

ALGORITMA J48 DAN LOGISTIC MODEL TREE UNTUK MEMPREDIKSI PREDIKAT KELULUSAN MAHASISWA: STUDI KASUS STT XYZ

J48 ALGORITHM AND LOGISTIC MODEL TREE FOR PREDICTING STUDENT GRADUATION PREDICATE: CASE STUDY STT XYZ

Meriska Defriani¹, Irsan Jaelani²

¹²Sekolah Tinggi Teknologi Wastukencana Purwakarta
meriska@stt-wastukencana.ac.id

ABSTRACT

One of the things assessed in the accreditation process is the average cumulative achievement index (GPA) of graduates in the last three years. Higher School of Technology XYZ of Technology has been accredited by BAN-PT, but the accreditation value obtained is not optimal. One of the factors that influence is the average GPA of graduates which are still low. Therefore, a data mining analysis was carried out with a classification model on the academic and non-academic data of students. The classification algorithm used is J48 and Logistic Model Tree (LMT). The purpose of this study is to compare the J48 and LMT algorithms in predicting the final GPA of students through the graduation predicate and find out what factors influence the final GPA of students. The data used is the data of undergraduate students who have been declared graduated in the 2015-2019 year of the year totaling 1050 data. The results of this study are the J48 algorithm produces an accuracy of 75.5%, while the LMT algorithm produces an accuracy of 73.5%. In both algorithms, the attributes that influence the predicate of student graduation are IPS of semester 2, IPS of semester 1, and IPS of semester 4. Algorithm J48 is able to better classify student graduation and do a simpler classification process which also requires shorter time.

Keywords: *Graduation Predicate, Classification, J48, LMT*

ABSTRAK

Salah satu hal yang dinilai dalam proses akreditasi adalah rata-rata Indeks Prestasi Kumulatif (IPK) lulusan dalam tiga tahun terakhir. Sekolah Tinggi Teknologi (STT) XYZ telah terakreditasi BAN-PT namun nilai akreditasi yang didapatkan belum maksimal. Salah satu hal yang mempengaruhi adalah rata-rata IPK lulusan yang masih rendah. Oleh karena itu, dilakukan analisis data mining dengan model klasifikasi pada data akademik dan non akademik mahasiswa. Algoritma klasifikasi yang digunakan adalah J48 dan Logistic Model Tree (LMT). Tujuan dari penelitian ini adalah untuk membandingkan algoritma J48 dan LMT dalam memprediksi IPK akhir mahasiswa melalui predikat kelulusan dan mengetahui faktor apa saja yang mempengaruhi IPK akhir mahasiswa. Data yang digunakan adalah data mahasiswa jenjang sarjana yang telah dinyatakan lulus pada yusidium tahun 2015 sampai 2019 sejumlah 1050 data. Hasil penelitian ini adalah algoritma algoritma J48 menghasilkan akurasi sebesar 75,5%, sedangkan algoritma LMT menghasilkan akurasi sebesar 73,5%. Pada kedua algoritma, atribut yang berpengaruh pada predikat kelulusan mahasiswa adalah IPS semester 2, IPS semester 1, dan IPS semester 4. Algoritma J48 mampu mengklasifikasikan predikat kelulusan mahasiswa dengan lebih baik dan melakukan proses klasifikasi dengan lebih sederhana juga membutuhkan waktu yang lebih singkat.

Kata kunci: *Predikat Kelulusan, Klasifikasi, J48, LMT*

PENDAHULUAN

Rata-rata Indeks Prestasi Kumulatif (IPK) mahasiswa merupakan salah satu hal yang menjadi penilaian dalam proses akreditasi, baik akreditasi institusi maupun program studi. Dalam matriks penilaian akreditasi program sarjana pada elemen luaran dan capaian tridarma indikator kinerja utama, salah satu hal yang dinilai adalah rata-rata IPK lulusan dalam tiga

tahun terakhir. Skor terbaik yang diberikan adalah jika rata-rata nilai IPK lebih atau sama dengan 3,25.

Sekolah Tinggi Teknologi (STT) XYZ merupakan perguruan tinggi swasta yang terletak di Kabupaten Purwakarta. STT XYZ memiliki dua jenjang studi, yaitu sarjana dan diploma. Program studi yang termasuk dalam jenjang sarjana, yaitu Teknik Industri, Teknik Mesin, dan Teknik

Informatika, sementara program studi yang termasuk dalam jenjang diploma adalah Teknik Tekstil dan Manajemen Industri. STT XYZ telah berstatus terakreditasi BAN-PT namun nilai akreditasi tersebut belum maksimal.

Berdasarkan hal tersebut, perlu dilakukan suatu analisa terhadap data-data mahasiswa yang tersedia, baik data yang berhubungan dengan akademik maupun non-akademik untuk mengetahui faktor apa saja yang mempengaruhi IPK akhir mahasiswa dan juga untuk memprediksi IPK akhir mahasiswa yang saat ini sedang aktif. IPK akhir mahasiswa dapat direpresentasikan dengan predikat kelulusan yang diberikan kepada mahasiswa yang telah dinyatakan lulus.

