TECHNO
Vol.20, No.2, Oktober 2019, Hal. 131~142
P-ISSN: 1410-8607, E-ISSN: 2579-9096

PREDIKSI KELULUSAN TEPAT WAKTU MENGGUNAKAN METODE C4.5 DAN *K-NN* (Studi Kasus : Mahasiswa Program Studi S1 Ilmu Farmasi, Fakultas Farmasi, Universitas Muhammadiyah Purwokerto)

Eko Purwanto[1], Kusrini [2], Sudarmawan[3]

[123] Department of Informatics, University AMIKOM Yogyakarta, Jl. Ring Road Utara, Condong Catur, Sleman, Yogyakarta, Central Java, 55283, Indonesia

Informasi Makalah

Dikirim, 14 Agustus 2019
Direvisi, 14 Oktober 2019
Diterima, 16 Oktober 2019

Kata Kunci:

Prediksi
Kelulusan tepat waktu
Mahasiswa
Algoritma
C.45
K-NN
Seleksi mundur
Variabel

INTISARI

Profil kelulusan merupakan salah satu element penting bagi standar akreditasi perguruan tinggi. Profil kelulusan mencerminkan kinerja sistem penyelenggaraan pendidikan yang dianut dalam jangka waktu tertentu. Semakin baik profil kelulusan, semakin baik pula nilai akreditasinya. Untuk mewujudkan hal tersebut, prediksi kelulusan dapat dilakukan pada basis data akademik mahasiswa. Hal ini penting dilakukan untuk menelusuri dan mengelompokkan data historikal ke dalam data latih dan data uji, untuk selanjutnya digunakan untuk memprediksi kelulusan tepat waktu. Langkah ini penting untuk membantu menentukan kebijakan manajemen yang lebih baik dari proses pembelajaran. Untuk itulah kajian ini dilakukan untuk menganalisa penggunaan variabel-variabel tertentu untuk memprediksi kelulusan tepat waktu dengan menggunakan metode algoritma C.45 dan K-Nearest Neighbour (K-NN). Penambangan data dilakukan pada basis data akademik mahasiswa program studi Farmasi, Fakultas Farmasi, Universitas Muhammadiyah Purwokerto, dengan menambahkan beberapa variabel tertentu ke dalam proses penambangan data. Data kemudian diklasifikasikan ke dalam data latih dan data uji. Seleksi mundur digunakan untuk menyeleksi variabel yang paling baik dan berpengaruh terhadap set data. Kajian lebih jauh menunjukkan bahwa dengan menggunakan algoritma C.45 dan seleksi mundur, keakuratan kelulusan mencapai diatas 84.75%. Hasil ini berbeda dari keakuratan yang ditunjukkan oleh algoritma K-NN dan seleksi mundur yang mencapai 89.14%. Hasil ini memberikan manfaat yang penting bagi program studi untuk membuat arah kebijakan yang lebih baik guna meningkatkan kualitas pelayanan, terutama pelayanan proses pembelajaran.

Keyword:

Prediction
On-time graduation
Students
Algoryhtm
C.45
K-NN
Backward selection
Variables

ABSTRACT

The graduation profile is one of the key elements for the accreditation standard of higher education. It mirrors the performance of the applied educational system within a period of time. The better it is, the better the accreditation will be. In support of this, a graduation prediction may be conducted to the academic database of the students. It is of pivotal to trace and classify the historical data into the data training and data testing, thus, to predict the on time-graduation. The step is importantly done to help decide the better management of learning processes. This study was therefore done to analyse certain variables applied to predict the on time-graduation using the algorythms of C.45 and K-Nearest Neighbour (K-NN). The data mining was done to the academic database of the students of the Pharmacy study programme, Pharmacy Faculty, Muhammadiyah University of Purwokerto by adding certain variables into the process. The data was then classified into the data training and data testing. Backward selection was done to select the best and most influential variables for the dataset. The study further resulted that by using the algorhythm of C.45 and backward selection, the accuracy of the graduation reached 84.75%. It is different from the acurracy the K-NN and backward selection showed that reached 89.14%. The result confirmed that the KNN showed the better accuracy than the C.45. It considerably benefitted

the study programme to make better decisions on increasing the quality of services, in particular that of leraning processes.

Korespondensi Penulis:

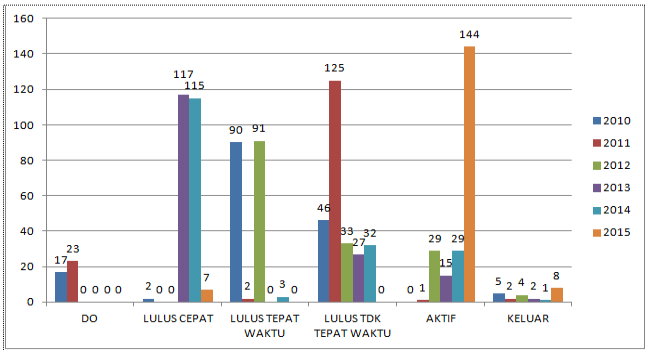
Eko Purwanto
Department of Informatics, University AMIKOM Yogyakarta,
Jl. Ring Road Utara, Condong Catur, Sleman, Yogyakarta,
Central Java, 55283, Indonesia

