8 (delapan) semester dan dapat di tempuh dalam waktu kurang dari 8 semester dan paling lama 10 semester. Penelitian dalam hal memprediksi kelulusan mahasiswa telah banyak dilakukan. C4.5 adalah satu dari sebagian algoritma dalam metode decision tree yang mengubah data menjadi pohon keputusan, kemudian dapat disimpulkan menjadi rule-rule. Sedangkan Naïve Bayes adalah metode yang menghitung probabilitas dari tingkat kemunculan data yang satu terhadap data yang lainnya. berdasarkan latar belakang tersebut peneliti bermaksud membandingkan dua metode tersebut yang akan di presentasikan untuk memprediksi kelulusan mahasiswa angkatan 2013. Berdasarkan hasil pengujian dengan mengukur kinerja kedua metode tersebut menggunakan metode pengujian Confusion Matrix dalam tool weka diketahui bahwa C4.5 memiliki nilai akurasi yang lebih baik yakni sebesar 100%, dengan akurasi 100%, Presisi 100% dan Recall 100%, sedangkan Naive Bayes memiliki nilai akurasi 89%, Presisi 94% dan Recall 100%. Kemudian berdasarkan perbandingan kinerja kedua metode tersebut, metode dengan pencapaian nilai akurasi terbaik akan presentasikan untuk pengujian prediksi kelulusan mahasiswa angkatan 2103. Sehingga di dapat hasil prediksi semakin meningkat pada angkatan 2013 dari evaluasi tahun – tahun sebelumnya Dengan jumlah lulus 144 mahasiswa dengan total mahasiswa angkatan 2013 sebanyak 263 mahasiswa, hasil ini dilihat dan di bandingkan dengan data historis angkatan 2010, 2011 dan 2012.

Kata kunci : C4.5, Naïve Bayes, Tool Weka, Prediksi Kelulusan Mahasiswa

BAB I PENDAHULUAN

1.1 Latar Belakang

Seiring dengan berkembangnya teknologi penyimpanan data dan informasi, semakin berkembang pula kemampuan seseorang dalam mengumpulkan dan mengolah data. Data yang terkumpul dan berukuran besar tersebut merupakan aset yang dapat dimanfaatkan untuk dianalisis yang hasilnya berupa pengetahuan atau informasi berharga untuk masa mendatang. Tidak hanya dunia bisnis, namun instansi seperti perguruan tinggi lainnya juga mengalami penumpukan data.

Jurusan Teknik Informatika Universitas Muhammadiyah Jember adalah program pendidikan perguruan tinggi yang memiliki beban studi sekurang - kurangnya 144 SKS (satuan kredit semester) yang dijadwalkan untuk 8 semester dan dapat ditempuh dalam waktu kurang dari 8 semester. Dari data kelulusan yang diperoleh dari sekretariat Jurusan Teknik Informatika Program Sarjana (S1) angkatan 2010, 2011, 2012, dan 2013 hanya Sedikit mahasiswa yang lulus dalam 8 semester. Hal ini menunjukkan bahwa masih banyak mahasiswa Program Sarjana (S1) reguler di Jurusan Teknik Informatika yang menempuh lama studi lebih dari 8 semester.

Melihat kondisi tersebut diperlukan penelitian untuk menggali data yang dimiliki oleh Jurusan Teknik Informatika

Universitas Muhammadiyah Jember. Data yang akan dimanfaatkan dalam penelitian ini adalah data mata kuliah umum, yaitu Pengantar Teknologi Informasi, Programan Terstruktur, Struktur Data, Sistem Informasi, Analisis Desain Berorientasi Objek, Programan Berorientasi Objek, Jaringan Komputer, Praktikum Jaringan Komputer, Programan Jaringan, Praktikum Programan Jaringan Komputer, dan nilai IPK semester 1-6, dan data nilai akademik mahasiswa baik yang sudah lulus, belum lulus atau sedang menempuh yang digunakan sebagai data training dan data testing untuk memprediksi masa studi mahasiswa. Penelitian ini dirasa perlu karena jika masa studi mahasiswa dapat diketahui lebih dini, maka pihak jurusan dapat melakukan tindakan-tindakan yang dirasa perlu supaya mahasiswa dapat lulus tepat waktu sekaligus meningkatkan kualitas jurusan itu sendiri. Dari data masalah dan pengalaman historis yang sudah terjadi dibutuhkan suatu teknik dan metode klasifikasi yang merupakan salah satu teknik dari klasifikasi untuk menganalisis data di Jurusan Teknik Informatika. untuk menganalisa kasus tersebut. Algoritma C4.5 dan Naive Bayes adalah salah satu metode klasifikasi yang dapat di uji untuk memproses dan memprediksi kemungkinan mahasiswa yang lulus lebih dari 8 semester dan lulus tepat waktu.

Algoritma C4.5 merupakan algoritma pengembangan dari algoritma ID3 yang digunakan untuk membentuk pohon keputusan. Pohon keputusan dapat membagi kumpulan data yang besar menjadi himpunan-himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan (Sofiana, 2012).

Naive Bayes merupakan teknik prediksi berbasis probabilitas sederhana yang berdasar pada penerapan aturan bayes dengan asumsi ketidaktergantungan yang kuat. Selain itu naive bayes juga dapat menganalisa variabel-variabel yang paling mempengaruhinya dalam bentuk peluang (Prasetyo, 2012).

Menurut Wu, et al. (2007), terdapat sejumlah algoritma fungsi data mining yang dipilih menjadi Top 10 algorithm data mining. Hasilnya, diperoleh bahwa algoritma klasifikasi menjadi mayoritas dalam Top 10 algorithm. Adapun algoritma klasifikasi yang terbaik dalam Top 10 algorithm data mining adalah salah satunya pohon keputusan C4.5, K-Nearest Neighbor (K-NN), Naive Bayes, dan Classification and Regression Trees (CART).

