

KAJIAN DATA MINING
PREDIKET KELULUSAN
MAHASISWA TEKNIK
INFORMATIKA UNIVERSITAS
ISLAM RIAU DENGAN
ALGORITMA C4.5

by Fakultas Teknik

Submission date: 19-Jan-2024 02:41PM (UTC+0700)

Submission ID: 2273776174

File name: Prosiding_10.pdf (256.38K)

Word count: 2444

Character count: 14888

KAJIAN DATA MINING PREDIKET KELULUSAN MAHASISWA TEKNIK INFORMATIKA UNIVERSITAS ISLAM RIAU DENGAN ALGORITMA C4.5

Des Suryani¹, Ause Labellapansa²

Jurusan Teknik Informatika Fakultas Teknik Universitas Islam Riau
Jalan Kaharuddin Nasution No. 113 Marpoyan Pekanbaru, Riau. Telp (0761) 674674
Email : d3ssuryani@gmail.com, ause.labella@eng.uir.ac.id

Abstrak

Kelulusan adalah target utama yang diharapkan mahasiswa dalam menempuh pendidikan di perguruan tinggi. Namun dalam kenyataannya masih banyak mahasiswa khususnya di program studi Teknik Informatika fakultas Teknik Universitas Islam Riau yang belum dapat mencapai standar prediket kelulusan. Standar prediket kelulusan mahasiswa sangat dipengaruhi oleh standar penerimaan mahasiswa. perguruan tinggi sebagai penyelenggara pendidikan juga harus memperhatikan standar dalam penerimaan mahasiswa. Berdasarkan hal tersebut, perlu dilakukan evaluasi terhadap kumpulan data kelulusan untuk menentukan pola kelulusan mahasiswa. Proses data mining menggunakan metode Klasifikasi dengan algoritma Decision Tree C4.5. Atribut yang digunakan dalam penentuan pola prediket kelulusan mahasiswa tersebut adalah jenis SLTA, jurusan di SLTA, waktu tunggu kuliah, nilai matematika di SLTA dan prediket kelulusan sebagai atribut label/class. Hasil kajian berupa pola prediket kelulusan yang dapat memberikan kebijakan yang lebih baik bagi jurusan Teknik Informatika dalam menentukan kriteria seleksi penerimaan mahasiswa baru sehingga dapat meningkatkan lulusan yang memenuhi standar kelulusan yang telah ditetapkan universitas.

Kata kunci : *data mining, decision tree C4.5, klasifikasi, kelulusan*

1. Pendahuluan

Perguruan Tinggi sebagai penyelenggara pendidikan dituntut untuk dapat melaksanakan Tridharma Perguruan Tinggi dengan baik. Keberhasilan sebuah perguruan tinggi salah satunya ditentukan dari lulusannya, dimana semakin pendek masa studi dan semakin tinggi nilai IPK maka kelulusan semakin baik. Untuk itu semua perguruan tinggi baik negeri maupun swasta harus mempersiapkan diri dalam penilaian penyelenggaraan pendidikan oleh pemerintah melalui proses akreditasi baik institusi maupun program studi. Hal ini dibuktikan banyaknya instansi maupun perusahaan mensyaratkan nilai akreditasi minimal B dalam penerimaan pegawai.

Mahasiswa merupakan salah satu aspek penting dalam evaluasi keberhasilan penyelenggaraan program studi pada suatu perguruan tinggi. Pemantauan mahasiswa yang masuk, peningkatan kemampuan mahasiswa, prestasi yang dicapai mahasiswa, rasio kelulusan terhadap jumlah total mahasiswa, dan kompetensi lulusan seyogyanya mendapatkan perhatian yang serius untuk memperoleh kepercayaan *stakeholder* dalam menilai dan menetapkan penggunaan lulusannya.