Analisa yang dapat dilakukan untuk mencapai tujuan tersebut adalah data mining. Data mining merupakan suatu teknik yang dapat digunakan untuk menemukan suatu informasi dari sekumpulan data yang besar. Hasil analisa tersebut berupa pola atau aturan yang dapat menghasilkan suatu pengetahuan dimana pengetahuan tersebut dapat digunakan untuk mendukung pengambilan suatu keputusan atau kebijakan. Terdapat tiga model data mining, yaitu klasifikasi, pengelompokan, dan asosiasi. Model yang dapat digunakan dalam penelitian ini adalah model klasifikasi. Klasifikasi adalah suatu proses menemukan sebuah fungsi atau model untuk membedakan data ke dalam kelas-kelas yang sebelumnya tidak diketahui.

Terdapat beberapa algoritma yang dapat digunakan dalam model klasifikasi, diantaranya adalah J48 dan *Logistic Model Tree* (LMT). Algoritma J48 merupakan implementasi dari algoritma C4.5 di dalam aplikasi WEKA. Algoritma ini menghasilkan pohon keputusan yang dapat digunakan untuk mengklasifikasikan suatu data ke dalam kelas-kelas dan juga memprediksi kelas dari suatu data (Agustiani, 2020). LMT merupakan algoritma data mining yang menyatukan dua algoritma klasifikasi yaitu *Decision*

Tree (DT) dan *Logistic Regression* (LR) (Chen, dkk., 2017).

Algoritma J48 sebelumnya digunakan untuk menganalisis Tingkat Kemiskinan di Indonesia. Algoritma J48 mampu menghasilkan model data mining dengan tingkat akurasi sebesar 88,6% (Kaunang, 2019). Sebuah penelitian di tahun 2018 melakukan analisis komparasi beberapa algoritma klasifikasi, yaitu algoritma *Simple Logistic*, *Instance-base k-nearest Neighbors* (IBK), *Naïve Bayes*, *Stochastic Gradient Descent* (SGD), *Logistic Model Tree* (LMT) dan *Sequential Minimal Optimization* (SMO). Hasil dari penelitian tersebut menyatakan bahwa LMT memiliki akurasi yang lebih baik dibandingkan dengan algoritma klasifikasi yang lain (Alaoui & Dkk, 2018). Pada tahun 2020, algoritma J48 dan LMT untuk memprediksi lama studi mahasiswa di Program Studi Teknik Informatika STT XYZ. Hasil penelitian tersebut, algoritma LMT memiliki akurasi yang lebih baik dibandingkan J48 (Maulana & Defriani, 2020).

Pada penelitian ini dilakukan analisa data mining untuk memprediksi IPK akhir melalui predikat kelulusan mahasiswa dengan menggunakan algoritma J48 dan LMT. Selain itu, analisa ini dilakukan juga untuk mengetahui faktor apa saja yang mempengaruhi IPK mahasiswa STT XYZ. Analisa dilakukan hanya pada jenjang sarjana yang terdiri dari tiga program studi, yaitu Teknik Industri, Teknik Mesin, dan Teknik Informatika. Harapannya dengan dilakukan analisa ini, IPK akhir mahasiswa aktif yang sedang duduk pada tahun ketiga perkuliahan dapat diprediksi sehingga hasil prediksi tersebut dapat digunakan oleh pihak institusi untuk membuat kebijakan terkait pelayanan institusi dan proses belajar mengajar yang dapat meningkatkan nilai mahasiswa dan juga mendukung peningkatan nilai akreditasi program studi maupun institusi.

METODE

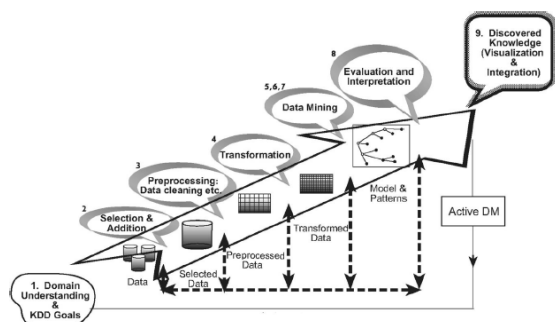
Metode yang digunakan dalam penelitian ini adalah *Knowledge Discovery in Database* (KDD) dengan algoritma yang digunakan adalah algoritma J48 dan *Logistic Model Tree* (LMT). Berikut ini adalah penjelasan dari metode dan algoritma yang digunakan:

Knowledge Discovery in Database (KDD)

Proses KDD terdiri dari sembilan tahapan seperti yang dapat dilihat pada Gambar 1 (Rokach & Maimon, 2015). Berikut ini adalah penjelasan dari setiap tahapan dalam proses KDD:

1. Mengembangkan pemahaman tentang domain aplikasi

Pada tahapan ini dilakukan persiapan awal untuk memahami lebih dalam permasalahan yang terjadi kemudian melakukan pengambilan keputusan terkait transformasi, algoritma, representasi, dan hal yang lainnya. Pada tahapan ini juga dilakukan penentuan tujuan pengguna akhir dan lingkungan dimana proses penemuan pengetahuan akan berlangsung.



