

PENERAPAN ALGORITMA K-NEAREST NEIGHBOUR (K-NN) UNTUK PENENTUAN MAHASISWA BERPOTENSI DROP OUT

Ni Luh Ratniasih

Program Studi Sistem Informasi
Institut Teknologi dan Bisnis STIKOM Bali
ratni@stikom-bali.ac.id

ABSTRACT

Drop out is a situation where students are expelled from college because of several factors, one of which is because the status of lectures is not active beyond 5 semesters for undergraduate students. The high level of success and low failure of students can reflect the quality of education in higher education. The high level of student drop outs can affect the value of Higher Education accreditation so that it will affect the level of public trust. Student data drop out becomes something important to be researched and analyzed, so it can be seen how the characteristics of students who have the potential to drop out as early as possible. The data of ITB STIKOM Bali students is very much so that they can utilize data mining techniques for data classification. In this study the K-NN method was implemented to classify students as potential drop outs and the student data used in this study were students of the 2014 Information Systems study program using 6 attributes, namely gender, age, religion, class status, practical work, and grades GPA. The results showed that the accuracy of the method was 81.50%.

Keywords: KNN, Drop Out, ITB STIKOM Bali

ABSTRAK

Drop out adalah suatu keadaan dimana mahasiswa dikeluarkan dari perguruan tinggi karena beberapa faktor salah satunya karena status perkuliahannya tidak aktif melebihi 5 semester untuk mahasiswa S1. Tingginya tingkat keberhasilan dan rendahnya kegagalan mahasiswa dapat mencerminkan kualitas pendidikan di perguruan tinggi. Tingginya tingkat drop out mahasiswa dapat mempengaruhi nilai akreditasi Perguruan Tinggi sehingga akan mempengaruhi tingkat kepercayaan masyarakat. Data mahasiswa drop out menjadi sesuatu hal yang penting untuk diteliti dan dianalisa, sehingga dapat diketahui bagaimana karakteristik mahasiswa yang berpotensi drop out sedini mungkin. Data mahasiswa ITB STIKOM Bali sangat banyak sehingga dapat memanfaatkan teknik data mining untuk klasifikasi data. Pada penelitian ini diimplementasikan metode K-NN untuk klasifikasi mahasiswa berpotensi drop out dan data mahasiswa yang digunakan dalam penelitian ini adalah mahasiswa program studi Sistem Informasi angkatan 2014 dengan menggunakan 6 atribut yaitu jenis kelamin, umur, agama, status kelas, kerja praktek, dan nilai IPK. Hasil penelitian menunjukkan bahwa tingkat akurasi metode sebesar 81.50%.

Kata kunci: KNN, Drop Out, ITB STIKOM Bali

PENDAHULUAN

Data mahasiswa *drop out* menjadi sesuatu hal yang penting untuk diteliti dan dianalisa, sehingga dapat diketahui bagaimana karakteristik mahasiswa yang berpotensi *drop out* sedini mungkin. Institut Teknologi dan Bisnis (ITB) STIKOM Bali merupakan salah satu perguruan tinggi di Bali di bidang ICT (*Information Communication Technology*) yang memiliki data sangat besar. Salah satu cara untuk melakukan pengolahan data yang besar adalah dengan metode penggalian data atau *data mining*. Penggalian data berdasarkan data pendidikan di perguruan tinggi dapat meningkatkan kualitas pembelajaran mahasiswa

di Perguruan Tinggi [1]. Menurut Turban, data mining diartikan sebagai suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam basis data [2]. Teknologi *data mining* dapat dimanfaatkan untuk menggali pengetahuan di basis data ITB STIKOM Bali untuk mengetahui model yang menggambarkan karakteristik mahasiswa yang berpotensi *drop out*.

