JUMANJI
ISSN (p): 2598-8050
ISSN (e): 2598-8069
| Vol. 5 | No. 2 | Halaman 89 – 98
Oktober 2021

Implementasi Data Mining untuk Prediksi Kelulusan Tepat Waktu Mahasiswa STIMIK ESQ Menggunakan Decision Tree C4.5

MITA NURUL YATIMAH[1]

[1]Jurusan Ilmu Komputer,
STIMIK ESQ
Jl. Tb. Simatupang, Cilandak, Jakarta Selatan
m.nurul.y@students.esqbs.ac.id

ABSTRAK

Kelulusan mahasiswa merupakan isu penting yang sering dibahas di perguruan tinggi. Hal ini berkaitan dengan penilaian akreditasi. Kualitas perguruan tinggi sangat dipengaruhi oleh proses penilaian akreditasi. Salah satu komponen penilaian akreditasi adalah kelulusan mahasiswa. Semakin banyak lulusan tepat waktu (empat tahun) maka akan semakin baik pula penilaian terhadap perguruan tinggi tersebut. Untuk itu penelitian ini akan melakukan prediksi secara dini kepada mahasiswa untuk melihat potensi kelulusan tepat waktu. Pada penelitian ini akan dilakukan di kampus STIMIK ESQ yang merupakan salah satu perguruan tingi yang ada di Jakarta. Hasil dari penelitian ini akan sangat bermanfaat bagi civitas akademik terutama para dosen untuk memperoleh hasil analisis yang sifatnya objektif, cepat dan terotomatisasi. Selain itu, hasil analisis ini dapat menjadi suatu informasi pendukung dalam memberikan penanganan-penanganan khusus terhadap mahasiswa. Penelitian ini menggunakan Decision tree C4.5 dalam melakukan prediksi tingkat kelulusan. Hasil akurasi yang diberikan penelitian ini sebesar 90% dengan menggunakan parameter jenis kelamin, usia, prodi, IPS1, SKS1, IPK1, IPS2, SKS2, IPK2, IPS3, SKS3, IPK3, IPS4, SKS4, IPK4, dan masa studi.

***Kata kunci**: Prediksi, Kelulusan, Decision tree C4.5, Data Mining*

ABSTRACT

Student graduation is an important issue that is often discussed in universities. This is related to the accreditation assessment. The quality of higher education is strongly influenced by the accreditation assessment process. One component of the accreditation assessment is student graduation. The more graduates on time (four years) the better the assessment of the college will be. For this reason, this study will make early predictions for students to see the potential for graduation on time. This research will be conducted at the STIMIK ESQ campus which is one of the universities in Jakarta. The results of this study will be very useful for the academic community, especially lecturers to obtain analytical results that are objective, fast and automated. In addition, the results of this analysis can be used as supporting information in providing special treatments for students. This study uses Decision tree C4.5 in predicting the graduation rate. The accuracy results

given by this study are 90% using the parameters of gender, age, study program, IPS1, SKS1, GPA1, IPS2, SKS2, GPA2, IPS3, SKS3, GPA3, IPS4, SKS4, GPA4, and the study period.

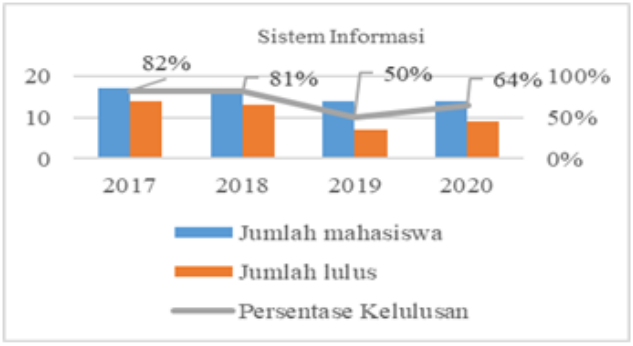
***Keywords**: Prediction, Graduation, Decision tree C4.5, Data Mining.*

1. PENDAHULUAN

Perkembangan dunia pendidikan di Indonesia telah memberikan dampak persaingan yang sangat ketat. Hal ini dipicu akibat semakin majunya pendidikan di perguruan tinggi. Salah satu dampak dari persaingan yaitu menghasilkan lulusan yang berkualitas. Adapun kriteria lulusan yang berkualitas diantaranya mampu menyelesaikan masa pembelajaran tepat waktu.