Galathiya, et al. (2012), melakukan penelitian dengan melakukan pengembangan pada algoritma pohon keputusan C4.5 dan melakukan perbandingan terhadap algoritma Naive Bayes, Sequential Minimal Optimization (SMO), dan multilayer perceptron dengan menggunakan sejumlah data sets dan pengujian akurasi berdasarkan kepada Correctly Classified Instances dan Relative Absolute Error. Hasilnya, algoritma pohon keputusan C4.5 yang dikembangkan memiliki tingkat akurasi sekitar 1-3% lebih baik pada 7 data sets dari total 9 data sets.

Othman & Yau (2007) melakukan penelitian berjudul "Comparison of Different Classification Techniques using WEKA for Breast Cancer". Metode pengujian yang digunakan berupa percentage split sebesar 75%, artinya 75% data digunakan sebagai data training dan 25% sisanya sebagai data testing. Pengujian akurasi berdasarkan kepada Correctly Classified Instances, Incorrectly Classified Instances, dan Time Taken. Hasilnya, algoritma Naive Bayes memiliki tingkat akurasi tertinggi dan waktu pembangunan model tercepat dibandingkan dengan Radial Basis Function, Decision Tree and Pruning, Single Conjunctive Rule Learner, dan Nearest Neighbors Algorithm.

Berdasarkan masalah tersebut di atas, penulis mencoba melakukan penelitian untuk menganalisa dan membandingkan akurasi dua metode algoritma yang merupakan bagian dari Top 10 algorithm data mining, yaitu algoritma C4.5 dan Naive Bayes pada sejumlah data sets. Adapun sumber data sets adalah data mahasiswa jurusan teknik informatika (Universitas Muhammadiyah Jember) dimana setiap data sets memiliki jumlah data (instances) dan jumlah atribut (attributes) yang berbeda. Perbandingan pengukuran akurasi yang di uji akan menghasilkan informasi yang efisien, akurat dan rill, Atas dasar inilah, penulis tertarik untuk mengambil judul tugas akhir : “ *Analisis*

Perbandingan Algoritma C4.5 Dan Algoritma Naive Bayes Untuk Prediksi Kelulusan Mahasiswa (Studi Kasus : Prodi Teknik Informatika, Universitas Muhammadiyah Jember) ”.

1.2 Perumusan Masalah

Permasalahan yang dijadikan objek penelitian pada tugas akhir ini berdasarkan latar belakang yang telah dipaparkan adalah :

1. Bagaimana Memproses algoritma C4.5 dan algoritma Naive Bayes berdasarkan implementasi algoritma.
2. Bagaimana membandingkan performance algoritma C4.5 dan algoritma Naive Bayes berdasarkan implementasi algoritma ?
3. Berapa persentase akurasi yang didapat dalam implementasi algoritma C4.5 dan algoritma Naive Bayes ?

1.3 Batasan Masalah

Untuk lebih memfokuskan pengerjaan penelitian ditetapkan pembahasan hanya dibatasi pada :

1. Parameter yang digunakan 11 parameter, yaitu mata kuliah seperti Pengenalan Teknologi Informasi, Pemrograman Terstruktur, Struktur Data, Sistem Informasi, Analisis & Desain Berorientasi Objek, Pemrograman Berorientasi Objek, Jaringan Komputer, Praktikum Jaringan Komputer, Programan Jaringan, Praktikum Programan Jaringan Komputer, dan IPK semester 1 – 6.
2. Data yang di gunakan sebagai pengujian yaitu data mahasiswa angkatan 2010, 2011, 2012, dan 2013 jurusan teknik Informatika Universitas Muhammadiyah Jember, traning data angkatan 2010, 2011 2012, dan 100 data angkatan 2013 sebagai testing atau data uji.
3. Metode yang di gunakan Algoritma C4.5 dan Algoritma Naive Bayes.
4. Hasil akhir pengujian adalah perbandingan dan akurasi berupa informasi yang Efisien dan Rill.

1.4 Tujuan Penelitian

Tujuan penelitian pada tugas akhir ini adalah

1. Menganalisa performance algoritma C4.5 dan algoritma Naive Bayes.
2. Menganalisa implementasi algoritma C4.5 dan algoritma Naive Bayes.
3. Melakukan perbandingan akurasi algoritma C4.5 dan algoritma Naive Bayes.

1.5 Manfaat Penelitian

Manfaat penelitian dibedakan menjadi dua jenis yaitu manfaat secara teoritis dan manfaat secara praktis.

1.5.1 Manfaat Teoritis

Penelitian ini dapat digunakan sebagai referensi bagi pengembangan penelitian sejenis dimasa yang akan datang.

1.5.2 Manfaat secara Praktis

a. Manfaat bagi akademik

Bagi akademik, manfaat yang dapat diperoleh dari proses pembuatan proposal skripsi ini adalah berupa sumbangsih laporan buku Tugas Akhir, yang di harapkan mampu digunakan dengan sebaik-baiknya, dan sebagai tambahan buku di perpustakaan, dan juga bisa lebih di kembangkan untuk masa yang akan datang.

b. Bagi Mahasiswa

Bagi mahasiswa, Tugas Akhir ini dapat dijadikan sebagai sumber referensi maupun pedoman bagi mahasiswa lain yang mengambil penelitian sejenis.

c. Bagi Universitas

Mampu memprediksi masa studi mahasiswa pada program sarjana (S1) Jurusan Teknik Informatikan (atau lainnya).

BAB II TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Sebelum melakukan penelitian penulis terlebih dahulu melakukan tinjauan pustaka dari penelitian lain yang berkaitan dengan prediksi masa studi mahasiswa, penelitian ini bukanlah penelitian yang pertama kalinya. Sebelumnya sudah ada penelitian yang membahas mengenai klasifikasi penjurusan sekolah menengah atas menggunakan metode algoritma Naive Bayes dan algoritma C4.5. Berikut adalah beberapa penelitian yang terkait dengan masalah tersebut.