Seiring dengan perkembangan teknologi, *data mining* dengan teknik klasifikasi data dikembangkan dengan metode – metode. Terdapat beberapa penelitian yang telah dilakukan terkait klasifikasi data menggunakan metode. Salah satu penelitian terkait adalah penelitian dalam menganalisis perbandingan

algoritma classification untuk *authentication* uang kertas, dimana hasil persentase akurasi yang sangat tinggi adalah menggunakan metode *Decision Tree C4.5* dengan nilai akurasi sebesar 98,5 %, sedangkan *Neural Network* sebesar 95%, dan *Navies Bayes* sebesar 85% [3]. Selanjutnya penelitian yang menerapkan teknik *data mining* menggunakan metode *Support Vector Machine* (SVM) untuk memprediksi siswa yang berpeluang *drop out* menghasilkan tingkat akurasi sebesar 43,33% [4]. Penelitian yang mengimplementasi pengolahan citra dan klasifikasi *K-Nearest Neighbour* untuk membangun aplikasi pembeda daging sapi dan daging babi berbasis web, penelitian tersebut berhasil melakukan klasifikasi perbedaan daging babi dan daging sapi berdasarkan ekstraksi ciri tekstur dengan akurasi 88,75%[5]. Penelitian terkait klasifikasi mahasiswa *drop out* telah dilakukan pada tahun 2018 dengan judul perancangan sistem klasifikasi mahasiswa *drop out* menggunakan algoritma *K-Nearest Neighbor* dengan tempat studi kasus STMIK STIKOM Bali. Pada penelitian ini komponen atribut yang digunakan adalah Indeks Prestasi Kumulatif (IPK), jurusan, penghasilan orang tua, semester dengan menggunakan nilai k yang berbeda. Hasil penelitian ini adalah sebuah sistem yang dapat digunakan untuk mengklasifikasikan mahasiswa *drop out* namun tidak membahas tingkat akurasi dari sistem tersebut [6].

Berdasarkan uraian latar belakang di atas maka dalam penelitian ini akan dilakukan penelitian untuk klasifikasi mahasiswa *drop out* dengan menerapkan algoritma *K-Nearest Neighbour* untuk mengetahui tingkat akurasi dari metode tersebut. Data mahasiswa yang digunakan dalam penelitian ini adalah mahasiswa program studi Sistem Informasi angkatan tahun 2014 dengan menggunakan atribut yaitu jenis kelamin, umur, agama, status kelas, kerja praktek, dan nilai IPK.

LANDASAN TEORI

Definisi Data Mining

Menurut Gartner Group, *data mining* adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika [7]. *Data mining* bukanlah suatu bidang yang sama sekali baru. Salah satu kesulitan untuk mendefinisikan *data mining* adalah kenyataan bahwa *data mining* mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dulu.

Berawal dari beberapa disiplin ilmu, *data mining* bertujuan untuk memperbaiki teknik tradisional sehingga bisa menangani:

1. Jumlah data yang sangat besar
2. Dimensi data yang tinggi
3. Data yang heterogen dan berbeda bersifat

METODOLOGI PENELITIAN

Metode K-Nearest Neighbour

Algoritma *K-Nearest Neighbor* merupakan metode klasifikasi yang mengelompokkan data baru berdasarkan jarak data baru itu kedalam beberapa data tetangga (*neighbour*) terdekat. Teknik *K-Nearest Neighbor* dengan melakukan langkah-langkah yaitu, *input* : data training, label data training, k, data testing [8].

Proses yang dilakukan dalam *K-NN* untuk mendapatkan kelas kategori adalah dengan menghitung kemiripan antara data baru dengan tiap – tiap data yang sebelumnya telah dikategorikan. Dengan kata lain algoritma ini bekerja berdasarkan dari data baru terhadap data yang sudah ada atau latih. Data latih tersebut kemudian diurutkan mulai dari data yang memiliki nilai kemiripan paling besar dengan data yang akan dikategorikan. Data dipilih sebanyak *k* data dengan nilai kemiripan terbesar, kemudian memprediksikan data yang baru ke dalam kategori data tersebut. Adapun rumus untuk melakukan perhitungan kedekatan antara dua kasus dapat dilihat pada persamaan 3 sebagai berikut :

$$\text{Similarity (T,S)} = X = \frac{\sum_{i=1}^n f(T_i, S_i) \cdot w_i}{w_i}$$

Keterangan:

T : kasus baru

S : kasus yang ada dalam penyimpanan

n : jumlah atribut dalam setiap kasus

i : atribut individu antara 1 s.d. n

f : fungsi *similarity* atribut *i* antara kasus T dan kasus S

w : bobot yang diberikan pada atribut ke-*i* kedekatan biasanya berada pada nilai antara 0 s.d. 1. Nilai 0 artinya kedua kasus mutlak tidak mirip, sebaliknya untuk nilai 1 kasus mirip dengan mutlak.