Masa pembelajaran tepat waktu sangat mempengaruhi kualitas dari perguruan tinggi. Kemampuan perguruan tinggi menghasilkan lulusan yang mampu menyelesaikan masa pembelajaran tepat waktu merupakan faktor yang mempengaruhi akreditasi perguruan tinggi. Hal ini sesuai dengan peraturan Badan Akreditasi Nasional Perguruan Tinggi Nomor 3 tahun 2019 tentang Instrumen Akreditasi Perguruan Tinggi yang menyatakan bahwa salah satu indikator penilaian akreditasi adalah persentase lulusan tepat waktu untuk setiap program dari perguruan tinggi. Untuk itu sangat penting mencari faktor yang mempengaruhi kelulusan tepat waktu di perguruan tinggi.

Studi kasus pada penelitian ini adalah STIMIK (Sekolah Tinggi Ilmu Manajemen Ilmu Komputer) ESQ yang berada di Jakarta Selatan. Adapun yang menjadi objek pada penelitian ini merupakan mahasiswa dari prodi manajemen dan sistem informasi. Gambar 1 di bawah merupakan grafik kelulusan mahasiswa program studi informasi tahun 2017 sampai dengan tahun 2020.



Gambar 1. Persentase Kelulusan Program Studi Sistem Informasi

Gambar 2 di bawah merupakan grafik kelulusan mahasiswa program studi manajemen tahun 2017 sampai dengan 2020.

Gambar 2. Persentase Kelulusan Program Studi Manajemen

Merujuk pada Gambar 1 dan Gambar 2 terlihat bahwa adanya pola kelulusan yang mengalamai penurunan ataupun kenaikan yang sangat signifikan. Hal ini terlihat pada program studi sistem informasi pada tahun 2018-2019 yang mengalami penurunan drastis sebanyak 31%. Hal ini juga terjadi pada program studi manajemen pada tahun 2017-2018 yang menunjukan bahwa adanya kenaikan yang signifikan yaitu sebanyak 26%. Namun selanjutnya mengalami penurunan terus menerus hingga tahun 2020. Dengan demikian penelitian ini akan melakukan analisis terhadap faktor-faktor yang mempengaruhi kelulusan tepat waktu dengan menerapkan data mining untuk proses pengoalhan data secara otomatis.

Analisis dan prediksi diharapkan mampu menemukan faktor-faktor yang mempengaruhi dan memperediksi kelulusan tepat waktu. Sehingga dapat dilakukan prediksi kelulusan mahasiswa lebih dini. Manfaat lainnya yaitu untuk menunjang nilai akreditasi serta sistem yang lebih terintegrasi. Data yang digunakan pada penelitian ini merupakan data kelulusan mahasisswa tahun 2017-2020 pada prodi sistem informasi dan manajemen bisnis. Peneliti fokus pada faktor-faktor yang mempengaruhi kelulusan mahasiswa dari semester 1- 4 agar prediksi dapat dilakukan lebih dini.

Penelitian tehadap kelulusan mahasiswa telah banayak dilakukan salah satunya yaitu "Prediksi Kelulusan Mahasiswa Diploma dengan Komparasi Algoritma Klasifikasi". Penelitian ini mengkomparasi 5 algoritma data mining yaitu Decision tree C4.5, Naive Bayes, K-NN, rule Induction, dan random forest dengan tujuan untuk menentukan metode yang paling akurat.

2. METODOLOGI

Langkah untuk melakukan prediksi kelulusan mahasiswa adalah sebagai berikut.

2.1. Seleksi Data

Pada umumnya data yang diperoleh dari database memiliki isian yang tidak sesuai atau tidak sempurna seperti missing value, data yang tidak valid atau hanya sekedar salah penulisan. Maka pada tahap ini dilakukan pembersihan data dari missing value serta data yang tidak relevan. Pembersihan data perlu dilakukan agar data yang diolah memiliki performa yang baik.

Dari total 186 data didapatkan 86 data yang bersih dari missing value. pembersihan data dilakukan secara manual dengan mengeliminasi data yang memiliki missing value dan data yang tidak relevan. Dari 86 data tersebut akan diambil sebagaian data sebagai sampel, proses pengambilan sampel dilakukan dengan menggunakan metode stratified random sampling sedangkan penentuan jumlah sampel mengacu pada tabel krejcie.