2.1.1 Penelitian oleh (Claudia Clarentina C.) Tahun 2014

Penelitian yang pertama adalah penelitian yang dilakukan oleh Claudia Clarentina Ciptohartono yang berjudul "Algoritma Klasifikasi Naive Bayes untuk Menilai Kelayakan Kredit. Metode yang digunakan pada penelitian ini adalah Algoritma Naive Bayes. Data yang digunakan adalah data dari nasabah perusahaan BCA Finance Jakarta tahun 2013. Tool yang digunakan untuk implementasi sistem adalah matlab. Penelitian ini menghasilkan bukti bahwa algoritma Naive Bayes bisa diterapkan dalam pemberian kelayakan kredit pada BCA Finance Jakarta. Tahap pengolahan data awal dapat menghasilkan akurasi yang tinggi dan menghasilkan akurasi akhir yang Excellent. Dalam penelitian kelayakan kredit ini pada data awal dengan melakukan pre-processing mendapatkan akurasi sebesar 85.57%, tetapi jika data awal dilakukan pengolahan dan dengan melakukan pre-processing dapat menghasilkan akurasi sebesar 92.53%. Jadi Algoritma Naive Bayes akan lebih unggul dalam pemberian penilian kelayakan kredit jika dilakukan proses pengolahan data awal, meskipun algoritma Naive Bayes adalah algoritma yang sanggup menangani data yang hilang.

2.1.2 Penelitian oleh (Arief Jananto) 2014

Penelitian yang kedua adalah penelitian yang dilakukan oleh Arief Jananto. Penelitian yang berjudul "Algoritma Naive Bayes untuk Mencari Perkiraan Waktu 7 Studi Mahasiswa" Metode yang digunakan pada peneliti ini adalah Algoritma Klasifikasi Naive Bayes. Data yang diperoleh adalah data yang berkaitan dengan keterangan diri dari

mahasiswa UNISBANK dan data nilai semua mata kuliah mahasiswa lulusan tahun 2004-2007. Kesimpulan dari penelitian tersebut adalah sebagai berikut : a. Ketepatan masa studi mahasiswa diprediksi berdasarkan latar belakang sekolah sebelumnya dan data akademik, serta pribadi saat berada diperguruan tinggi. b. Dalam memprediksi ketepatan studi mahasiswa menggunakan data training dan testing dengan memanfaatkan fungsi prediksi dari teknik data mining yaitu menggunakan algoritma klasifikasi Naive Bayes. c. Fungsi klasifikasi yang digunakan dalam prediksi ketepatan masa studi mahasiswa ini mempunyai tingkat kesalahan sebesar 20% sampai 34%, kesalahan tersebut dapat dipengaruhi oleh jumlah data training dan testing serta tingkat konsistensi data yang digunakan

2.2.3. Klasifikasi

Klasifikasi merupakan salah satu tugas yang penting dalam data mining. mengorganisasikan serta mengelompokan data ke dalam kelas-kelas yang berbeda merupakan tujuan utama dari klasifikasi. Definisi pengklasifikasian adalah sebuah fungsi yang bersifat prediksi dan menggolongkan data item tertentu ke dalam sebuah kelas. Sebuah pengklasifikasian dibuat dari sekumpulan data latih dengan kelas yang telah ditentukan dan dikenal ciri-cirinya sebelumnya. Performa pengklasifikasian biasanya diukur dengan ketepatan.

2.2. Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Dan mereka juga dapat diekspresikan dalam bentuk bahasa basis data seperti Structured Query Language untuk mencari record pada kategori tertentu. pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variable input dengan sebuah variable target. Algoritma C4.5 ini mempunyai input berupa training samples dan samples. Training samples berupa data contoh yang akan digunakan untuk membangun sebuah pohon yang telah diuji kebenarannya. Sedangkan samples merupakan field-field data yang nanti akan kita gunakan sebagai parameter dalam melakukan klasifikasi data. Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut.

- Pilih atribut yang akan digunakan sebagai akar
- Buatlah sebuah cabang untuk setiap nilai
- Bagilah kasus dalam sebuah cabang
- Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama. Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung

gain digunakan rumus seperti yang tertera dalam persamaan 1 berikut.

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

Sehingga di sederhanakan :

$$Entropy(S) \equiv -p_+ \log_2 p_+ - p_- \log_2 p_-$$

Keterangan :

- S : himpunan kasus
- A : Atribut
- n : jumlah partisi atribut
- | Si | : jumlah kasus pada partisi ke -i
- | S | : jumlah kasus dalam S

Sementara entropy dapat dilihat pada persamaan 2 berikut.

$$Gains(S, A) = Entropy - \sum_{i=1}^n \frac{|Si|}{|S|} * Entropy(Si)$$

$$Gain(S, A) = Entropy(S) - \sum \frac{|Sv|}{|S|} Entropy(Sv)$$

Keterangan :

- S : Himpunan kasus
- A : Fitur
- n : Jumlah partisi S
- pi : Proporsi dari Si terhadap S

Untuk mengetahui analisis prediksi masa studi mahasiswa lulus tepat waktu dan tidak tepat waktu adalah dengan melakukan klasifikasi dari kumpulan data nilai mata kuliah umum tersebut. Salah satu metode klasifikasi adalah dengan membuat pohon keputusan. Berikut penjelasan lebih terperinci mengenai tiap-tiap langkah dalam pembentukan pohon keputusan dengan menggunakan algoritma C4.5 untuk menyelesaikan permasalahan.

2.3 Algoritma Naive Bayes

Klasifikasi Bayes adalah klasifikasi statistik yang dapat memprediksi kelas suatu anggota probabilitas. Algoritma ini memanfaatkan teori probabilitas yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas dimasa depan berdasarkan pengalaman dimasa sebelumnya. Dua kelompok peneliti, satu oleh Pantel dan Lin, dan yang lain oleh Microsoft Research memperkenalkan metode statistic *bayesianzi*. Tetapi yang membuat Naive Bayesian ini populer adalah pendekatan yang dilakukan oleh Paul Graham.