Rapid Miner

Rapid miner adalah aplikasi data mining yang berbasis *open source*. *Open source* rapid miner berlisensi AGPL (*GNU Affero General Public License*) versi 3. Penelitian mengenai *tools* ini dimulai sejak tahun 2001 oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer di Artificial Intelligence Unit dari *University of Dortmund* yang kemudian diambil alih oleh SourceForge sejak tahun 2004. Rapid miner

memperoleh peringkat satu sebagai *tools* data mining untuk proyek nyata pada poll oleh KDnuggets, sebuah koran data-mining pada 2010-2011.

Dalam penerapannya, rapid miner menyediakan prosedur *data mining* dan *machine learning* termasuk : ETL (*extraction, transformation, loading*), data preprocessing, visualisasi, modelling dan evaluasi. Proses data mining tersusun atas operator-operator yang nestable, dideskripsikan dengan XML, dan dibuat dengan GUI. *Tools* rapid miner ditulis dalam bahas pemrograman Java dan juga mengintegrasikan proyek data mining Weka dan statistika [9].

Beberapa solusi yang diusung oleh rapid miner antara lain :

- Integrasi data, Analitis ETL, Data Analisis, dan pelaporan dalam satu suite tunggal.
- Powerfull tetapi memiliki antarmuka pengguna grafis yang intuitif untuk desain analisis proses.
- Repositori untuk prose, data dan penanganan meta data.
- Hanya solusi dengan transformasi meta data: lupakan trail and error dan memeriksa hasil yang telah diinspeksi selama desain.
- Hanya solusi yang mendukung *on-the-fly* kesalahan dan dapat melakukan perbaikan dengan cepat. Lengkap dan fleksibel: ratusan *loading* data, transformasi data, pemodelan data dan metode visualisasi data.

HASIL DAN PEMBAHASAN

Praproses Data

Berdasarkan hasil observasi ke Bagian Pusat Komputer dan Jaringan ITB STIKOM Bali yang telah dilakukan, data mahasiswa Program Studi Sistem Informasi disimpan pada sebuah *database*. Data awal yang diperoleh dari Bagian Pusat Komputer dan Jaringan ITB STIKOM Bali adalah berupa *excel* (.xlsx) dengan jumlah data 667 *record* data mahasiswa. Data yang diambil disesuaikan dengan komponen atribut yang digunakan yaitu kerja praktek, jenis kelamin, umur, status kelas, agama, nilai IPK dan tahun lulus seperti yang terlihat pada Gambar 1.