2.2. Preprocessing

Preprocesing bertujuan untuk memilih variabel yang mempengaruhi kelulusan tepat waktu. analisis dilakukan dengan menggunakan bantuan aplikasi SPSS dengan teknik Regresi Logistik biner. Regresi Logistik Biner merupakan suatu metode analisis statistik yang berguna untuk menganalisis hubungan antar suatu variabel respon dengan beberapa prediktor.

Adapun variabel yang diuji dengan menggunakan metode regresi logistik biner adalah Jenis Kelamin, Usia saat masuk yang diketahui dari pengurangan tahun masuk dengan tahun lahir, Prodi, Indeks prestasi semester satu sampai dengan semester 4, Satuan Kredit semester 1 sampai dengan semester 4, Indeks prestasi kumulatif semester 1 sampai dengan semester 4 dan Masa studi sebagai label.

Enam belas variabel tersebut diuji untuk menentukan atribut yang paling berpengaruh terhadap kelulusan tepat waktu, pengujian didasarkan pada nilai signifikasi, atribut dikatakan berpengaruh jika memiliki nilai signifikasi <=0,05. Hasil pengujian menunjukan bahwa nilai sig >= 0,05 yang artinya secara parsial variabel bebas tidak mempengaruhi model. Dari ke-15 variabel tersebut yang mendekati nilai <=0.05 adalah variabel SKS4, SKS3, IPK1 dan seterusnya. Hal tersebut menunjukan bahwa SKS4 memiliki nilai yang paling berpengaruh terhadap kelulusan mahasiswa. Pengujian tersebut juga menunjukan bahwa model yag dibuat sudah fit.

2.3. Requirements Definition

Data dari setiap atribut ditransformasikan ke dalam bentuk kategori agar data dapat diproses dan berada di rentang nilai yang sama, kemudian dinotasikan dengan rentang angka 0-4. Adapun transformasi yang dilakukan adalah sebagai berikut:

1) Jenis Kelamin, atribut ini memiliki dua nilai yaitu laki-laki dan perempuan, dimana laki-laki dinotasikan dengan 0 dan perempuan dinotasikan dengan 1.

2) Prodi, atribut ini memiliki dua nilai yaitu sistem informasi dan manajemen, dimana manajemen dinotasikan dengan 1 dan sistem informasi dengan 0.

3) Usia, atribut usia ditransformasikan pada Tabel 1 berikut.

Tabel 1. Transformasi Variabel Usia

No	Usia	Transformasi
1	<=15	3
2	16-18	2
3	19-21	1

4) Indeks Prestasi Semester dan Indeks Prestasi Kumulatif, atribut ini memiliki rentang nilai dari 0.00 sampai dengan 4.00. yang ditransformasikan pada Tabel 2.

Tabel 2. Transformasi Variabel Indeks Prestasi

No	Indeks Prestasi	Notasi
1	≥3.51	0
2	3.01-3.50	1
3	2.76-3.00	2

5) Satuan Kredit Semester (SKS), atribut ini memiliki rentang nilai antara 0 sampai dengan 24 yang ditransformasikan pada Tabel 3.

Tabel 3. Transformasi Variabel SKS

No	Indeks Prestasi	Notasi
1	22-24	0
2	19-21	1
3	16-18	2

6) Masa Studi, atribut ini memiliki dua nilai yaitu "Tepat" dan terlambat, dimana tepat dinotasikan dengan 1 dan terlambat dengan 0.

2.4. Proses Data Mining

Berdasarkan data dan atribut yang telah didapat dari proses sebelumnya, proses selanjutnya adalah melakukan pengolahan data meggunakan metode Decision tree C4.5. Decision Tree merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon yang berisi alternatif-alternatif pemecahan dari suatu permasalahan (Hamidah et al., 2019). Pengertian tersebut sejalan dengan pengertian Decision Tree menurut Kamal dan kawan-kawan (Kamal et al., 2017) yang menyatakan bahwa Decision Tree merupakan salah satu metode klasifikasi yang direpresentasikan dengan struktur pohon (Tree) atribut direpresentasikan oleh node, nilai dari atribut direpresentasikan oleh cabang, kelas direpresentasikan oleh daun.

Node akar (root) merupakan level teratas yang biasanya berupa atribut yang memiliki pengaruh paling besar pada suatu kelas tertentu, konsep dari Decision Tree adalah mengubah data menjadi suatu model pohon keputusan yang kemudian diubah menjadi sebuah rule (Setio et al., 2020).