Naive Bayes berasumsi bahwa efek dari suatu pada kelas yang diberikan adalah independen terhadap nilai atribut yang lainnya. Asumsi ini biasa disebut dengan class conditional independence. Itu dibuat untuk menyederhanakan komputasi yang terkait dan dalam hal ini disebut sebagai "naive".(Rezdy Anugrah Perdana hal.7)

Naive Bayes Classifier (NBC) merupakan sebuah pengklasifikasian probabilitas sederhana yang mengaplikasikan Teorema Bayes dengan asumsi ketidak tergantungan (independent) yang tinggi. Keuntungan penggunaan NBC adalah metode ini hanya membutuhkan jumlah data pelatihan (training data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. (Rezdy Anugrah Perdana hal.8)

Banyak aplikasi menghubungkan antara atribut set dan variable kelas yang non deterministic. Dengan kata lain, label kelas test record tidak dapat diprediksi dengan peristiwa tertentu meskipun atribut set identik dengan beberapa contoh training. Situasi ini makin meningkat karena noisy data atau kehadiran factor confounding tertentu yang mempengaruhi klasifikasi tetapi tidak termasuk didalam analisis.

Naive Bayes klasifikasi adalah metode yang berdasarkan probabilitas dan Teorema Bayesian dengan asumsi bahwa setiap variable bersifat bebas (independence) dan mengamsusikan bahwa keberadaan sebuah fitur (variable) tidak ada kaitanya dengan beradaan fitur (variable) yang lain. Naive Bayes adalah model penyederhanaan dari metode bayes. Naive bayes inilah yang digunakan di dalam machine learning sebagai metode untuk mendapatkan hipotesis untuk suatu keputusan. (Basuki, 2006) Salah satu penerapan Teorema Bayes adalah dalam klasifikasi *Naive Bayes* sebagai berikut. (Kusumadewi, 2003)

$$p(H_i|E) = \frac{p(E|H_i) * (p(H_i))}{\sum_{k=1}^n p(E|H_k) * (p(H_k))}$$

- P (H_i | E) = Probabilitas hipotesis H_i benar jika diberikan evidence (fakta) E.
- P (E | H_i) = Probabilitas munculnya evidence (fakta) E jika diketahuhipotesis H_i benar
- P(H_i) = Probabilitas hipotesis H_i (menurut hasil sebelumnya) tanpa memandang evidence (fakta) apapun.
- n = jumlah hipotesis yang mungkin

Naive Bayes Classifier umumnya memiliki karakteristik sebagai berikut :

- a. Kokoh untuk titik noise yang diisolasi seperti titik yang dirata – ratakan ketika mengestimasi peluang bersyarat data. Naive Bayes Classifier dapat menangani missing value dengan megabaikan contoh selama pembuatan model dan klasifikasi.
- b. Kokoh untuk atribut tidak relevan, jika X_i adalah atribut yang tidak relevan, maka p(X_i | Y) menjadi hampir didistribusikan seragam. Peluang kelas bersyarat untuk X_i tidak berdampak pada keseluruhan perhitungan peluang posterior.
- c. Atribut yang dihubungkan dapat menurunkan performa Naive Bayes Classifier karena asumsi independen bersyarat tidal lagi menangani atribut tersebut.

2.4 Precision, Recall dan Akurasi

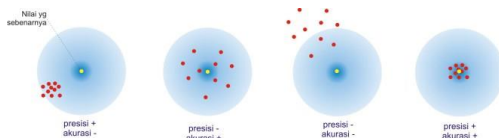
Dalam “dunia” pengenalan pola (pattern recognition) dan temu kembali informasi (information retrieval), precision dan recall adalah dua perhitungan yang banyak digunakan untuk mengukur kinerja dari sistem / metode yang digunakan. **Precision** adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. **Recall** adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. **Accuracy** didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai actual. Tahap berikutnya melakukan pengujian terdapat tiga bagian

1. Akurasi didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai actual.
2. Presisi tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem.
3. Recall tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi.

Sedangkan di “dunia lain” seperti dunia statistika dikenal juga istilah *accuracy*. Akurasi didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai actual. Ilustrasi berikut ini memberikan gambaran perbedaan antara akurasi dan presisi.

Tabel 2.10 Presisi, Akurasi dan Recall

		Nilai Sebenarnya	
		TRUE	FALSE
Nilai Prediksi	TRUE	TP (True Positive) Corect result	FP (False Positive) Unexpected result
	FALSE	FN (False Negative) Missing result	TN (True Negative) Corect result



Gambar 2.5 Perbdaan Akurasi, Presisi dan Recall

Secara umum presisi, recall dan akurasi dapat dirumuskan sebagai berikut:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

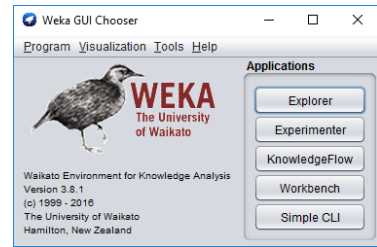
$$Presisi = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

2.5. Tool Weka Dalam Klasifikasi

Weka adalah aplikasi data mining open source berbasis Java. Aplikasi ini dikembangkan pertama kali oleh Universitas Waikato di Selandia Baru sebelum menjadi bagian dari Pentaho. Weka terdiri dari koleksi algoritma machine learning yang dapat digunakan untuk melakukan generalisasi /formulasi dari sekumpulan data sampling. Walaupun kekuatan Weka terletak pada algoritma yang makin lengkap dan canggih, kesuksesan data mining tetap terletak pada faktor pengetahuan manusia implementornya. Tugas pengumpulan data yang berkualitas tinggi dan pengetahuan

pemodelan dan penggunaan algoritma yang tepat diperlukan untuk menjamin keakuratan formulasi yang diharapkan.