Universitas Islam Riau (UIR) merupakan universitas swasta yang mempunyai jumlah mahasiswa terbesar di Kopertis Wilayah X Sumbar, Riau, Jambi dan Kepri. Perguruan tinggi ini memiliki 9 (sembilan) fakultas dan satu program Pasca Sarjana dan salah satu fakultasnya adalah Fakultas Teknik (FT) yang memiliki 6 (enam) program studi. Untuk program studi Teknik Informatika (TI) termasuk program studi yang sangat diminati. Namun data kelulusannya, masih banyak yang belum memenuhi standar kelulusan yang ditetapkan institusi yaitu IPK minimal 2.76 dengan prediket kelulusan Sangat Memuaskan.

Beberapa penelitian telah banyak dilakukan dengan menggunakan teknik *data mining* untuk menggali berbagai informasi dari sebuah *database* mahasiswa. Penelitian data mining untuk menentukan berapa banyak dan mata kuliah apa yang harus dipilih oleh mahasiswa baru. Algoritma yang digunakan adalah C4.5. Terdapat beberapa atribut yang digunakan yaitu jumlah matakuliah yang diambil secara simultan, nama matakuliah dan IPK. Sebagai contoh rule yang dihasilkan yaitu *courses enrolled in >5 and courses enrolled in <=6* , *Name of Course =dinamika kontrol*, $13,43 < IPK < 13,0$ maka kelas yang harus diambil adalah APROB dengan nilai 96,0% [5]. Vialardi, dkk (2011) melakukan penelitian data mining untuk membimbing mahasiswa dalam proses pendaftaran matakuliah yang akan diambil semester

berikutnya dilihat dari performa akademik mahasiswa tersebut. Terdapat dua atribut utama dalam proses persiapan data. Pertama yaitu tingkat kesulitan mata kuliah yang akan diambil. Data ini diperoleh dari nilai mata kuliah tersebut sebelumnya. Kedua yaitu potensi dari mahasiswa yang dinyatakan dalam bentuk numerik berupa keahlian mahasiswa terutama pada matakuliah yang utama. Pada penelitian ini digunakan 4 metode yaitu C4.5, KNN, Naive Bayes dan Bagging and Boosting dengan hasil algoritma yang paling efisien yaitu C4.5[6].

Berdasarkan hal tersebut, perlu dilakukan kajian dalam penelitian terhadap data lulusan untuk mengetahui urutan atribut penentu dalam mencapai standar kelulusan agar mahasiswa yang diterima berikutnya sesuai dengan pola yang diperoleh dari hasil kajian tersebut. Penelitian ini menggunakan metode klasifikasi dengan algoritma *Decision Tree* C4.5 untuk menemukan pengetahuan berupa pola/model kelulusan mahasiswa TI tersebut. Data yang digunakan dalam penelitian ini berupa data kelulusan mahasiswa TI FT-UIR dari angkatan 2009 s/d 2011 dengan atribut : jenis SLTA, jurusan di SLTA, waktu tunggu kuliah, nilai matematika di SLTA dan prediket kelulusan sebagai atribut *class/label*. Diharapkan hasil penelitian ini dapat membantu jurusan untuk menentukan kebijakan yang lebih baik dalam seleksi penerimaan mahasiswa baru dan meningkatkan jumlah lulusan yang memenuhi standar kelulusan yang ditetapkan.

2. Metode Penelitian

Penelitian adalah usaha menemukan dengan proses yang metodis untuk memperkaya pengetahuan itu sendiri serta dengan yang lainnya, oleh penemuan fakta dan wawasan tidak biasa. Ada empat metode penelitian yang umum digunakan, diantaranya: Experiment, Action Research, Survey dan Case Study.[1]

Penelitian ini menggunakan metode eksperimen yaitu penelitian yang menyangkut tindakan pada parameter atau variabel terkait dan memakai tes yang dikendalikan oleh si peneliti itu sendiri.