1. PENDAHULUAN

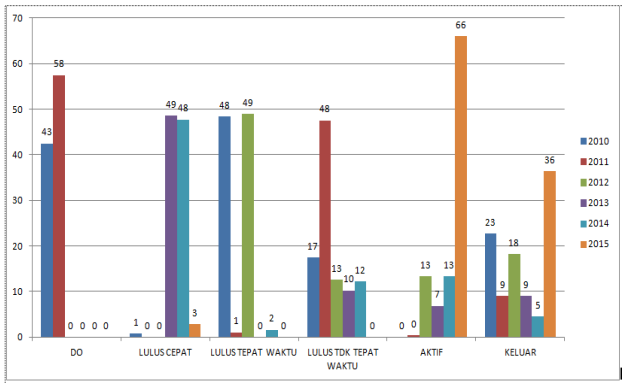
Salah satu parameter penting dalam dunia pendidikan, baik dasar, menengah, maupun tinggi adalah tingkat kelulusan siswa dan atau mahasiswa. Bagi para siswa di pendidikan dasar dan menengah,kelulusan adalah bagian penting untuk melanjutkan pendidikan mereka ke jenjang yang lebih tinggi. Bagi seorang mahasiswa, kelulusan menjadi bagian penting dan awal dari karir dan bahkan kelanjutan untuk menuju ke jenjang yang semakin tinggi. Bagi perguruan tinggi, tingkat kelulusan mahasiswa menjadi salah satu aspek penting penilaian akreditasi. Hal ini seperti dijelaskan dalam Standar 3 pada Buku V Panduan Akreditasi Progam Studi yang dikeluarkan oleh Badan Akreditasi Nasional, tentang Kemahasiswaan dan Lulusan. Dalam standar ini, salah satu butir penilaian diambil dari profil mahasiswa dan lulusan yang mencakup di antaranya (dalam tahun) rerata jumlah mahasiswa, rerata masa studi, dan rerata Indeks Prestasi Kumulatif (IPK). Jika suatu program studi mimiliki rerata masa studi dan IPK yang baik, maka hal ini tentu akan berpengaruh pada *institutional branding* yang juga akan semakin baik.Sesuai acuan akreditasi, rerata masa studi yang baik adalah yang sesuai atau tepat waktu; sedangkan IPK yang baik adalah ≥ 3,00.

Sehubungan dengan hal di atas, diperlukan suatu strategi untuk mempredikasi masa studi atau waktu kelulusan mahasiswa dari awal. Hal ini dimaksudkan untuk membantu mahasiswa meningkatkan kualitas belajarnya dan membantu program studi memperbaiki kualitas penyelenggaraan pendidikan. Dengan demikian, upaya untuk mencapai rerata masa studi tepat waktu dan menghindari *drop-out* serta meningkatkan mutu layanan pendidikan dapat terwujud.

Ketepatan masa studi juga menjadi hal penting bagi mahasiswa prodi S1 Ilmu Farmasi, Fakultas Farmasi, Universitas Muhammadiyah Purwokerto. Berdasarkan kurikulum prodi dan SNPT, persyaratan kelulusan ditentukan oleh total sks minimum yang harus ditempuh, yaitu 144 sks dan IPK minimum yaitu, 2,76. Ketepatan waktu lulusan adalah 4 tahun atau 8 semester. Dari data akademik yang diambil dari 5 tahun terakhir mahasiswa dengan status lulus yaitu angkatan 2010 hingga angkatan 2015, tren kelulusan menunjukkan pola yang fluktuatif, seperti terlihat pada gambar1 dan 2 berikut.



Gambar 1. Profil jumlah lulusan Program Studi S1 Ilmu Farmasi, Fakultas Farmasi, Universitas Muhammadiyah Purwokerto 5 tahun terakhir mahasiswa dengan status lulus (tahun masuk angkatan 2010-2015.)



Gambar 2. Persentase profil lulusan Program Studi S1 Ilmu Farmasi, Fakultas Farmasi, Universitas Muhammadiyah Purwokerto 5 tahun terakhir mahasiswa dengan status lulus (tahun masuk angkatan 2010-2015.)

Pada kedua gambar di atas, diperoleh data bahwa persentase kelulusan tepat waktu (4 tahun) mahasiswa prodi S1 Ilmu Farmasi masih sangat rendah untuk 3 tahun terakhir, tetapi dua tahun terakhir sangat tinggi untuk lulus dengan cepat (kurang dari 4 tahun). Sementara itu, untuk data kelulusan tidak tepat waktu cenderung rendah pada 3 tahun terakhir. Hal ini menunjukkan bahwa data ini tidak beraturan karena didasarkan pada rasio jumlah mahasiswa yang masuk dan yang keluar (lulus), tidak didasarkan pada kualitas *input*, proses, dan kualitas luaran (*output*).

Untuk itulah penelitian ini akan dilakukan untuk mengetahui variabel-variabel penentu yang akan digunakan untuk memprediksi masa studi tepat waktu dan menentukan kebijakan yang lebih tepat untuk meningkatkan waktu kelulusan tepat dan cepat waktu dan mengurangi angka *drop-out* (DO) dan tidak tepat waktu. Penelitian ini akan dilakukan dengan data mining melalui algoritma C4.5 dan *K-Nearest Neighbour* (*K-NN*).

2. LANDASAN TEORI

2.1. Data Mining

[1] mendefinisikan *data mining* sebagai suatu kajian tentang pengumpulan, pembersihan, pemrosesan, analisis, dan pemerolehan beragam pengetahuan yang bermanfaat dari suatu data. Kajian ini berhubungan dengan domain permasalahan, aplikasi, formulasi dan representasi data yang ditemukan dalam aplikasi nyata. Dengan kata lain, "*data mining*" adalah "payung umum" ("*broad umbrella*") yang digunakan untuk mendeskripsikan keempat aspek berbeda ini dalam pemrosesan data [3] data mining sebagai kebutuhan yang sangat penting karena berhubungan dengan pergerakan dan pengumpulan data besar yang begitu cepat setiap harinya. Data mining juga dapat dikatakan sebagai evolusi teknologi informasi dari sistem pemrosesan file primitif menuju sistem basis data yang kuat dan canggih. Lebih jauh, data mining merujuk pada istilah "*knowledge mining from data*, *knowledge extraction*, *data/pattern analysis*, *data archaeology*, dan *data dredging"*. Umumnya orang lebih mudah mengenalnya sebagai "*knowledge discovery from data"* atau **KDD** (penemuan pengetahuan dari data) [3].
Sementara itu, [4] menyebut *data mining* sebagai solusi permasalahan dengan menganalisa data yang sudah ada dalam basis data. Dengan demikian, ini juga disebut sebagai proses penemuan pola dalam basis data yang diharapkan dapat memberikan keuntungan secara ekonomis [4].