NIM	NAMA	KERJA PRAKTEK	JENIS KELAMIN	TGL LAHR	STATUS KELAS	AGAMA	IPK	LULUS
140030001	YULI EKO PRASETYO	Sudah Entry Nilai	LAKI	7/29/1995	Reguler	ISLAM	3.64	-
140030002	CHRISTIAN ALFI RAHMANANDA	Sudah Entry Nilai	LAKI	9/19/1996	Reguler	KRISTEN	3.70	-
140030003	DICKY PANI ANGGARA	Sudah Entry Nilai	LAKI	2/4/1996	Reguler	KRISTEN	3.93	-
140030004	PUTU PRISCA KRISTINA	Sudah Entry Nilai	LAKI	12/5/1995	Reguler	HINDU	3.27	2018
140030005	MADE YUDHA WIRAKUSUMA	Sudah Entry Nilai	LAKI	4/30/1996	Reguler	HINDU	3.67	2017
140030006	PUTU NGURAH RIZKY ANDIKA	Sudah Entry Nilai	LAKI	1/2/1996	Reguler	HINDU	3.46	-
140030007	NI MADE INTAN LESTARI	Sudah Entry Nilai	PEREMPUNAN	7/30/1995	Reguler	HINDU	3.05	2018
140030008	MADE TESNA WARDANA	Belum KP	LAKI	6/2/1996	Reguler	HINDU	3.46	-
140030009	MADE ARTA WIGUNA	Belum KP	LAKI	03/28/1995	Reguler	HINDU	0.00	-
140030010	AGUS ANANDA RAHMAT RAHA	Sudah Entry Nilai	LAKI	11/25/1989	Reguler	HINDU	0.00	-
140030011	MADE GANGGAS ARI WAHYU	Sudah Entry Nilai	LAKI	10/25/1996	Reguler	HINDU	3.00	-
140030012	PX PUTU IRVAN RIVALDY ANGG	Sudah Entry Nilai	LAKI	3/26/1996	Reguler	KATOLIK	3.42	-
140030013	WISMA HADI PURWANTI	Sudah Entry Nilai	PEREMPUNAN	3/21/1996	Reguler	ISLAM	3.36	2018
140030014	PUTU WAHYU SEDANA YOGA	Sudah Entry Nilai	LAKI	12/14/1996	Reguler	HINDU	3.46	2017
140030015	WAYAN ANGGA CIPTA PRATA	Sudah Entry Nilai	LAKI	11/2/1996	Reguler	HINDU	3.36	2017
140030016	LENI MARTHA DEWI	Sudah Entry Nilai	PEREMPUNAN	1/4/1996	Reguler	HINDU	3.29	2017
140030017	PUTU WIWEN KRISYANTI	Sudah Entry Nilai	PEREMPUNAN	10/24/1996	Reguler	HINDU	3.42	2017
140030018	PETER ALLAN CORNELIUS SERO	Sudah Entry Nilai	LAKI	3/20/1996	Bayar Dimuka	KATOLIK	3.65	2018
140030019	YOGA PRATAMA	Sudah Entry Nilai	LAKI	3/21/1995	Reguler	ISLAM	3.54	2017
140030020	KE WAHYUNING TIAS	Sudah Entry Nilai	PEREMPUNAN	3/21/1996	Reguler	ISLAM	3.32	-
140030021	AA ISTRI INDRYARTHI DEWI	Sudah Entry Nilai	PEREMPUNAN	12/16/1996	Reguler	HINDU	3.73	2017
140030022	PUTU AGUS DARMA ADI	Sudah Entry Nilai	LAKI	7/21/1994	Reguler	HINDU	3.14	2018
140030023	BAJU GEM ASMORO	Sudah Entry Nilai	LAKI	10/22/1996	Reguler	ISLAM	3.63	2017
140030024	NI LUH GEDE BUDI RAHAYU	Sudah Entry Nilai	PEREMPUNAN	12/28/1994	Reguler	HINDU	3.21	2017

Gambar 1. Data Awal

Penyesuaian data awal dengan atribut – atribut yang digunakan dalam penelitian dilakukan sebagai berikut:

- Atribut **kerja praktek** merupakan atribut yang menjelaskan apakah mahasiswa sudah menyelesaikan kerja praktek atau tidak. Pada *database* atribut kerja praktek (KP) diperoleh dua status yaitu “Sudah Entry Nilai” dan “Belum Kerja Praktek” sehingga status akan disesuaikan. Jika status adalah “Sudah Entry Nilai” maka status ditulis “SUDAH” sedangkan status “Belum Kerja Praktek” status ditulis “TIDAK”.
- Atribut **jenis kelamin** terdiri dari dua jenis type yaitu “Laki – laki” dan “Perempuan” yang diperoleh lgs dari data awal *database*.
- Atribut **umur** tidak secara spesifik diperoleh dikarenakan yang tersimpan pada *database* adalah tanggal lahir sehingga umur dihitung menggunakan formula yang ada pada *Microsoft Excel*.
- Atribut **status kelas** merupakan status yang dipilih oleh mahasiswa pada saat mendaftar pertama kali, dimana pada *database* tersimpan empat jenis status kelas yaitu “Reguler”, “Bayar Dimuka”, ‘eksekutif” dan “Investasi”. Jika status kelas adalah “Reguler” maka status ditulis “TIDAK”, sedangkan jika status kelas “Investasi” / “Bayar Dimuka” / “Eksekutif” maka status akan ditulis “INVESTASI”.
- Atribut **agama** terdiri dari lima status diantaranya “Islam”, “Kristen”, “Hindu”, “Katolik”, “Budha” yang diperoleh langsung dari *database*.
- Atribut **IPK** merupakan nilai Indeks Prestari Kumulatif di semester terakhir dari masing – masing mahasiswa.
- Atribut **lulus** merupakan keterangan yang diperoleh dari tahun lulus masing – masing mahasiswa dimana jika mahasiswa lulus tahun 2018 maka statusnya adalah “TEPAT” jika tahun lulus di atas tahun 2018 maka statusnya “TIDAK”.