2.1 Langkah-Langkah Penelitian

Sebagai suatu rangkaian proses, *data mining* dapat dibagi menjadi beberapa tahap proses. Tahap-tahap dalam penelitian adalah sebagai berikut:

1. Himpunan data (*Data set*)

Pada tahap ini perlu dilakukan pemahaman dan pengolahan data dengan cara:

- a. Pembersihan data (*data cleaning*). Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan.
- b. Integrasi data (*data integration*)
Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru. Seluruh data kelulusan digabungkan mulai dari angkatan 2009 sampai 2011.
- c. Seleksi data (*data selection*)
Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*. Dalam hal ini data yang diperlukan data kelulusan mahasiswa TI FT-UIR dari angkatan 2009 s/d 2011 dengan atribut : kelompok SLTA, jurusan di SLTA, waktu tunggu kuliah, nilai matematika di SLTA dan prediket kelulusan
- d. Transformasi data (*data transformation*)
Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*. Format atribut data yang digunakan adalah jenis SLTA (kode-SLTA), jurusan di SLTA (kode-jrs), waktu tunggu kuliah (waktu-tunggu), nilai matematika di SLTA (kode-Mtk) dan prediket kelulusan (prediket) sebagai atribut label/class.

2. Proses *mining*

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Dalam hal ini menggunakan metode klasifikasi dengan algoritma *Decision Tree* C4.5 yang menghasilkan pengetahuan berupa pola/model/tree/rule.

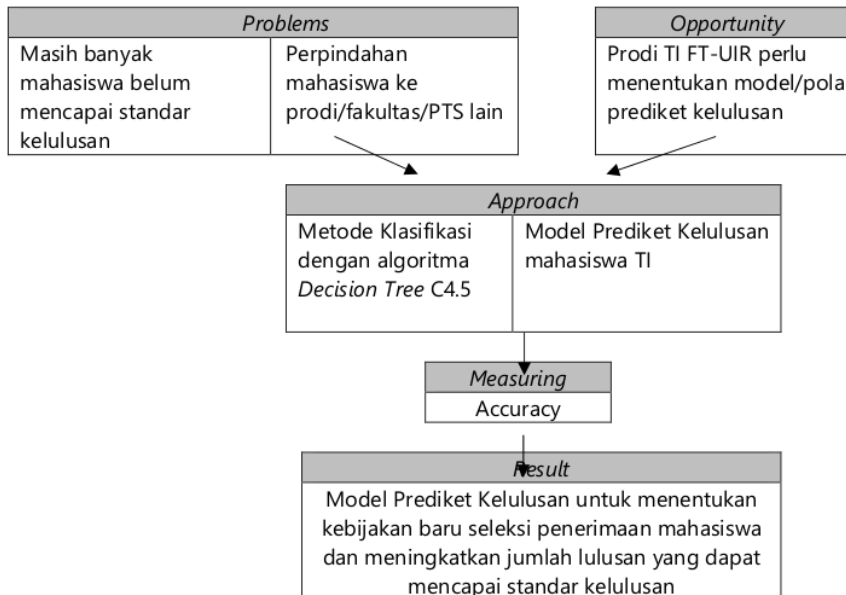
3. Evaluasi pola (*pattern evaluation*)

Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan. Evaluasi yang digunakan adalah *Accuracy*.

4. Presentasi pengetahuan (*knowledge presentation*)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

Untuk menentukan alur logika jalannya penelitian ini perlu dibuatkan kerangka pemikiran dalam penelitian ini seperti gambar 1.



Gambar 1. Kerangka Pemikiran Penelitian

2.2 Data Penelitian

Data yang diperoleh dalam penelitian ini bersumber dari berbagai informasi mengenai permasalahan yang dibahas, penulis melakukan studi pustaka, membaca dan mempelajari dokumen-dokumen, buku-buku acuan serta sumber lainnya yang berkaitan dengan penelitian untuk dijadikan referensi. Data yang digunakan dalam penelitian ini berupa data kelulusan mahasiswa TI FT-UIR dari angkatan 2009 s/d 2011 dengan atribut : kelompok SLTA, jurusan di SLTA, waktu tunggu kuliah, nilai matematika di SLTA dan prediket kelulusan sebagai atribut *class/label*.