Berdasarkan pengertian di atas, *data mining* dapat disimpulkan sebagai:

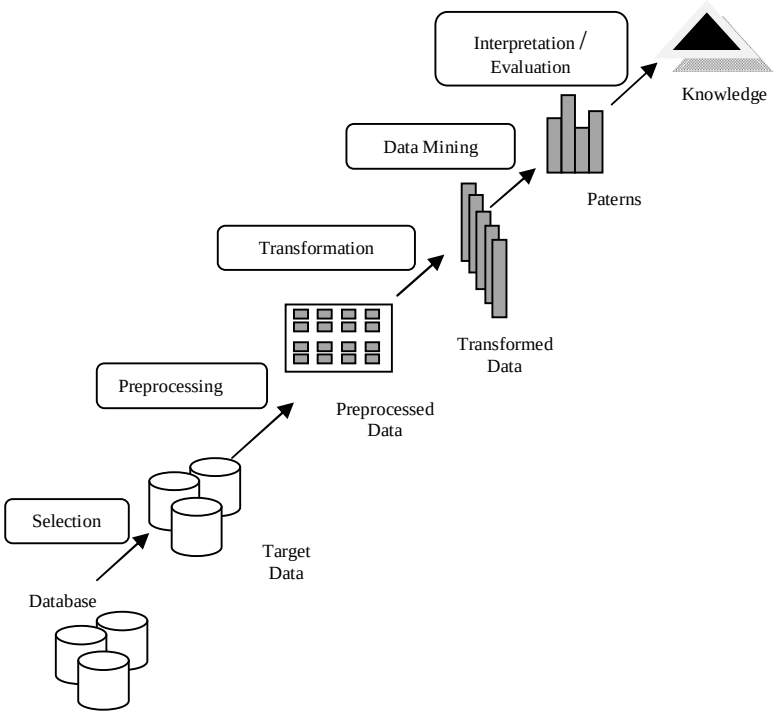
a. Suatu kajian tentang proses pengolahan data
b. Suatu evolusi teknologi informasi
c. Suatu proses ekstraksi atau penambangan pengetahuan atau informasi dari basis data besar yang berguna untuk pengambilan keputusan di masa depan, memperkecil pengeluaran, dan memberikan keuntungan

Proses penemuan pengetahuan atau informasi dalam *data mining* dapat dilakukan dengan beberapa tahap, yaitu:

1 *Data cleaning* (untuk menghilangkan data yang mengganggu dan tidak konsisten)
2 *Data integration* (mengkombinasikan sumber-sumber data yang banyak/*multiple*)
3 *Data selection* (pengambilan data dari basis data yang relevan dengan tugas analisis)
4 *Data transformation* (data diubah dan dikonsolidasikan ke dalam bentuk-bentuk yang sesuai untuk ditambang melalui operasi agregasi dan ringkasan)
5 *Data mining* (proses penting mengekstraksi pola-pola data)
6 *Pattern evaluation* (mengidentifikasi pola-pola yang benar-benar menarik yang merepresentasikan pengetahuan berdasarkan "*interestingness-measures*")

7 *Knowledge representation* (visualisasi dan teknik representasi pengetahuan yang telah ditambang kepada pengguna) [3].

Selanjutnya dapat dilihat dalam Gambar 3 di bawah ini.



Gambar 3.Tahap-tahap ekstraksi informasi dari basis data.

Data mining dalam penelitian ini akan menggunakan algoritma C4.5 dan K-NN untuk memprediksi kelulusan tepat waktu mahasiswa prodi Farmasi Universitas Muhammadiyah Purwokerto Tahun Lulusan 2010-2015.

2.2. C4.5

Algoritma C4.5 adalah salah satu teknik dari data mining yang berupa klasifikasi data.Tujuannya adalah untuk membuat model-model pendeskripsian kelas-kelas data yang penting.Algoritma ini adalah salah satu cabang pohon keputusan yang mirip seperti struktur pohon, berisi node internal (bukan daun), yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas. Jika diuji dengan sejumlah data, misalnya X di mana kelas data X ini belum diketahui, maka pohon keputusan ini akan bekerja menelusuri data mulai dari akar sampai node dan setiap nilai dari atribut sesuai data X diuji apakah sesuai dengan aturan pohon keputusan, kemudian pohon keputusan akan memprediksi kelas dari tupel X [10].

Algoritma C4.5 dalam penelitian akan digunakan untuk mengklasifikasikan data akademik mahasiswa prodi Farmasi yang selanjutnya akan digunakan untuk memprediksi kelulusan tepat waktu. Hasil prediksi akan dijadikan dasar pengambilan kebijakan oleh pihak universitas untuk semakin mendorong para mahasiswa menyelesaikan studi mereka tepat waktu.

Tahap-tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5, adalah sebagai berikut [2]:

1. Menyiapkan data training/latih. Data training diambil dari data histori yang sudah terjadi sebelumnya dan dikelompokan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung gain dari atribut, hitung dahulu nilai entropy menggunakan persamaan 1 sebagai berikut:

$$Entrophy\ (s) = \sum_{i=1}^{n} - pi * \log_2 pi \qquad (1)$$

Keterangan:
S : himpunan kasus
A : atribut

n : jumlah partisi S
pi : proporsi dari Si terhadap S

3. Menentukan nilai gain dengan metode informasi gain:

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^{n} \frac{Si}{S} * Entropy(Si) \quad (2)$$

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

4. Ulangi langkah ke-2 hingga semua kasus terpartisi.
5. Proses partisi pohon keputusan akan berhenti saat:
   a. Semua kasus dalam node N mendapat kelas yang sama.
   b. Tidak ada atribut di dalam kasus yang dipartisi lagi.
   c. Tidak ada kasus di dalam cabang yang kosong.

2.3. K-Nearest Neighbour (K-NN)

Berbeda dari algoritam C4.5, algoritma *K-Nearest Neighbour* (K-NN) merupakan salah satu teknik klasifikasi data yang kuat dengan mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama melalui pencocokan bobot [8].

K-NN adalah metode algoritma *supervised learning* yang mendasarkan hasil klasifikasi pada kelas yang paling banyak muncul [7].