Setelah data disesuaikan maka dilakukan praproses data. Praproses data dilakukan untuk

mendapatkan data yang relevan dan siap digunakan. Ada beberapa tahapan dalam melakukan praproses data yaitu pembersihan data, pemilihan data, integrasi serta transformasi data. Pembersihan data dilakukan untuk menghilangkan *missing value* atau membuang data yang tidak konsisten. Pemilihan data merupakan pengambilan data yang relevan untuk dijadikan fitur integrasi data dilakukan dengan menggabungkan data dari sumber yang berbeda yang kemudian siap diolah untuk tahap selanjutnya. Terdapat beberapa data dihilangkan karena nilai IPK 0 seperti yang terlihat pada Gambar 2.

NIM	NAMA	KP	JENIS KE...	UMUR	STATUS	AGAMA	IPK	LULUS
140030051	OVI AYU TREESNANIE	SUDAH	PEREMPUAN	22	TIDAK	HINDU	3,54	TIDAK
140030052	NI NYOMAN DARMA YANTI	SUDAH	PEREMPUAN	22	TIDAK	HINDU	3,54	TEPAT
140030053	WELLY RACHMAT KUNCOR	TIDAK	LAKI	22	TIDAK	ISLAM	3,75	TEPAT
140030054	GEDE FARINDRA ADITYA	SUDAH	LAKI	23	TIDAK	HINDU	3,00	TIDAK
140030055	LINGGA WEGIG GABIS SULLU	SUDAH	LAKI	23	TIDAK	ISLAM	3,71	TEPAT
140030056	I MADE YUDHI SEMARA	SUDAH	LAKI	23	TIDAK	HINDU	3,25	TEPAT
140030059	I NYOMAN YUDI MAHENDRI	TIDAK	LAKI	23	TIDAK	HINDU	3,18	TEPAT
140030060	I KETUT WAHYU RYANANT	SUDAH	LAKI	21	TIDAK	HINDU	1,29	TIDAK
140030061	YOSEPH KURNIADY NUYA	SUDAH	LAKI	23	TIDAK	KATOLIK	3,46	TIDAK
140030062	I PUTU ARIE WILLIYANTAR	SUDAH	LAKI	24	TIDAK	HINDU	3,27	TEPAT
140030063	I DW GEDE AGUS WARDAN	SUDAH	LAKI	24	TIDAK	HINDU	2,92	TEPAT
140030064	I PUTU MAYDA SAVITRI	SUDAH	PEREMPUAN	23	TIDAK	HINDU	3,55	TIDAK
140030065	I PUTU AGUS EKA WIBAWA	TIDAK	LAKI	23	TIDAK	HINDU	1,73	TIDAK
140030066	MEGA WIDI LANINGRUM	SUDAH	PEREMPUAN	23	TIDAK	ISLAM	3,00	TIDAK
140030067	IDA BAGUS ANANDA S. NAI	SUDAH	LAKI	23	TIDAK	HINDU	3,17	TIDAK
140030068	SERVINA LUTFI GUNTARY	SUDAH	PEREMPUAN	26	TIDAK	ISLAM	5,83	TEPAT
140030069	AYU KARTIKA SARI	SUDAH	PEREMPUAN	23	TIDAK	ISLAM	3,75	TEPAT
140030070	NOER MOCHAMAD RIVAN	SUDAH	LAKI	23	TIDAK	ISLAM	3,50	TEPAT
140030071	NI KADEK NILA FRIANA PU	SUDAH	PEREMPUAN	23	TIDAK	HINDU	3,42	TEPAT
140030072	I GEDE PUTU PUTRA WIRAN	SUDAH	LAKI	23	TIDAK	HINDU	3,75	TEPAT
140030073	I PUTU GEDE CANDRA UDA	SUDAH	LAKI	22	TIDAK	HINDU	3,23	TEPAT
140030074	I NYOMAN RESTU BAWAH	TIDAK	LAKI	24	TIDAK	HINDU	3,83	TIDAK
140030075	DIAN TRISNAWATI	SUDAH	PEREMPUAN	37	INVESTASI	ISLAM	3,30	TIDAK
140030076	RISKA NOPITA SARI	SUDAH	PEREMPUAN	23	TIDAK	ISLAM	3,05	TIDAK
140030078	NI KD WINDA PATRIANING	SUDAH	PEREMPUAN	22	INVESTASI	HINDU	3,79	TEPAT