3. HASIL PENELITIAN DAN ANALISIS

Populasi yang digunakan pada penelitian ini sebanyak 86 data sedangkan sampel yang digunakan sebanyak 70 data yang terdiri dari 35 data mahasiswa lulus tepat waktu dan 35

data mahasiswa lulus tidak tepat waktu atau terlambat. Penentuan jumlah sampel didasarkan pada tabel Krejcie Morgan.

Variabel yang digunaan pada penelitian ini sebanyak 16 data yang terdiri dari variabel bebas yaitu Usia, Jenis Kelamin, Prodi, SKS 1-4, IPS 1-4, dan IPK 1-4. Sedagkan variabel terikatnya adalah variabel masa studi. Untuk menguji apakah variabel bebas secara bersama-sama mempengaruhi model atau tidak, variabel bebas secara sendiri-sendiri (parsial) mempengaruhi model atau tidak dan menentukan model yang dibuat apakah fit atau tidak, dilakukan pengujian dengan menggunakan metode regresi logistik biner. Pengujian yang didapatkan pada Tabel 4, Tabel 5 dan Tabel 6 berikut.

1) Tabel Omnibus Test.

Tabel 4. Omnibus Test

No	Step	df	Sig.	Chi-square
1	Step	55,209	14	,000
2	Block	55,209	14	,000
3	Model	55,209	14	,000

Dapat dilihat bahwa nilai sig <0,05 yang artinya secara bersama-sama variabel bebas terbukti mempengaruhi model.

2) Variabel in the equation

Tabel 5. Variabel in the equation

No	Variabel	Sig.
1	JK(1)	,653
2	USIA	,369
3	PRODI(1)	,570
4	IPS1	,895
5	IPK1	,033
6	IPS2	,630
7	SKS2	,754
8	IPK2	,898
9	IPS3	,427
10	SKS3	,027
11	IPK3	,818
12	IPS4	,532
13	SKS4	,008
14	IPK4	,269
15	Constant	,069

Dapat dilihat nilai sig >= 0,05 yang artinya secara parsial variabel bebas tidak mempengaruhi model. Dari ke-15 variabel tersebut yang mendekati nilai <=0,05 adalah variabel SKS4, SKS3, IPK1, dan seterusnya. Hal tersebut menunjukan bahwa SKS4 memiliki nilai yang paling berpengaruh terhadap kelulusan mahasiswa.

3) Tabel Hosmer and Lomeshow Test

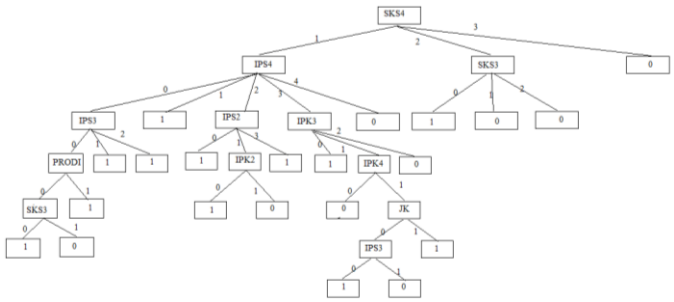
Tabel 6. Hosmer and Lomeshow Test

No	Step	df	Sig.	Chi-square
1	1	5,266	8	,729

Pada Tabel 4.8 terlihat bahwa nilai sig > 0.05 yang artinya model yang dibuat terbukti fit.

Berdasarkan data dan atribut yang telah didapat dari proses sebelumnya, proses selanjutnya adalah melakukan pengolahan data dengan metode *Decision tree*. Untuk melakukan prediksi dibutuhkan data latih dan data uji. Data latih berfungsi sebagai data pembelajaran dan data uji berfungsi untuk menguji model yang dihasilkan dari data latih. Pada penelitian ini data latih yang digunakan sebanyak 70% dan data ujinya sebanyak 20% atau setara dengan 21 data.

Pengujian data kelulusan mahasiswa untuk memprediksi kelulusan mahasiswa tepat waktu dengan metode *Decision tree* dilakukan dengan menggunakan bahasa pemrograman python pada *tools Google Colab*. Gambar 3 merupakan gambar pohon keputusan prediksi kelulusan tepat waktu. Keterangan pada Gambar 3 dimuat pada Tabel 7.