Alur dalam algoritma K-NN adalah:
1. Menentukan parameter K (jumlah tetangga paling dekat), Parameter K pada testing ditentukan berdasarkan nilai K optimum padasaat training.
2. Menghitung kuadrat jarak euclid (*euclidean distance*) masing-masing objek terhadap data sampel yang diberikan.
3. Mengurutkan objek-objek tersebut kedalam kelompok yang mempunyai jarak *Euclidian* terkeci
4. Mengumpulkan kategori Y (klasifikasi *nearest neighbour*)
5. Dengan menggunakan kategori mayoritas, maka dapat hasil klasifikasi [6].

Dalam pendefinisian jarak antara x dan y, digunakan rumus jarak *Euclidian* pada persamaan (1) [7] berikut.

$$D(x,y) = \sqrt{\sum_{k}^{n} = 1 (X_{training} - Y_{testing})^2} \quad (3)$$

Keterangan :
Xtraining : data training ke-*i*,
Y testing : data testing,
i : record (baris) ke-*i* dari table,
n : jumlah data training.

Seperti algoritma C4.5, algoritma K-NN digunakan untuk memprediksi klasifikasi data kelulusan tepat waktu mahasiswa prodi Farmasi, UMP. Kedua algoritma akan diuji efektifitasnya atau keakurasiannya dalam merepresentasikan data kelulusan tepat waktu dari 5 tahun kelulusan terakhir selanjutnya akan digunakan untuk menentukan strategi yang tepat dalam memperbaiki dan meningkatkan waktu kelulusan mahasiswa.

2.4. Fitur Backward Selection/Elimination

Metode backward, adalah suatu metode pemilihan variabel yang berpengaruh dalam analisa data dengan caramemasukkan semua variabel / prediktor, kemudian mengeliminasi satu persatu hingga tersiswa prediktor yang signifikan saja, yaitudengan melakukan pengujian terhadap parameter -parameternya dengan menggunakan partial F test. Nilai partial F-test (FL) terkecil dibandingkan dengan F0 table:

- Jika FL < F0, maka X yang bersangkutan dikeluarkan dari model dan dilanjutkan dengan pembuatan model baru tanpa variable tersebut
- Jika FL>F0, maka proses dihentikan dan persamaan terakhir tersebut yang digunakan/dipilih

3. METODE PENELITIAN

3.1. Metode Pengumpulan Data

Tahap dalam penelitian ini adalah dengan pengumpulan data dilakukan melalui observasi langsung, yaitu mengamati basis data akademik mahasiswa prodi Farmasi UMP dari 5 tahun kelulusan terakhir untuk tujuan identifikasi masalah dan seleksi dataset yang akan diklasifikasikan ke dalam data training dan data testing. Selain itu juga melakukan studi pustaka untuk mendapatkan data sekunder berupa informasi yang penting untuk menjawab rumusan masalah.

3.2. Metode Analisis Data

Analisis data dalam penelitian ini akan dilakukan dengan menggunakan teknik klasifikasidata mining melalui algoritma C4.5 dan K-NN dengan menambahkan fitur backward selection. Atribut/variabel yang digunakan adalah: Jenis kelamin, Status Sekolah, Jurusan, NEM, nilai TPA, Indek prestasi semester (IPs), Indek prestasi komulatif (IPK), Jml SKS, dan Riwayat pembayaran SPP untuk mendapatkan keakuratan data prediksi kelulusan tepat waktu mahasiswa prodi Farmasi UMP TA 2010 sampai dengan 2019 (sekarang).

Dalam melakukan perbandingan analisa menggunakan algoritma C4.5 dan K-NN untuk menghasilkan keakuratan data, dilakukan beberapa cara, yaitu :

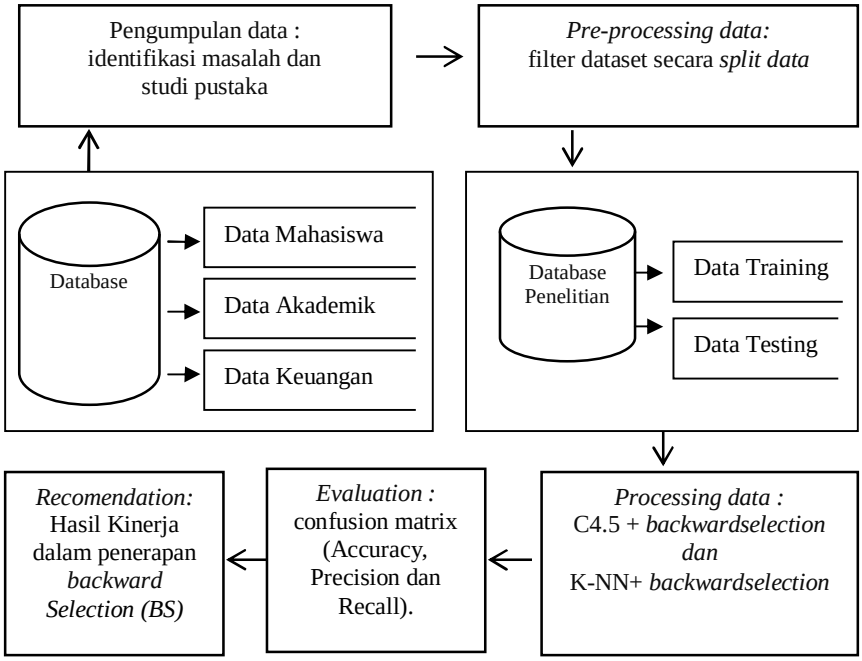
1. Melakukan uji data menggunakan beberapa pasang variabel yang berbeda pada masing-masing algoritma
2. Melakukan uji data dengan data training satu angkatan atau beberapa angkatan, atau seluruh angkatan
3. Melakukan uji data dengan hasil pemilihan variabel yang paling berpengaruh menggunakan fitur backward selection
4. Melakukan uji data dengan beberapa jumlah tetangga/ N pada algoritma K-NN

3.3. Alur Penelitian

Alur penelitian ini dilakukan dengan tahap-tahap sebagai berikut:

1. Pengumpulan data melalui observasi langsung yang bertujuan untuk identifikasi masalah dan studi pustaka
2. *Pre-processing data* yang dilakukan dengan menyeleksi dataset secara *split data*untuk selanjutnya dilakukan mining yaitu dengan mengklasifikasikan dataset ke dalam data training dan data testing menggunakan algoritma klasifikasi dengan menambahkan fitur *backward selection*dilengkapi atribut yang telah ditentukan untuk memperoleh klasifikasi data dan parameter evaluasi
3. *Processing data* dilakukan untuk mining klasifikasi data yang telah diperoleh dengan menambahkan fitur *backwardselection* dilengkapi atribut yang telah ditentukanmenggunakan algoritma C4.5 dan K-NN untuk mendapatkan optimatisasi kedua algoritma dalam memprediksi potensi kelulusan tepat waktu mahasiswa prodi Farmasi UMP.
4. *Evaluation* dilakukan untuk memperoleh optimalisasi perbandingan keakuratan kedua algoritma, keunggulan dan kelemahannya guna perumusan rekomendasi bagi pengambilan keputusan yang lebih strategik di masa mendatang, yaitu bagi peningkatan potensi kelulusan tepat waktu mahasiswa prodi Farmasi UMP. Pada tahap evaluasi ini dilakukan berbagai macam skenario untuk mendapatkan hasil yang mendekati keakuratan yang maksimal.
5. *Recommendation* dilakukan untuk memberikan pertimbangan terhadap prosessing data setelah dilakukan evaluasi dan mengetahui keakuratannya.

Gambar 4 berikut ini adalah alur penelitian yang dimaksud.



Gambar 4. Alur Penelitian yang dilakukan

4. HASIL DAN PEMBAHASAN

4.1. Dataset Mahasiswa

Setelah proses pengumpulan data dilakukan, proses klasifikasi diawali dengan penentuan dataset yang disimpan dalam format excel (*.xls) seperti pada gambar 5 berikut:

NIM	KEL	STS SEKOLAH	JURUSAN	NEM	TPA	SKS SMT 1	SKS SMT 2	SKS SMT 3	SPP SMT 3	SPP SMT 4	SPP SMT 5	SPP SMT 6	SPP SMT 7	SPP SMT 8	IPK	SKS	PREDIKAT KELULUSAN
1308010088	L	SWASTA	IPS	6	41	20	22	18	20141	20142	20151	20152	20161	20162	2,70	146	TEPAT WAKTU
1408010105	P	NEGERI	IPA	6	61	20	22	20	20151	20152	20161	20162	20171	20172	3,34	146	TEPAT WAKTU
1308010137	L	NEGERI	IPA	7	44	20	22	18	20141	20142	20151	20152	20161	20162	2,75	146	TEPAT WAKTU
1408010189	L	NEGERI	IPA	7	55	20	22	20	20151	20152	20161	20162	20171	20172	3,41	146	TEPAT WAKTU
1408010115	P	NEGERI	IPA	7	50	20	22	18	20151	20152	20161	20162	20171	20172	2,95	146	TEPAT WAKTU
1408010124	P	NEGERI	IPA	7	56	20	22	20	20151	20152	20161	20162	20171	20172	3,42	146	TEPAT WAKTU
1308010041	P	NEGERI	IPA	8	46	20	22	18	20141	20142	20151	20152	20161	20162	3,01	146	TEPAT WAKTU
1308010066	P	NEGERI	FARMASI	8	42	20	22	21	20141	20142	20151	20152	20161	20162	3,23	149	LAMBAT
1208010128	P	SWASTA	IPA	8	54	20	22	21	20131	20132	20141	20142	20151	20152	3,00	147	TEPAT WAKTU
1308010054	L	SWASTA	FARMASI	8	44	20	22	21	20141	20142	20151	20152	20161	20162	2,95	146	TEPAT WAKTU
1108010061	P	SWASTA	FARMASI	25	98	20	22	21	20121	20122	20131	20132	20141	20142	3.38	147	TEPAT WAKTU
1008010149	P	SWASTA	FARMASI	26	106	20	20	21	20111	20112	20121	20122	20131	20132	3.32	147	TEPAT WAKTU
1108010153	P	SWASTA	FARMASI	32	95	20	18	18	20121	20122	20131	20132	20141	20142	2,53	147	LAMBAT
1108010100	P	SWASTA	FARMASI	33	92	20	18	15	20121	20122	20131	20132	20141	20142	2,37	147	LAMBAT
1108010090	P	SWASTA	FARMASI	33	98	20	22	21	20121	20122	20131	20132	20141	20142	3.35	148	TEPAT WAKTU
1108010121	P	SWASTA	FARMASI	34	97	20	22	21	20121	20122	20131	20132	20141	20142	3.31	147	TEPAT WAKTU
1208010054	P	SWASTA	FARMASI	34	45	20	22	21	20131	20132	20141	20142	20151	20152	2,89	147	TEPAT WAKTU
1008010043	L	NEGERI	IPS	40	95	20	15	21	20111	20112	20121	20122	20131	20132	2.40	147	LAMBAT
1108010060	P	SWASTA	FARMASI	40	94	20	15	18	20121	20122	20131	20132	20141	20142	2,32	147	LAMBAT
1208010152	P	NEGERI	IPA	41	34	20	22	21	20131	20132	20141	20142	20151	20152	3,11	147	TEPAT WAKTU
1108010083	P	NEGERI	IPA	41	95	20	22	18	20121	20122	20131	20132	20141	20142	2.50	148	TEPAT WAKTU
1208010123	L	NEGERI	IPA	42	50	20	19	18	20131	20132	20141	20142	20151	20152	2,44	147	LAMBAT
1208010102	P	NEGERI	IPA	42	54	20	20	18	20131	20132	20141	20142	20151	20152	2,59	147	LAMBAT
1108010068	P	SWASTA	FARMASI	44	103	20	22	21	20121	20122	20131	20132	20141	20142	3.70	147	TEPAT WAKTU

Gambar 5. Potongan Data Status Mahasiswa

Variabel yang digunakan untuk proses data dan tipe data attribut dan label pada data indikator dataset mahasiswa dalam penelitian ini dapat dijabarkan seperti pada table 1 berikut:

Tabel 1. Pembagian Variabel, Atribut, Kelas dan Tipe Data

Variabel	Nama Atribut	Kelas Data	Tipe Data
V1	NIM	Nomor Induk Mahasiswa	Character
V2	Jenis Kelamin	Laki-laki ; Perempuan	Character
V3	Status Sekolah Asal	Negeri ; Swasta	Character
V4	Jurusan	Juruan SLTA	Character
V5	NEM	Nilai Ebtanas Murni	Numeric
V6	Nilai TPA	Nilai Tes Potensi Akademik	Numeric
V7	Indek Prestasi Smt 1	0 s/d 4,00	Numeric

V8	Indek Prestasi Smt 2	0 s/d 4,00	Numeric
V9	Indek Prestasi Smt 3	0 s/d 4,00	Numeric
V10	Indek Prestasi Smt 4	0 s/d 4,00	Numeric
V11	Indek Prestasi Smt 5	0 s/d 4,00	Numeric
V12	Indek Prestasi Smt 6	0 s/d 4,00	Numeric
V13	Indek Prestasi Smt 7	0 s/d 4,00	Numeric
V14	Indek Prestasi Smt 8	0 s/d 4,00	Numeric
V15	Registrasi/Aktif Smt 1	Aktif ; Tidak Aktif ; Lulus	Character
V16	Registrasi/Aktif Smt 2	Aktif ; Tidak Aktif ; Lulus	Character
V17	Registrasi/Aktif Smt 3	Aktif ; Tidak Aktif ; Lulus	Character
V18	Registrasi/Aktif Smt 4	Aktif ; Tidak Aktif ; Lulus	Character
V19	Registrasi/Aktif Smt 5	Aktif ; Tidak Aktif ; Lulus	Character
V20	Registrasi/Aktif Smt 6	Aktif ; Tidak Aktif ; Lulus	Character
V21	Registrasi/Aktif Smt 7	Aktif ; Tidak Aktif ; Lulus	Character
V22	Registrasi/Aktif Smt 8	Aktif ; Tidak Aktif ; Lulus	Character
V23	SKS	0 s/d 150	Numeric
V24	IPK	0 s/d 4,00	Numeric

4.2. Implementasi K-NN

Dari jumlah data lulusan mulai angkatan 2010-2015 sebanyak 719, ditentukan untuk data latih/training angkatan 2010-2013 sebanyak 543 dan untuk data uji/test angkatan 2014-2015 sebanyak 176

Contoh proses perhitungan menggunakan algoritma K-NN menggunakan excel dengan menentukan jumlah tetangga(N) antara 1 s/d 21 ditunjukan pada gambar 6 berikut :

Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK
Nim	Ek Distance	N1	N3	N5	N7	N9	N11	N13	N15	N17	N19	N21
1408010002	10.66442216											
	2.320818821								TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU
	1.523909446			LAMBAT	LAMBAT	LAMBAT	LAMBAT	LAMBAT	LAMBAT	LAMBAT	LAMBAT	LAMBAT
	2.270066078							TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU
	8.085146876											
	8.54141089											
	3.120689026											TEPAT WAKTU
	2.259137003							TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU
	10.16754149											
	2.6813243										LAMBAT	LAMBAT
	6.114392856											
	3.656760862											
	7.250420677											
	1.016661202		TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU
	0.84172442		TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU
	2.219143979						TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU	TEPAT WAKTU
	6.575979015											
	2.83961265										LAMBAT	LAMBAT
	7.075436382											
	6.229975923											
	1.529215485				CEPAT	CEPAT	CEPAT	CEPAT	CEPAT	CEPAT	CEPAT	CEPAT

Gambar 6. Potongan proses perhitungan menggunakan algoritma K-NN

4.3. Implementasi C4.5

Pada implementasi C4.5, dilakukan proses klasifikasi terhadap atribut dan jumlah nilai pada masing-masing atributnya seperti pada Tabel 2, selanjutnya hitunglah enteropi dan Gainnya menggunakan rumus Enteropi dan Gain seperti pada rumus persamaan ke (1) dan (2).

Tabel 2. Menghitung Jumlah kasus setiap atribut dan menghitung Enteropi dan Gain masing-masing atribut pada node pertama.

Atribut	Nilai	Jml Kasus	Cepat	Tepat Waktu	Lambat	Entropi	Gain
Jenis Kelamin							0.030536835
	Laki-laki	124	9	54	61	1.30042028	
	Perempuan	419	15	283	121	1.07184678	
	Total	543	24	337	182	1.15458088	
Status Sekolah Asal							0.006038652
	Negeri	344	17	202	125	1.19612341	
	Swasta	199	7	135	57	1.06629135	
	Total	543	24	337	182	1.15458088	
Jurusan							0.079508994
	IPA	346	19	215	112	1.18319517	
	IPS	27	0	12	15	0	
	Farmasi	161	5	105	51	1.08309631	