Gambar 1 Nilai IPK 0.00

Setelah dilakukan praproses data diperoleh *data cleaning* yang siap dipakai pada penelitian ini sebanyak 519 *record* dengan 9 atribut seperti yang terlihat pada Gambar 5.

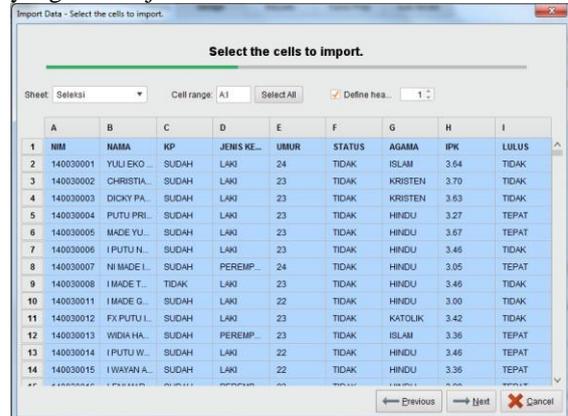
A	B	C	D	E	F	G	H	I
NIM	NAMA	KP	JENIS KELAMIN	UMUR	STATUS	AGAMA	IPK	LULUS
140030001	YULI EKO PRASETYO	SUDAH	LAKI	24	TIDAK	ISLAM	3,64	TIDAK
140030002	CHRISTIAN ALFI RAHMANS	SUDAH	LAKI	23	TIDAK	KRISTEN	3,70	TIDAK
140030003	DICKY PANJI ANGGARA	SUDAH	LAKI	23	TIDAK	KRISTEN	3,63	TIDAK
140030004	PUTU PRISCA KRISTINA	SUDAH	LAKI	23	TIDAK	HINDU	3,27	TEPAT
140030005	MADE YUDHA WIRAKUSUN	SUDAH	LAKI	23	TIDAK	HINDU	3,67	TEPAT
140030006	I PUTU NGURAH RIZKY ANI	SUDAH	LAKI	23	TIDAK	HINDU	3,46	TIDAK
140030007	NI MADE INTAN LESTARI	SUDAH	PEREMPUAN	24	TIDAK	HINDU	3,05	TEPAT
140030008	I MADE TRESNA WARDAN	TIDAK	LAKI	23	TIDAK	HINDU	3,46	TIDAK
140030011	I MADE GANGGAS ARI WA	SUDAH	LAKI	22	TIDAK	HINDU	3,00	TIDAK
140030012	FX PUTU IRVAN RIVALDY A	SUDAH	LAKI	23	TIDAK	KATOLIK	3,42	TIDAK
140030013	WIDIA HADI PURWANTI	SUDAH	PEREMPUAN	23	TIDAK	ISLAM	3,36	TEPAT
140030014	PUTU WAHYU SEDANA YESU	SUDAH	LAKI	22	TIDAK	HINDU	3,46	TEPAT
140030015	I WAYAN ANGGA CIPTA PR	SUDAH	LAKI	22	TIDAK	HINDU	3,36	TEPAT
140030016	LENI MARTHA DEWI	SUDAH	PEREMPUAN	23	TIDAK	HINDU	3,29	TEPAT
140030017	PUTU WIWEN KRISTYANTI	SUDAH	PEREMPUAN	22	TIDAK	HINDU	3,42	TEPAT
140030018	PIETER ALLAN CORNELIS S	SUDAH	LAKI	23	INVESTASI	KATOLIK	3,65	TEPAT
140030019	YOGA PRATAMA	SUDAH	LAKI	24	TIDAK	ISLAM	3,54	TEPAT
140030020	IKE WAHYUNING TIAS	SUDAH	PEREMPUAN	23	TIDAK	ISLAM	3,32	TIDAK
140030021	AA ISTRI INDIYARTHI DEW	SUDAH	PEREMPUAN	22	TIDAK	HINDU	3,73	TEPAT
140030022	PUTU AGUS DARMA ADI	SUDAH	LAKI	25	TIDAK	HINDU	3,14	TEPAT
140030023	BAYU GENI ASMORO	SUDAH	LAKI	22	TIDAK	ISLAM	3,63	TEPAT
140030024	NI LUH GEDE BUDI RAHAY	SUDAH	PEREMPUAN	24	TIDAK	HINDU	3,21	TEPAT
140030025	JESICHA PUTRI HANDAYAN	SUDAH	PEREMPUAN	22	TIDAK	KRISTEN	3,83	TEPAT