Gambar 2. Persentase Kelulusan Program Studi Manajemen

Tabel 7. Keterangan Gambar 3

No	Variabel	Keterangan
1	JK (Jenis Kelamin)	0 (Laki-laki) dan 1 (Perempuan)
2	Usia	0 (>=21), 1 (19-21), 2 (16-21), dan 3 (<=15)
3	Prodi	0 (Sistem informasi) dan 1 (Manajemen)
4	IPK dan IPS (1-4)	0 (>=3.51), 1 (3.01-3.50), 2 (2.76-3.00), 3 (2.00-2.75), dan 4 (<=1.99)
5	SKS (1-4)	0 (22-24), 1 (19-21), 2 (16-18), dan 3 (<=15)
6	Masa Studi	0 (Terlambat) dan 1 (Tepat)

Berdasarkan pemodelan yang dilakukan dengan *Decision tree* C4.5 dengan data sebanyak 70 data. Dimana data uji yang digunakan sebanyak 30% atau setara dengan 21 data dan data latih sebanyak 70% atau setaa dengan 49 data. Model tersebut menghasilkan akurasi sebesar 90% dengan SKS4 sebagai root atau sebagai atribut yang paling berpengaruh. Adapun interpretasi dari model yang telah dibuat disajikan pada Tabel 8.

Tabel 8. Interpretasi Model

No	**Program Studi**	**Rules**
1	Sistem Informasi	Jika SKS4 sebanyak 19-21 SKS, IPS4 lebih besar atau sama dengan dari 3.51, IPS3 lebih besar atau sama dengan 3.51, Prodi Sistem Informasi, SKS3 sebanyak 22-24 maka diprdiksi "lulus tepat waktu"
2	Manajemen	Jika SKS4 sebanyak 19-21 SKS, IPS4 lebih besar atau sama dengan dari 3.51, IPS3 lebih besar atau sama dengan 3.51, Prodi Sistem Informasi, SKS3 sebanyak 19-21 maka diprediksi "tidak lulus tepat waktu atau terlambat"
3	Sistem Informasi dan Manajemen	Jika SKS4 sebanyak 19-21 SKS, IPS4 lebih besar atau sama dengan dari 3.51, IPS3 lebih besar atau sama dengan 3.51, Prodi Manajemen maka diprediksi "lulus tepat waktu" Jika SKS4 sebanyak 19-21 SKS, IPS4 lebih besar atau sama dengan dari 3.51, IPS3 berada diantara rentang nilai 3.01-3.50 maka diprediksi "lulus tepat waktu" Jika SKS4 sebanyak 19-21 SKS, IPS4 lebih besar atau sama dengan dari 3.51, IPS3 berada diantara rentang nilai 2.76-3.00 maka diprediksi "lulus tepat waktu" Jika SKS4 sebanyak 19-21 SKS, IPS4 berada pada rentang nilai 3.01-3.50 maka diprediksi "lulus tepat waktu" Jika SKS4 sebanyak 19-21 SKS, IPS4 berada pada rentang nilai 2.76-3.00, dan IPS2 lebih dari atau sama dengan 3.51 maka diprediksi "lulus tepat waktu" Jika SKS4 sebanyak 19-21 SKS, IPS4 berada pada rentang nilai 2.76-3.00, dan IPS2 lebih dari atau sama dengan 3.51, IPK2 lebih dari 3.51 maka diprediksi "lulus tepat waktu" Jika SKS4 sebanyak 19-21 SKS, IPS4 berada pada rentang nilai 2.76-3.00, dan IPS2 lebih dari atau sama dengan 3.51, IPK2 berada pada rentang nilai 3.01-3.50 21 maka diprediksi "tidak lulus tepat waktu atau terlambat" Jika SKS4 sebanyak 19-21 SKS, IPS4 berada pada rentang nilai 2.76-3.00, dan IPS2 berada pada rentang niali 2.00-2.75 maka diprediksi "lulus tepat waktu" Jika SKS4 sebanyak 19-21 SKS, IPS4 berada pada rentang nilai 2.00-2.75, IPK3 lebih dari atau sama dengan 3.51 maka diprediksi "lulus tepat waktu" Jika SKS4 sebanyak 19-21 SKS, IPS4 berada pada rentang nilai 2.00-2.75, IPK3 berada pada rentang nilai 3.01-3.50 ,IPK4 lebih dari atau sama dengan 3.51 maka diprediksi "lulus tepat waktu" Jika SKS4 sebanyak 19-21 SKS, IPS4 berada pada rentang nilai 2.00-2.75, IPK3 berada pada rentang nilai 3.01-3.50, IPK4 berada pada rentang nilai 3.01-3.50, JK (Jenis Kelamin) laki-laki, IPS3 lebih dari 3.51 maka diprediksi "lulus tepat waktu" Jika SKS4 sebanyak 19-21 SKS, IPS4 berada pada rentang nilai 2.00-2.75, IPK3 berada pada rentang nilai 3.01-3.50 , IPK4 berada pada rentang nilai 3.01-3.50, JK (Jenis Kelamin) laki-laki, IPS3 berada pada rentang nilai 3.01-3.51 maka diprediksi "tidak lulus tepat waktu atau terlambat" Jika SKS4 sebanyak 19-21 SKS, IPS4 berada pada rentang nilai 2.00-2.75, IPK3 berada pada rentang nilai 3.01-3.50 , IPK4 berada pada rentang nilai 3.01-3.50, JK (Jenis Kelamin) maka diprediksi "lulus tepat waktu" Jika SKS4 sebanyak 19-21 SKS, IPS4 berada pada rentang nilai 2.00-2.75, IPK3 berada pada rentang nilai 2.76-3.00 maka diprediksi "lulus tepat waktu"