	Perkantoran	1	0	0	1	0	
	Akuntansi	1	0	1	0	0	
	Bahasa	1	0	0	1	0	
	Bangunan	1	0	0	1	0	
	Jaringan	1	0	1	0	0	
	Science	4	0	3	1	0	
	Total	543	24	337	182	1.15458088	
NEM							0.030375633
	Sangat Baik	107	7	77	23	1.07571023	
	Baik	324	14	179	131	1.19699690	
	Cukup	111	3	81	27	0.96860776	
	Kurang	1			1	0	
	Total	543	24	337	182	1.15458088	
Nilai TPA							0.075313425
	Sangat Baik	218	10	112	96	1.21864564	
	Baik	26	4	21	1	0.84510642	
	Cukup	165	7	130	28	0.89865411	
	Kurang	134	3	74	57	1.12034914	
	Total	543	24	337	182	1.15458088	
IP Smt 1							0.610137174
	Istimewa (≥ 3.51)	55	11	41	3	1.00920940	
	Sangat Baik (≥ 3.01)	173	12	141	20	0.86737474	
	Baik (≥ 2.76)	103	1	77	25	0.87447171	
	Kurang (< 2.76)	212	0	78	134	0	
	Total	543	24	337	182	1.15458088	
IP Smt 2							0.762462044
	Istimewa (≥ 3.51)	35	8	27		0	
	Sangat Baik (≥ 3.01)	129	11	107	11	0.82948723	
	Baik (≥ 2.76)	112	5	86	21	0.94568461	
	Kurang (< 2.76)	267		117	150	0	
	Total	543	24	337	182	1.15458088	
IP Smt 3							0.211018018
	Istimewa (≥ 3.51)	39	8	30	1	0.89548638	
	Sangat Baik (≥ 3.01)	151	12	125	14	0.83412954	
	Baik (≥ 2.76)	99	2	76	21	0.88106211	
	Kurang (< 2.76)	254	2	106	146	1.04036203	
	Total	543	24	337	182	1.15458088	
IP Smt 4							0.183161574
	Istimewa (≥ 3.51)	73	10	61	2	0.75154704	
	Sangat Baik (≥ 3.01)	164	11	118	35	1.07870729	
	Baik (≥ 2.76)	102	2	78	22	0.88449526	
	Kurang (< 2.76)	204	1	80	123	1.00730999	
	Total	543	24	337	182	1.15458088	
IP Smt 5							0.599232815
	Istimewa (≥ 3.51)	58	9	46	3	0.90335923	
	Sangat Baik (≥ 3.01)	181	12	148	21	0.85754028	
	Baik (≥ 2.76)	99	3	74	22	0.94893306	
	Kurang (< 2.76)	205		69	136	0	
	Total	543	24	337	182	1.15458088	
IP Smt 6							0.432628548
	Istimewa (≥ 3.51)	86	11	73	2	0.70637010	
	Sangat Baik (≥ 3.01)	224	12	175	37	0.93355461	
	Baik (≥ 2.76)	92		54	38	0	
	Kurang (< 2.76)	141	1	35	105	0.86635501	
	Total	543	24	337	182	1.15458088	
IP Smt 7							0.442025551
	Istimewa (≥ 3.51)	103	6	84	13	0.85572232	
	Sangat Baik (≥ 3.01)	265	18	183	64	1.12746470	
	Baik (≥ 2.76)	128		61	67	0	
	Kurang (< 2.76)	47		9	38	0	
	Total	543	24	337	182	1.15458088	
IP Smt 8							0.722476738
	Istimewa (≥ 3.51)	114	16	96	2	0.70871107	
	Sangat Baik (≥ 3.01)	198	7	177	14	0.58531301	
	Baik (≥ 2.76)	48	1	39	8	0.79057320	
	Kurang (< 2.76)	183		25	158	0	
	Total	543	24	337	182	1.15458088	
Registrasi/ Aktif Smt 1							0
	Aktif	543	24	337	182	1.15458088	
	Tidak Aktif		0	0	0	0	
	Total	543	24	337	182	1.15458088	

Registrasi/ Aktif Smt 2							0
	Aktif	543	24	337	182	1.15458088	
	Tidak Aktif					0	
	Total	543	24	337	182	1.15458088	
Registrasi/ Aktif Smt 3							0
	Aktif	543	24	337	182	1.15458088	
	Tidak Aktif					0.00000000	
	Total	543	24	337	182	1.15458088	
Registrasi/ Aktif Smt 4							0
	Aktif	543	24	337	182	1.15458088	
	Tidak Aktif					0	
	Total	543	24	337	182	1.15458088	
Registrasi/ Aktif Smt 5							0
	Aktif	543	24	337	182	1.15458088	
	Tidak Aktif					0	
	Total	543	24	337	182	1.15458088	
Registrasi/ Aktif Smt 6							0
	Aktif	543	24	337	182	1.15458088	
	Tidak Aktif	0	0	0	0	0	
	Total	543	24	337	182	1.15458088	
Registrasi/ Aktif Smt 7							0.008340698
	Aktif	542	23	337	182	1.14835502	
	Tidak Aktif					0	
	Lulus	1	1			0	
	Total	543	24	337	182	1.15458088	
Registrasi/ Aktif Smt 8							0.034048192
	Aktif	539	20	337	182	1.12884833	
	Tidak Aktif					0	
	Lulus	4	4			0	
	Total	543	24	337	182	1.15458088	
SKS							0.007316331
	>=147	470	18	300	152	1.12036963	
	>=144	73	6	37	30	1.32042364	
	Total	543	24	337	182	1.15458088	
IPK							0.695028759
	Istimewa (≥ 3.51)	57	12	43	2	0.94957537	
	Sangat Baik (≥ 3.01)	235	12	193	30	0.83153620	
	Baik (≥ 2.76)	122		72	50	0	
	Kurang (< 2.76)	129		29	100	0	
	Total	543	24	337	182	1.15458088	

Dalam menentukan entropy dan gain prosesnya akan berhenti sampai tidak ada atribut lainnya yang dapat digunakan untuk mempartisi sampai lebih lanjut. Dan selanjutnya menentukan pohon keputusannya berdasarkan nilai gain tertinggi.

4.4. Analisa Hasil

Dari perhitungan K-NN dan C4.5, diperoleh keakuratan data yang ditunjukan pada tabel 3 sebagai berikut :

Tabel 3. Keakuratan data menggunakan Algoritma K-NN dan C4.5

Algoritma	Jumlah Data Training	Jumlah Data Testing	Keakuratan
K-NN	543	76	82,26%
C4.5	543	76	80,68%

Untuk meningkatkan hasil kekauratan data, dilakukan penambahan seleksi atribut yang paling menentukan dan membuang atribut yang tidak berperangaruh menggunakan seleksi fitur Backward eliminiation dengan regresi linear. Penentuan variabel yang berpengaruh menggunakan aplikasi statistik regresi linear melalui tahap atau model 9 proses. Proses pertama ditunjukan pada tabel 4 berikut :