Gambar 2 Data Siap Diolah

a. Pengujian Algoritma

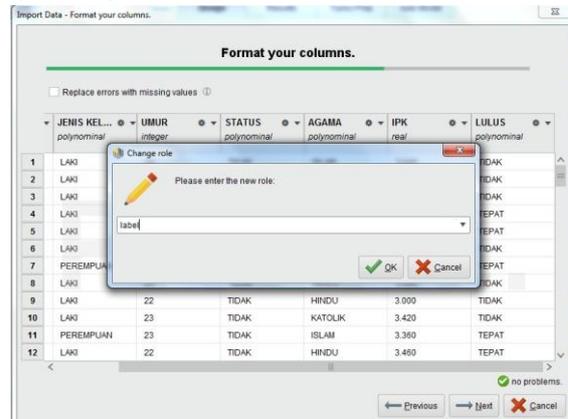
Pada tahap ini dilakukan pengujian algoritma *K-Nearest Neighbour* untuk mengetahui tingkat akurasi metode tersebut. Perhitungan tingkat akurasi dilakukan menggunakan aplikasi *RapidMiner Studio*. Dari tingkat akurasi yang dihasilkan dapat menentukan tinggi rendahnya kesalahan prediksi pada adat baru. semakin tinggi tingkat akurasi maka semakin rendah kesalahan prediksi pada data baru demikian sebaliknya. Proses perhitungan tingkat akurasi dilakukan dengan metode *Confussion Matrix*.

Tahap awal pengujian adalah *import data* yang akan diproses seperti Gambar 4, kemudian pemilihan cell range dari cell C sampai I selanjutnya menentukan kategori dan atribut yang akan dijadikan label.

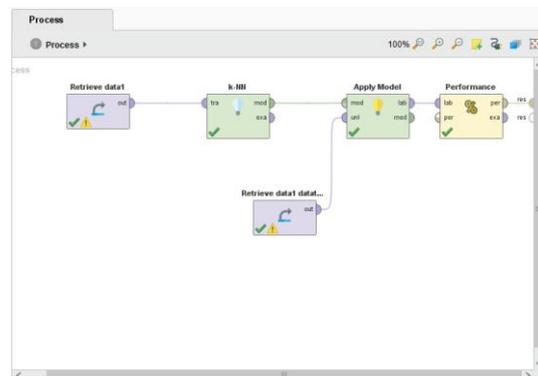


Gambar 3 Import Data

Atribut yang dijadikan sebagai label adalah atribut "LULUS" yang berada di cell I seperti Gambar 5. Tahap selanjutnya adalah melakukan pengujian dengan algoritma *K-Nearest Neighbour* sehingga terbentuk pemodelan *K-Nearest Neighbour* seperti yang terlihat pada Gambar 6.



Gambar 4 Menentukan Label



Gambar 5 Pemodelan Algoritma K-Nearest Neighbour

Pada Gambar 5 terdapat parameter dan operator yang digunakan pada model algoritma *K-Nearest Neighbour* sebagai berikut :

- a. *Retrieve data* dan *retrieve data dataset* adalah operator yang digunakan untuk mengimport dataset yang akan digunakan, pada penelitian ini data diimport dari file excel.
- b. *K-Nearest Neighbour* adalah metode klasifikasi yang digunakan dalam penelitian ini.
- c. *Apply Model* adalah operator yang digunakan dalam penelitian ini Algoritma *K-Nearest Neighbour*.
- d. *Performance* adalah operator yang digunakan untuk mengukur performa akurasi dari model.