Gambar 2.7 Halaman awal Weka

BAB III METODOLOGI PENELITIAN



Gambar 3.1. Tahapan Penelitian

3.1 Studi Literatur

Tahap studi literatur ini mempelajari tentang semua data dan informasi yang berkaitan dengan algoritma C4.5 dan Naif Bayes dan juga semua materi yang berhubungan dengan masalah yang akan dibahas, dalam penelitian ini refrensi diambil dari berbagai sumber, seperti buku, jurnal, *e-book*, serta sumber-sumber lain yang dinilai dapat memberi tambahan wawasan untuk penelitian ini.

3.2 Pengumpulan Data

Data yang digunakan untuk penelitian ini adalah data mahasiswa jurusan teknik informatika (universitas muhammadiyah jember) yang di dapat dari Pusat Data dan Informasi (PDI) angkatan 2010, 2011, 2012, dan 2013 sdata tersebut data yang sudah melalui tahapan pengolahan kembali oleh peneliti.

3.3 Preprocessing Data

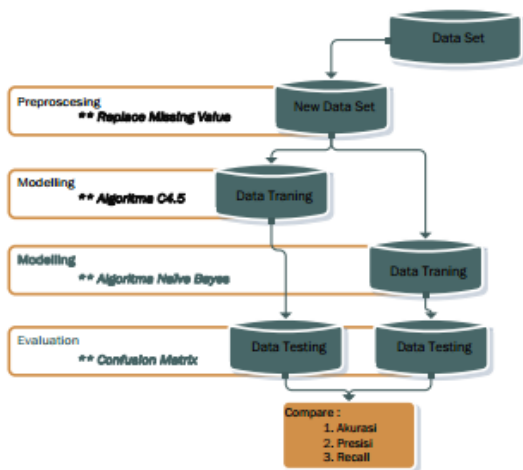
Berisi pemerosan data mulai data mentah, yang di olah dan di lakukan konversi data sampai data tersebut siap untuk di ujikan.



Gambar 3.3 Alur Pemrosesan Data

Konversi nilai :

- ✓ 0 – 2,50 = Kurang, 2,51 – 2,75 = Cukup, 2,76 – 3,49 = Baik, dan 3,50 – 4,00 = Baik Sekali
- ✓ A = A, B = B, C = C, C = D, E = D, T = D, dan 0 = D



Gambar 3.4 Model Pemrosesan Tool Weka

3.4. Implementasi dan Pengujian

Implementasi dilakukan untuk menyelesaikan masalah sesuai dengan perancangan proses pembentukan algoritma C4.5 dan Naive Bayes. Dan dilakukan pengujian dengan menginputkan data uji mahasiswa prodi teknik Informatika Universitas Muhammadiyah Jember angkatan 2013. Pengujian ini menerapkan metode C4.5 dan Naive Bayes Dalam Menentukan Masa Studi mahasiswa teknik Informatika angkatan 2013. Nilai akhir implementasi dan pengujian data yaitu nilai perbandingan dan akurasi 2 metode yg di tunjukan pada tabel sebagai berikut :

Tabel 3.12 Rencana Analisis Perbandingan Algoritma C4.5 & Naive Bayes

Algoritma	Perbandingan			
	Akurasi	Presisi	Recall	Time Taken
Algoritma C4.5	?	?	?	?
Algoritma Naive Bayes	?	?	?	?
Rata – Rata %	?	?	?	?

3.5. Analisis

Hasil dari pengujian tersebut dianalisis menggunakan pendekatan atau Teknik analisis data deskriptif merupakan teknik analisis yang dipakai untuk menganalisis data dengan mendeskripsikan atau menggambarkan data-data yang sudah dikumpulkan seadanya tanpa ada maksud membuat generalisasi dari hasil penelitian. Yang termasuk dalam teknik analisis data statistik deskriptif diantaranya seperti penyajian data kedalam bentuk grafik, tabel, presentase, frekwensi, diagram, grafik, mean, modus dll.

BAB IV HASIL DAN PEMBAHASAN

4.1. Analisis Spesifikasi

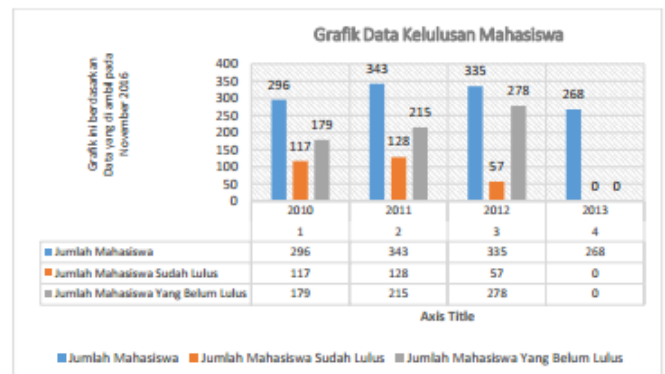
Pada analisis ini membahas tentang aspek – aspek yang di dinilai berpengaruh dalam pengujian data khususnya dalam Taken Time dalam peroses data, antara lain sebagai berikut :

Table. 4.1 Spesifikasi Hardware & Software

Hardware	Software
1. Processor : Intel® Core™ i3-3110M CPU @ 2.4GHz	1) Weka 3.8
2. Operating System : Windows 10 Pro	2) Visio 2013
3. Display Type : 14" HD WLED	
4. System Graphics: Integrated Graphic	
5. RAM : 2 GB (1x2096) RAM	
6. Hard Drive : 500GB HDD	
7. Battery : 4 Cells	
8. Weight : 2.2kg	
9. USB: 1x2.0; 2x3.0	
10. Web-cam/HDMI/DVD-SM/VGA	
11. LAN/BT/Card Reader/Speakers	