2.3 Lokasi Penelitian

Penelitian bertempat di Program Studi Teknik Informatika Fakultas Teknik Universitas Islam Riau, Jalan Kaharuddin Nasution No. 113 Marpoyan Pekanbaru.

2.4 Metode dan Algoritma Penelitian

2.4.1 Decision Tree

Classification adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Model itu sendiri bisa berupa aturan "jika-maka", berupa *decision tree*, formula matematis atau *neural network*.

Decision tree adalah salah satu metode *classification* yang paling populer karena mudah untuk diinterpretasi oleh manusia. Disini setiap percabangan menyatakan kondisi yang harus dipenuhi dan tiap ujung pohon menyatakan kelas data. Algoritma *decision tree* yang paling terkenal adalah C4.5.

Proses *classification* biasanya dibagi menjadi dua fase : *learning* dan *test*. Pada fase *learning*, sebagian data yang telah diketahui kelas datanya diumpangkan untuk membentuk model perkiraan. Kemudian pada fase *test* model yang sudah terbentuk diuji dengan sebagian data lainnya untuk mengetahui akurasi dari model tersebut. Bila akurasinya mencukupi model ini dapat dipakai untuk prediksi kelas data yang belum diketahui.

2.4.2 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 Karena pohon keputusan memadukan antara eksplorasi data dan pemodelan, dia sangat bagus sebagai langkah awal dalam proses pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain. Sebuah pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Sebuah model pohon keputusan terdiri dari sekumpulan aturan untuk membagi sejumlah populasi yang heterogen menjadi lebih kecil, lebih *homogen* dengan memperhatikan pada variabel tujuannya. Sebuah pohon keputusan mungkin dibangun dengan seksama secara manual, atau dapat tumbuh secara otomatis dengan menerapkan salah satu atau beberapa algoritma pohon keputusan untuk memodelkan himpunan data yang belum terklasifikasi. Variabel tujuan biasanya dikelompokkan dengan pasti dan model pohon keputusan lebih mengarah pada perhitungan probabilitas dari masing-masing record terhadap kategori kategori tersebut, atau untuk mengklasifikasi record dengan mengelompokkannya dalam satu kelas.

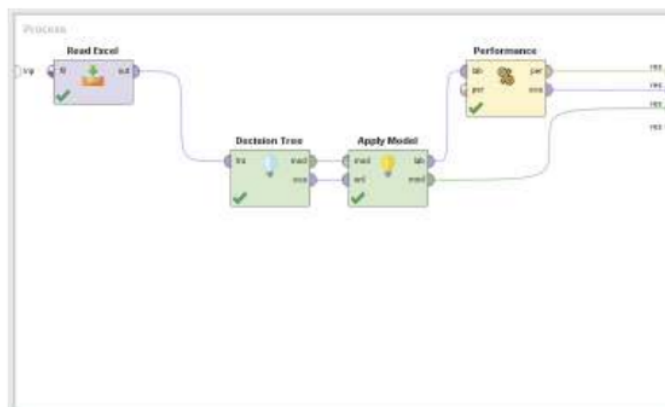
Pohon keputusan juga dapat digunakan untuk mengestimasi nilai dari variabel *continue*, meskipun ada beberapa teknik yang lebih sesuai untuk kasus ini. Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan antara lain ID3, CART dan C4.5.

Algoritma C4.5 merupakan pengembangan dari algoritma ID3. Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan record. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan pohon.