Tabel 4. Korelasi antar variabel dalam model 1

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	22.282	4.644		4.798	.000	13.158	31.405
	KE1	.038	.057	.022	.671	.503	-.074	.150
	STS SEKOLAH	.127	.056	.083	2.285	.023	.018	.237
	JURUSAN	.011	.036	.011	.296	.767	-.060	.082
	NEM	.012	.028	.014	.412	.681	-.044	.067
	TPA	.005	.001	.160	4.300	.000	.003	.008
	IPS SMT 1	.014	.113	.009	.120	.905	-.208	.235
	IPS SMT 2	-.186	.109	-.129	-1.715	.087	-.400	.027
	IPS SMT 3	.039	.121	.025	.319	.750	-.199	.277
	IPS SMT 4	.017	.107	.012	.162	.871	-.193	.228
	IPS SMT 5	-.408	.108	-.276	-3.770	.000	-.620	-.195
	IPS SMT 6	.021	.087	.014	.245	.807	-.150	.193
	IPS SMT 7	-.302	.083	-.159	-3.640	.000	-.464	-.139
	IPS SMT 8	-.394	.031	-.513	-12.551	.000	-.456	-.332
	SPP SMT 7	-1.385	.622	-.080	-2.226	.026	-2.607	-.163
	SPP SMT 8	-.247	.318	-.028	-.776	.438	-.872	.378
	IPK	.540	.337	.265	1.601	.110	-.123	1.202
	SKS	-.106	.031	-.111	-3.392	.001	-.168	-.045

Pada model pertama (1), ditunjukan dengan penghilangan/penghapusan beberapa variabel/atribut yang tidak berpengaruh karena datanya bersifat homogen. Selanjutnya potongan proses model ke 9 (terakhir) ditunjukan pada tabel 5 berikut :

Tabel 5. Korelasi antar variabel dalam model 9

9	(Constant)	22.275	4.455		5.000	.000	13.523	31.027
	STS SEKOLAH	.134	.048	.087	2.796	.005	.040	.227
	TPA	.006	.001	.169	5.375	.000	.004	.008
	IPS SMT 2	-.167	.101	-.116	-1.659	.098	-.366	.031
	IPS SMT 5	-.401	.101	-.271	-3.969	.000	-.600	-.203
	IPS SMT 7	-.296	.081	-.157	-3.671	.000	-.454	-.138
	IPS SMT 8	-.394	.031	-.512	-12.800	.000	-.454	-.333
	SPP SMT 7	-1.673	.535	-.097	-3.127	.002	-2.723	-.622
	IPK	.600	.220	.294	2.722	.007	.167	1.032
	SKS	-.105	.030	-.110	-3.479	.001	-.165	-.046

a. Dependent Variable: lamastudi

Dalam model ke 9, diperoleh variabel-variabel yang lebih berpengaruh terhadap proses pengolahan data prediksi kelulusan. Selanjutnya perhitungan ulang algoritma K-NN dan C4.5 meggunakan variabel/atribut tersebut, dan menghasilkan kekauratan data yang ditunjukan pada tabel 6 berikut.

Tabel 6. Keakuratan data menggunakan Algoritma K-NN+Backward Elimination dan C4.5+Backward Elimination

Algoritma	Jumlah Data Training	Jumlah Data Testing	Kekauratan
K-NN + Backward Elimination	543	76	89,14%
C4.5 + Backward Elimination	543	76	84,75%

5. KESIMPULAN

Pengujian data prediksi kelulusan tepat waktu menggunakan perbandingan algoritma K-NN dan C4.5 menghasilkan data keakuratan yang tinggi diatas 80%. Keakuratan data dapat ditingkatkan dengan menambahkan fitur Backward Selection/Elimination sehingga meningkatan keakuratan data pada penelitian ini. Dengan keakuratan yang tinggi, diharapkan metode ini dapat membantu menentukan arah kebijakan pada institusi dalam rangka meningkatkan mutu kelulusan dengan cepat atau tepat waktu.

6. DAFTAR PUSTAKA.

[1] Aggarwal, Charu J. 2015. *Data Mining: The Textbook.* Springer. New York.
[2] Kusrini dan Luthfi, E. T., 2009. Algoritma Data Mining. Yogyakarta : Penerbit Andi.

[3] Han, Jiawei, Kamber, Micheline, dan Pei, Jian. 2012. *Data Mining: Concepts and Techniques*.Edisi ketiga.Morgan Kaufmann Publishers. USA.
[4] Witten, Ian H. dan Frank, Eibe. 2005. *Data Mining: Practical Machine Learning Tools and Techniques.*Edisi kedua.Morgan Kaufmann Publishers. San Fransisco.
[5] Atma, Yeyen Dwi dan Setyanto, Arief. 2018. "Perbandingan Algoritma C4.5 Dan K-NN Dalam Identifikasi Mahasiswa Berpotensi Drop Out". *Jurnal Metik.* Vol. 2 No.2.hal.31-37.
[6] Mustafa, M. S. and Simpen, I. W., 2014."Perancangan Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik Data Mining (Studi Kasus: Data Akademik Mahasiswa STMIK Dipanegara Makassar)".*Citec J.*Vol. 1. hal. 270–281.
[7] Mustakim,Giantika O, 2016 "Algoritma K-Nearest Neighbor Classification Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa."*J. Sains dan Teknol. Ind.* Vol. 13. No. 2.hal.195–202.
[8] Ndaumanu, R. I. Arief, M. R., dan Kusrini. 2014. "Analisis Prediksi Tingkat Pengunduran Diri Mahasiswa dengan Metode K-Nearest Neighbor," *Jatisi*. Vol. 1, No. 1.hal.1–15.
[9] Putri, Ratna Puspita Sari dan Waspada, Indra.Juni 2018."Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatik".*Khazanah Informatika*, Vol. 4 No. 1, Juni, 2018.hal.1-7.
[10] Risnawati.Juni 2018 "Analisis Kelulusan Mahasiswa Menggunakan Algoritma C.45". *Jurnal Mantik Penusa.*Volume 2, No. 1.hal.71-76.
[11] Rohmawan, Eko Prasetiyo. April 2018. "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode *Decision Tree* dan *Artificial Neural Network"*.Jurnal Ilmiah *Matrik* Vol.20 No.1.