4.2. Analisis Data

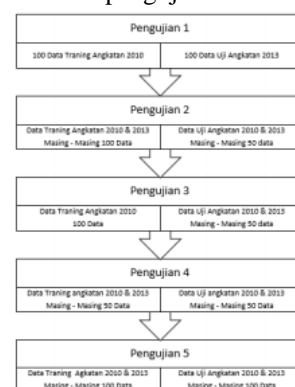
Data yang dianalisa dan diproses oleh peneliti adalah data mahasiswa angkatan 2010, 2011, 2012 dan 2013 Jurusan Teknik Informatika Universitas Muhammadiyah Jember, 100 Data Angkatan 2010 sebagai data training, dan 100 Data Angkatan 2013 Sebagai Testing dan data yang di gunakan sebagai pembelajaran atau bisa juga di sebut dengan data tranning untuk prediksi kelulusan mahasiswa adalah data angkatan 2010, 2011 dan 2012, kemudian data trestingnya adalah data angkatan 2013. Data tersebut sudah melalui tahapan beberapa proses pengelompokan dan konversi data yang bertujuan untuk mempermudah dalam Pemrosesan dan implementasi algoritma, dan juga dalam pengujian data nantinya akan di uji dengan beberapa kombinasi data yang sudah di skenarioikan oleh peneliti.



Grafik 4.1 Data Kelulusan Mahasiswa

4.3. Skenario Pengujian

Setelah data di proses dan siap untuk di uji, peneliti menggunakan beberapa skenario untuk pengujian data yang sudah di olah dan di proses untuk di ujian ke masing - masing algoritma yaitu algoritma C4.5 dan Algoritma Naive Bayes. Berikut gambar skenario pengujian data di bawah ini :



Gambar 4.3 Skenario Pengujian

4.4. Implementasi dan Pengujian Algoritma C4.5

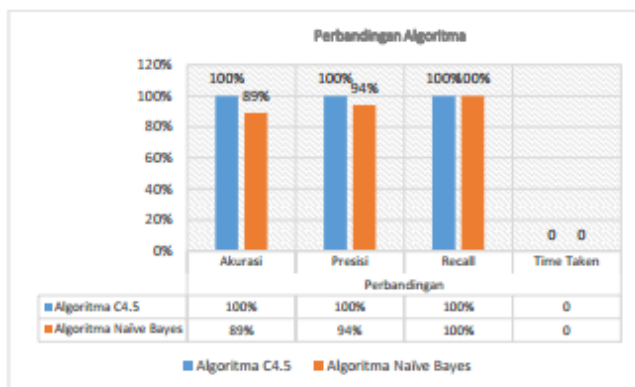
Table 4.7 Hasil Pengujian C4.5

Pengujian	C4.5			
	Presisi	Recall	Akurasi	Taken Tame
1	100%	100%	100%	0.02 (s)
2	100%	100%	100%	0 (s)
3	100%	100%	100%	0 (s)
4	100%	100%	100%	0 (s)
5	100%	100%	100%	5.05 (s)
Rata " %	100%	100%	100%	0 (s)

Table 4.13 Hasil Pengujian Naïve Bayes

Pengujian	Naïve Bayes			
	Presisi	Recall	Akurasi	Taken Tame
1	84%	100%	92%	0 (s)
2	96%	100%	98%	0 (s)
3	87%	100%	93%	0 (s)
4	80%	100%	90%	0,02 (s)
5	100%	100%	100%	0,03 (s)
Rata " %	89%	100%	94%	0 (s)

5.1. Perbandingan Algoritma C4.5 dan Naive Bayes



Grafik 4.2 Perbandingan Algoritma

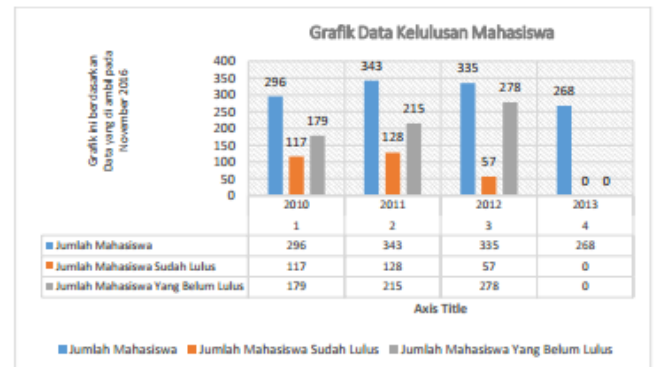
Dari hasil grafik 4.2 perbandingan algoritma dapat disimpulkan bahwa metode C4.5 lebih baik dan lebih akurat dari pada metode Naive Bayes dalam permasalahan ini. Metode J48 (decision tree) memang merupakan algoritma yang digunakan untuk pengelompokan class. Namun untuk waktu pemrosesan (running time) yang paling cepat yaitu Naive Bayes. Kemudian di lihat dari kualitas dan kuantitas dari algoritma C4.5 dan Naive Bayes ini sama-sama memuaskan kalau di liat dari pengujian data yang sudah di lakukan, Karna dari grafik perbandingan diatas hanya selisih beberapa angka saja yang menurut saya selisih tersebut bisa terjadi pada kesalahan data atau error data.

5.2. Prediksi Kelulusan Mahasiswa

Karna dalam pengujian beberapa data yang sudah dilakukan algoritma C4.5 yang paling unggul dan efektif. Maka untuk memprediksi Kelulusan angkatan 2013 menggunakan algoritma C4.5 dengan data taning angkatan 2010, 2011 dan 2012 dan data testing 2013.