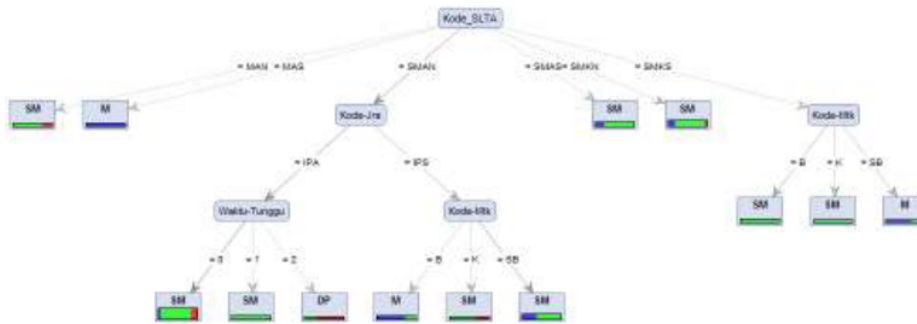
3. Hasil dan Pembahasan

Setelah diadakan eksperimen dengan melakukan pengolahan data menggunakan algoritma *Decision Tree C4.5* terhadap data kelulusan mahasiswa program studi Teknik Informatika FT-UIR mulai angkatan 2009 sampai 2011 sebanyak 107 record dengan 4 (empat) atribut regular yaitu jenis SLTA (kode-SLTA), jurusan di SLTA (kode-jrs), waktu tunggu kuliah (waktu-tunggu), nilai matematika di SLTA (kode-Mtk) dan 1 (satu) atribut *label/class* yaitu prediket kelulusan (prediket).

Proses data mining terhadap himpunan data (dataset) ini dilakukan pengolahan dengan menggunakan algoritma C4.5 melalui aplikasi RapidMiner Studio versi 7.0 seperti pada gambar 2 yang membentuk pohon keputusan (*Decision Tree*) seperti pada gambar 3.



Gambar 2. Proses Pemodelan dalam Rapidminer



Gambar 3. Pohon Keputusan (*Decision Tree*) Prediket Kelulusan Mahasiswa TI FT-UIR

Berdasarkan pohon keputusan pada gambar 3 dapat dijelaskan rule yang dihasilkan sebagai berikut :

Tree	RULE :
Kode_SLTA = MAN: SM {M=0, SM=3, DP=1}	1. IF Kode_SLTA = MAN THEN Prediket = SM
Kode_SLTA = MAS: M {M=4, SM=0, DP=0}	2. IF Kode_SLTA = MAS THEN Prediket = M
Kode_SLTA = SMAN	3. IF Kode_SLTA = SMAN THEN
Kode-Jrs = IPA	IF Kode-Jrs = IPA
Waktu-Tunggu = 0: SM {M=3, SM=29, DP=6}	IF Kode-WT = 0 THEN Prediket = SM
Waktu-Tunggu = 1: SM {M=0, SM=5, DP=0}	IF Kode-WT = 1 THEN Prediket = SM
Waktu-Tunggu = 2: DP {M=0, SM=1, DP=2}	IF Kode-WT = 2 THEN Prediket = DP
Kode-Jrs = IPS	IF Kode-Jrs = IPS
Kode-Mtk = B: M {M=3, SM=1, DP=0}	IF Kode-Mtk = B THEN Prediket = M
Kode-Mtk = K: SM {M=0, SM=2, DP=1}	IF Kode-Mtk = K THEN Prediket = SM
Kode-Mtk = SB: SM {M=5, SM=8, DP=0}	IF Kode-Mtk = SB THEN Prediket = SM
Kode_SLTA = SMAS: SM {M=2, SM=6, DP=0}	4. IF Kode_SLTA = SMAS THEN Prediket = SM
Kode_SLTA = SMKN: SM {M=3, SM=12, DP=1}	5. IF Kode_SLTA = SMKN THEN Prediket = SM
Kode_SLTA = SMKS	6. IF Kode_SLTA = SMKS
Kode-Mtk = B: SM {M=0, SM=4, DP=0}	IF Kode-Mtk = B THEN Prediket = SM
Kode-Mtk = K: SM {M=0, SM=2, DP=0}	IF Kode-Mtk = K THEN Prediket = SM
Kode-Mtk = SB: M {M=2, SM=1, DP=0}	IF Kode-Mtk = SB THEN Prediket = M

Tingkat akurasi yang diperoleh dari 107 record data latih yang digunakan dalam pemodelan *Decision Tree* C4.5 dapat dilihat pada gambar 4.