Berdasarkan hasil *performance* algoritma pada diperoleh tingkat akurasi (*accuracy*) sebesar 81.50%.

accuracy: 81.50%

	true TIDAK	true TEPAT	class precision
pred. TIDAK	99	32	75.57%
pred. TEPAT	64	324	83.51%
class recall	60.74%	91.01%	

Dibuktikan dengan perhitungan akurasi menggunakan *confussion matrix* :

Tabel 1 Aturan Confussion Matrix

	Memenuhi	Tidak Memenuhi
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

- a. *True Positive* (TP), data yang diprediksi masuk kategori TIDAK, dan data sebenarnya memang TIDAK.
- b. *True Negative* (TN), data yang diprediksi masuk kategori TEPAT, dan data sebenarnya TEPAT.
- c. *False Positive* (FP), data yang diprediksi masuk TIDAK, dan data sebenarnya adalah TEPAT
- d. *False Negative* (FN), data yang diprediksi masuk kategori TEPAT, dan data sebenarnya adalah TDAK

Berikut rumus perhitungan nilai akurasi:

$$\frac{TP + TN}{TP + TN + FP + FN} \dots\dots\dots(5.1)$$

$$\begin{aligned} \text{Akurasi} &= \frac{99 + 324}{324 + 32 + 99 + 64} \times 100\% \\ &= \frac{423}{519} \times 100\% \\ &= 81.50\% \end{aligned}$$

SIMPULAN

Penerapan algoritma *K-Nearest Neighbour* untuk klasifikasi mahasiswa yang berpotensi drop out telah dilakukan menggunakan aplikasi RapidMiner Studio, dengan tingkat akurasi sebesar 81.50%.

DAFTAR PUSTAKA

- [1] Tair, M. Mohammed Abu dan El-halees, Alaa M., “Mining Educational Data to Improve Students Performance: A Case Study”, *International Journal of Information and Communication Technology Research*, Vol. 2, No. 2, 140-146, 2012.
- [2] Turban, Efraim, Aronson, Jay E. dan Liang, Ting-Peng, “Decision Support System and Intelligent System”, Prentice-Hall of India: New Delhi, 7th ed, 2007.
- [3] Khairul Sani, Wing Wahyu Winarno, Silmi Fauziati, 2016, Analisis Perbandingan Algoritma Classification Untuk Authentication Uang Kertas (Studi Kasus: Banknote Authentication), *Jurnal Informatika* Vol. 10, No. 1.
- [4] Fiska, Ryci Rahmatil, 2017, Penerapan Teknik Data Mining dengan Metode Support Vector Machine (SVM) untuk Memprediksi Siswa yang Berpeluang Drop Out (Studi Kasudi SMKN 1 Sutera), *Jaringan Sistem Informasi Robotik* Vol. 1, No. 01.
- [5] Dea Alverina, A. R. Chrismanto, and R. G. Santosa, 2018, Perbandingan Akurasi Algoritma C4.5 dan CART dalam Memprediksi Kategori Indeks Prestasi Mahasiswa, *Jurnal Teknologi dan Sistem Komputer*, vol. 6, no. 2, Apr. 2018. doi: 10.14710/jtsiskom.6.2.2018.76-83, ISSN:2338-0403.
- [6] Ramayasa, I Putu. 2018. “Perancangan Sistem Klasifikasi Mahasiswa Drop Out Menggunakan Algoritma K-Nearest Neighbor”. *Seminar Nasional Sistem Informasi dan Teknologi Informasi 2018*. STMIK Pontianak. 585 – 589.
- [7] Larose, D.T, 2006. *Discovering Knowledge in Data: An Introduction to Data mining*. John Willey & Sons, Inc.
- [8] Santosa B. 2007. *Data Mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis, Teori dan Aplikasi*. Graha Ilmu Yogyakarta.
- [9] Rapid-I GmbH. (2008). *Rapidminer-4.2-tutorial*. Germany: Rapid-I