Table. 4.13 Hasil Pengujian Prediksi Kelulusan Mahasiswa

Data		Hasil Pengujian C4.5					
Traning	Angkatan	Testing	Angkatan	Presisi	Recall	Akurasi	Taken Tame
974	2010, 2011, 2012	268	2013	99 %	93 %	95 %	0.08 (s)



Grafik 4.1 Data Kelulusan Mahasiswa

Dari hasil percobaan diatas yang di gambarkan dalam grafik 4.3 menunjukkan bahwa prediksi kelulusan mahasiswa dengan data traning angkatan 2010, 2011, 2013 dan data testing angkatan 2013 tambah meningkan berdasarkan data yang sudah di ujikan.

5.3. Hasil dan Analisis Algoritma C4.5 dan Naïve Bayes

Hasil pengujian dan proses yang dilakukan dalam memprediksi kelulusan mahasiswa berdasarkan parameter mata kuliah umum Pengantar Teknologi Informasi, Programan Terstruktur, Struktur Data, Sistem Informasi, Analisis Desain Berorientasi Objek, Programan Berorientasi Objek, Jaringan Komputer, Praktikum Jaringan Komputer, Programan Jaringan, Praktikum Programan Jaringan Komputer, dan nilai IPK Semester 1 -6, Fakultas Teknik Prodi Teknik Informatika Universitas Muhammadiyah Jember, dihasilkan suatu pola, informasi, dan pengetahuan baru sesuai dengan tujuan peneliti yaitu pola perhitungan data yang berisi data training dan data testing serta mencari probabilitas dari setiap atribut berdasarkan data training dan data testing untuk menghasilkan suatu informasi baru, pada data mahasiswa Fakultas Teknik Prodi Teknik Informatika Universitas Muhammadiyah Jember akan di ketahui lebih dini dalam memprediksi kelulusan mahasiswa berdasarkan parameter yang sudah di tentukan, kemudian data testing yang akan di cari dengan memanfaatkan data traning sebagai data historis atau pembelajara untuk mengetahui dan meprediksi kelulusan mahasiswa kedepannya dan data testing sebagai data uji untuk merealisasikan suatu prediksi, sehingga setelah di uji nilai akhir dari penelitian ini selain untuk menguji data ke dalam algoritma C4.5 dan Naïve Bayes yaitu nilai akhir berupa akurasi, presisi, recall dan taken time yang mempresentasikan prediksi kelulusan mahasiswa. Kemudian untuk menguji tingkat keakurasiannya maka digunakan Tool Weka sebagai alat bantu dalam proses pengujian data. Untuk tingkat akurasi berdasarkan proses klasifikasi menggunakan algoritma C4.5 dan algoritma naïve bayes, dengan melalui beberapa tahapan dan pengujian yang

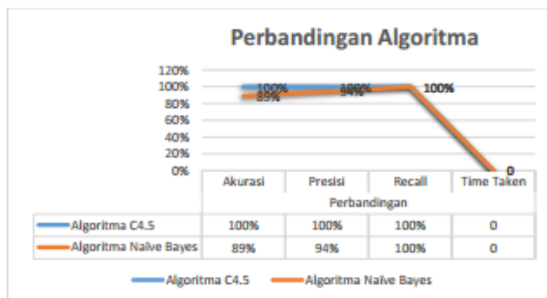
semua tahapan dipastikan tidak ada bagian – bagian penting yang terlewatkan, sehingga dihasilkan tingkat akurasi sebesar 100%, Presisi 100% , Recall 100% dan Teken Time 0 (s) untuk Algoritma C4.5. Kemudian untuk Naïve Bayes dihasilkan tingkat akurasi sebesar 89%, Presisi 100% , Recall 94% dan Teken Time 0 (s), Berdasarkan hasil proses dan pengujian data tingkat akurasi dengan menggunakan Tool Weka untuk percobaan tersebut dapat di bandingkan bahwa nilai akurasi algoritma terbaik adalah Algoritma C4.5 yaitu dengan akurasi sebesar 100%, Presisi 100% , Recall 100% dan Teken Time 0 (s), dari hasil tersebut bisa juga disebabkan oleh beberapa faktor – faktor yaitu : faktor dalam pemilihan parameter, faktor dalam banyangknya dan pemilihan prameter, faktor banyaknya data yang di ujiakan merkaitan data traning dan data testingnya, faktor spesifiksi bahan dan alat yang juga mempegaruhi kecepatan dalam prosesing data, dan bisa juga kurang kompleksitas data yang mengakibatkan model dapat memprediksi cukup akurat. Faktor – faktor diatas tersebut sangatlah mempengaruhi dalam pemrosesan data pada sistem.

BAB V KESIMPULAN DAN SARAN

5.1 Kesimpulan

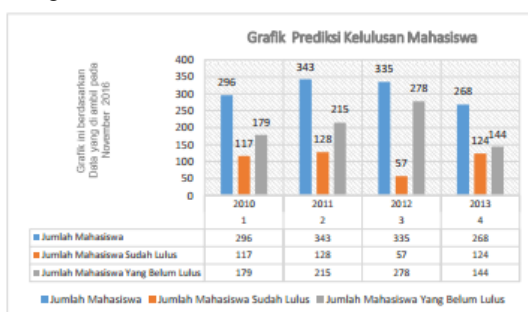
Berdasarkan penelitian dan pengujian data dalam memprediksi kelulusan mahasiswa dapat ditarik kesimpulan sebagai berikut :

1. Dari perbandinga algoritma C4.5 dan Algoritma Naïve Bayes Metode yang efektif digunakan untuk prediksi kelulusan mahasiswa adalah algoritma C4.5 dengan hasil akurasi akurasi sebesar 100%, Presisi 100% , Recall 100% dan Teken Time 0 (s).



Grafik 5.1 Perbandingan Algoritma

2. Karena Algoritma C4.5 yang paling efektif pada perbandingan diatas tersebut maka untuk menemukan hasil prediksi kelulusan mahasiswa angkatan 2013 di lakukan pengujian ulang denga data traning angkatan 2010, 2011, 2012 dan data uji angkatan 2013. Dari pengujian tersebut dihasilkan sebagai berikut.