Kategori	Rata-Rata	Rata-Dev	Rata-DB	Tingkat akurasi
pred IP	0	0	0	81.52%
pred SM	10	11	9	76.84%
pred DP	0	1	0	88.07%
rata-rata	40.21%	35.36%	38.12%	

Gambar 4. Tingkat Akurasi Pemodelan *Decision Tree* C4.5

Berdasarkan gambar di atas, terdapat 9 record diprediksi secara benar memperoleh prediket Memuaskan (M), 71 record diprediksi secara benar prediket Sangat Memuaskan (SM) dan 2 record diprediksi benar Dengan Pujian (DP) dengan tingkat akurasi = $(9+71+2)/107=76.64\%$.

4. Simpulan

Algoritma Decision Tree (C4.5) merupakan salah satu algoritma yang digunakan dalam metode klasifikasi yang kuat dan terkenal. dan memiliki performa yang baik. Hasil training yang dilakukan terhadap data kelulusan dapat membentuk pengetahuan berupa model pohon yang mengubah data yang besar menjadi pohon keputusan yang merepresentasikan aturan. Berdasarkan aturan yang dihasilkan ternyata dari 107 record data lulusan, urutan lulusan yang dapat memenuhi standar kelulusan yaitu kelas SMAN jurusan IPA (40.19%), SMKN (12,15%), SMAN jurusan IPS (11,21%), SMKS (6,5%), SMAS (5,6%), MAN (3,72%) dan MAS (0%). Tingkat akurasi yang diperoleh dari pengolahan data mencapai 76,64%. Dengan hasil ini dapat memberikan kebijakan baru dalam penerimaan mahasiswa baru. Adapun saran untuk pengembangan atau lanjutan penelitian berikutnya dapat membandingkan dengan algoritma lainnya.

Daftar Pustaka

- [1] Dawson, C. Project in Computing and Information Systems. England: Pearson Education. 2009.
- [2] Han, J., M. Kambetior. Data Mining: Concepts and Techniques. *Second Edition*. Morgan Kaufmann Publishers. San Francisco. 2006.
- [3] Larose, D.T. Discovering Knowledge in Data: An Introduction to Data Mining. John Willey & Sons, Inc. 2005
- [4] McLeod, Jr.R., G.P. Schell. Management Information System. 10 th ed. Pearson Education, Inc. Ali Akbar Yulianto dan Afia R. Fitriati (penterjemah). 2008. Sistem Informasi Manajemen. Edisi 10. Nina Setyaningsih (editor). Salemba Empat. Jakarta. 2007.
- [5] Vialardi, C., Bravo, J., Shafti, L., Ortigosa, A. *Recommendation in Higher Education Using Data Mining Techniques*. In International Conference on Educational Conference. Cordoba, Spain. 2009:190-198.
- [6] Vialardi, C., Chue, J., Peche, J., Alvarado, G., Vinatea, B., Estrella, J., & Ortigosa, L. *A Data Mining Approach to Guide Students Through The Enrollment Process Based on Academic Performance*. User Modeling and User - Adapted Interaction, 2011:21(1-2), 217.

KAJIAN DATA MINING PREDIKET KELULUSAN MAHASISWA TEKNIK INFORMATIKA UNIVERSITAS ISLAM RIAU DENGAN ALGORITMA C4.5

ORIGINALITY REPORT

22%

SIMILARITY INDEX

17%

INTERNET SOURCES

12%

PUBLICATIONS

12%

STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

2%

★ Nugroho Dwi Saputro, Setyoningsih Wibowo,
Mochamad Fadjar Darmaputra. "The Application of
K-Nearest Neighbours Algorithms for the
Classification of Fashion Trend", JUSS (Jurnal Sains
dan Sistem Informasi), 2019

Publication

Exclude quotes On

Exclude matches < 1%

Exclude bibliography On