No	Program Studi	Rules
		Jika SKS4 sebanyak 19-21 SKS, IPS4 lebih kecil dari atau sama denga 1.99 maka diprediksi "tidak lulus tepat waktu atau terlambat" Jika SKS4 sebanyak 19-21 SKS, SKS3 berada pada rentang 22-24 SKS maka diprediksi "lulus tepat waktu" Jika SKS4 sebanyak 19-21 SKS, SKS3 berada pada rentang 19-21 "tidak lulus tepat waktu atau terlambat" Jika SKS4 sebanyak 19-21 SKS, SKS3 berada pada rentang 16-18 maka diprediksi "tidak lulus tepat waktu atau terlambat" Jika SKS4 kurang atau sama dengan 15 maka diprediksi "tidak lulus tepat waktu atau terlambat"

4. KESIMPULAN

Berdasarkan seluruh hasil dari tahapan penelitian yang telah dilakukan pada penerapan metode Decision tree C4.5 untuk prediksi kelulusan mahasiswa STIMIK ESQ dapat ditarik kesimpulan sebagai berikut:

1) Faktor yang mempengaruhi kelulusan tepat waktu di STIMIK ESQ berdasarkan pohon keputusan decision tree C4.5 yaitu SKS4, IPS4, SKS3, IPS3, IPS2, IPK3, Prodi, IPK2, IPK4, dan Jenis Kelamin.

2) Prediksi kelulusan mahasiswa dapat dilakukan dengan menggunakan pendekatan data mining dengan metode decision tree C4.5 yang menghasilkan 21 rules dengan tingkat akurasi sebesar 90.

DAFTAR RUJUKAN

Rujukan Jurnal:

Maulana, M., Sabarudin, R., & Nugraha, W. (2019). Prediksi Ketepatan Kelulusan Mahasiswa Diploma dengan Komparasi Algoritma Klasifikasi. *Jurnal Sistem dan Teknologi Informasi*, *7*(3), 202.

Rahman, A., Sorikhi., & Wartulas, S. (2020). Prediksi kelulusan mahasiswa menggunakan algoritma c4.5 (studi kasus di universitas peradaban). *Indonesian Journal of Informatics and Research, 1*(2), 70-77.

Hamidah, M., Fitriyah, H., & Arwani, I. (2019). Implementasi Decision Tree pada Penentuan Kondisi Ruang Berasap Menggunakan Multi-Sensor Berbasis Arduino Uno. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer, 3*(4), 3845-3854

Rujukan Prosiding:

Kamal, I., Hendro, T., & Ilyas, R. (2017). Prediksi Penjualan Buku Menggunakan Data Mining Di Pt. Niaga Swadaya. *Seminar Nasional Teknologi Informasi dan Multimed*ia. 2, 49-54.

Setio, P. B., Saputro, A., & Winarno, B. (2020). Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5. *Prisma : Prosiding Seminar Nasional Matematika.* 3, 64-71.