Grafik 4.3 Prediksi Kelulusan Mahasiswa

Dari jumlah data traning 974 data mahasiswa dan 268 data uji menggunakan metode C4.5 ternyata didapatkan hasil bahwa prediksi kelulusan mahasiswa angkatan 2013 semakin meningkat di lihat dari data historis yang sudah ada.

5.2 Saran

Bagi peneliti - peneliti selanjutnya yang akan melakukan penelitian hampir serupa dan mengembangkan penelitian ini adalah : 1. Mencoba menggunakan aplikasi selain Weka dalam analisa data dan mencoba menggunakan metode lain selain C4.5 dan Naive Bayes. 2. Mencoba lebih banyak record dan attribute dan parameter dalam pemrosesan data. 3. Diperlukan tingkat ketelitian yang tinggi dan pembersihan data yang sempurna supaya tidak terjadinya Error data dalam pemrossan data dalam Tool Weka. 4. Data perlu menyesuaikan dengan kurikulum yang terbaru. Dibuatkan grafik jumlah lulusan setiap tahunnya agar mengetahui ada kenaikan atau tida.

DAFTAR PUSTAKA

- Sofiana, Fiqih. (2012). “ Algoritma C4.5”. Diakses dari [<http://fiqihsofiana.blogspot.com/2012/01/algoritma-c45.html>] (diakses tanggal 26 November 2015).
- Prasetyo, Eko. (2012). Data Mining konsep dan aplikasi menggunakan matlab. Yogyakarta: Andi.
- Kusrini, luthfi taufiq Emha, (2009), Algoritma Data Mining, Penerbit Andi, Yogyakarta.
- Menurut Wu, et al. (2007), “ Fungsi data mining yang dipilih menjadi Top 10 algorithm data mining. Hasilnya, diperoleh bahwa algoritma klasifikasi menjadi mayoritas dalam Top 1 0 algorithm”.
- Galathiya, etal. (2012), “Pengembangan Algoritma pohon keputusan C4.5 dan melakukan perbandingan terhadap algoritma Naive Bayes, Sequential Minimal Optimization (SMO) ”.
- Othman & Yau (2007), “Comparison of Different Classification Techniques using WEKA for Breast Cancer”.
- Claudia Clarentia Ciptohartono, “Algoritma Klasifikasi Naive Bayes Untuk Menilai Kelayakan Kredit,” 2014
- Faid Ari Prastya, (2015) “Penerapan Algoritma C4.5 untuk Prediksi Jurusan siswa SMAN 3 Rembang,”2015.
- Arief Jananto, (2014). “Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa,” Jurnal Teknologi Informasi DINAMIK, vol. 18, no. 1, halaman 09-16, Januari 2013

- Septian Nugroho, (2014)., “*Data Mining Menggunakan Algoritma Naive Bayes untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro*,”2014.
- Lorena Selvia (2014). “*Analisis Dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Memprediksi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik*”. Diakses dari [\[http://repository.akprind.ac.id/sites/files/A263-272%20silvia%20lorena.pdf\]](http://repository.akprind.ac.id/sites/files/A263-272%20silvia%20lorena.pdf) (tanggal 28 Desember 2016)
- Yuda Septian Nugroho, (2014). “*Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro* “. Diakses dari [\[http://eprints.dinus.ac.id/5542/1/13789.pdf\]](http://eprints.dinus.ac.id/5542/1/13789.pdf) (tanggal, 27 Desember 2016)
- Erdi Susanto (2012). “*Data Mining Menggunakan Wika*”. Di akses di[\[http://www.erdisusanto.com/2012/06/data-mining-menggunakan-weka.html\]](http://www.erdisusanto.com/2012/06/data-mining-menggunakan-weka.html) (diakses tanggal 23 Desember 2016).
- Shofiyatin, Lilik. (2011). “*Penerapan Metode Decision Dengan Menggunakan ID3 Untuk Pembuatan Sistem Penilaian Kinerja Guru*”. Skripsi.Malang : Fakultas Sains Dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim.
- Metode Naïve Bayes, (2016) “*Metode Naive Bayes, Sebuah Penjelasan Sederhana*”. di akses [\[https://lukisankeabadian.blogspot.co.id/2015/02/metode-naive-bayes.html\]](https://lukisankeabadian.blogspot.co.id/2015/02/metode-naive-bayes.html) (tanggal 20 Desember 2016).
- Perhitungan Naïve Bayes, (2015). “*Perhitungan Sederhana Naïve Bayes untuk Predeksi Kelulusan* ”. di akses [\[http://Educationprogrammer.blogspot.co.id/2013/01/perhitungan-sederhana-naive-bayes-untuk.html\]](http://Educationprogrammer.blogspot.co.id/2013/01/perhitungan-sederhana-naive-bayes-untuk.html) (tanggal 20 Desember 2016).
- Metode C4.5, (2015) “*Contoh Soal algoritma C4.5 dan Perhitungannya*”. di akses [\[http://mardani-g4ul.blogspot.co.id/2011/12/contoh-soal-algoritma-c45.html\]](http://mardani-g4ul.blogspot.co.id/2011/12/contoh-soal-algoritma-c45.html) (tanggal 25 Desember 2016).
- Website WEKA. <http://www.cs.waikato.ac.nz/ml/weka/>.
- Tool Weka, (2012). “*Data Mining Menggunakan Weka* “. di akses [\[http://www.erdisusanto.com/2012/06/data-mining-menggunakan-weka.html\]](http://www.erdisusanto.com/2012/06/data-mining-menggunakan-weka.html) (tanggal, 29 Desember 2016)
- Suhada, Dewi Rafiq, Saifullah, Rahmat Widia Sembiring, (2012). “*Studi Kajian Web Usage Mining Berbasis Open Source*”. Seminar Nasional : Manajemen Informatika, Politeknik Negeri Medan.