Sumber : (Rokach & Maimon, 2015)

Gambar 1. Proses KDD

2. Membuat dataset dimana proses penemuan pengetahuan akan dilakukan

Pada tahapan ini dilakukan pengumpulan dan pengintegrasian data ke dalam dataset. Tahapan ini merupakan proses yang sangat penting karena data mining mempelajari dan menemukan pola-pola baru dari data yang tersedia yang kemudian menjadi dasar untuk pembuatan model.

3. Persiapan Awal dan Pembersihan

Pada tahapan ini kualitas data akan ditingkatkan melalui pembersihan data seperti menangani nilai yang hilang dan

menghilangkan noise atau outlier. Proses persiapan awal dan pembersihan memungkinkan untuk melibatkan metode statistik yang kompleks atau menggunakan algoritma data mining yang spesifik.

4. Transformasi Data

Pada tahapan ini data disiapkan dan dikembangkan dengan lebih baik sehingga lebih siap untuk dilakukan pemodelan data mining. Hal yang dapat dilakukan dalam tahapan ini adalah reduksi dimensi seperti pemilihan fitur dan ekstraksi sampel data. Selain itu juga dapat dilakukan transformasi atribut seperti mengubah atribut numerik menjadi atribut diskrit dan transformasi fungsional.

5. Memilih Model Data Mining yang Sesuai

Pada tahapan ini dilakukan pemilihan model data mining yang sesuai dengan kebutuhan. Terdapat beberapa model data mining yang dapat dipilih, yaitu klasifikasi, regresi, atau pengelompokan. Terdapat dua tujuan utama dalam data mining, yaitu prediksi dan deskripsi. Prediksi sering kali diistilahkan dengan *supervised* data mining, sedangkan data mining deskriptif mencakup *unsupervised* data mining dan visualisasi data mining. Sebagian besar teknik data mining didasarkan pada pembelajaran induktif dimana sebuah model dibangun secara eksplisit atau implisit dengan melakukan generalisasi dari sejumlah contoh pelatihan yang memadai. Asumsi yang mendasari pendekatan induktif adalah bahwa model yang terlatih berlaku untuk kasus-kasus di masa depan.

6. Memilih Algoritma Data Mining

Pada tahapan ini dilakukan pemilihan metode spesifik yang akan digunakan untuk mencari pola. Contohnya, algoritma jaringan syarai dan pohon keputusan lebih baik digunakan jika yang menjadi pertimbangan adalah ketepatan dibandingkan dengan pemahaman.

7. Menggunakan metode data mining

Pada tahapan ini algoritma data mining dapat digunakan berulang kali hingga memperoleh hasil yang sesuai. Pada

tahapan ini juga dilakukan pengaturan parameter kontrol algoritma seperti jumlah minimum instansi dalam satu daun dari pohon keputusan.

8. Evaluasi

Pada tahapan ini dilakukan evaluasi dan penafsiran pola yang diekstraksi seperti aturan, keandalan, dan hal yang lainnya, sehubungan dengan tujuan yang telah ditentukan pada langkah awal. Tahapan ini berfokus pada kelengkapan dan kegunaan dari model yang diinduksi. Pada tahapan ini dilakukan pendokumentasian pengetahuan yang ditemukan untuk penggunaan lebih lanjut.

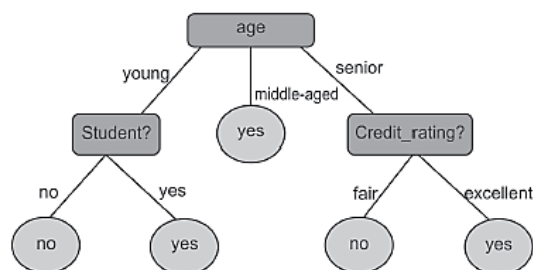
9. Menggunakan pengetahuan yang telah ditemukan

Pada tahapan ini, pengetahuan yang ditemukan telah siap untuk diintegrasikan dengan sistem lain untuk pemanfaatan lebih lanjut. Pengetahuan menjadi aktif dalam arti bahwa perubahan dapat dilakukan pada sistem dan dapat diukur efeknya. Bahkan, keberhasilan langkah ini menentukan keefektifan seluruh proses. Misalnya, sebuah pengetahuan ditemukan dari sampel data tapi kemudian data menjadi dinamis. Struktur data dapat berubah ketika atribut tertentu menjadi tidak tersedia dan domain data dapat dimodifikasi (Rokach & Maimon, 2015).

Decision Tree J48

Pada algoritma *Decision Tree* atau pohon keputusan, prediksi dibuat dengan beberapa kondisi jika-maka yang mirip dengan kontrol pada berbagai bahasa pemrograman. Struktur pohon keputusan hasil pemodelan data mining terdiri dari simpul akar, cabang, dan simpul daun. Setiap simpul cabang mewakili suatu kondisi pada beberapa atribut masukan, sementara masing-masing cabang menentukan hasil dari kondisi dan setiap simpul daun memiliki label kelas. Contoh pohon keputusan dapat dilihat pada Gambar 2. Gambar tersebut menjelaskan apakah seorang pelanggan akan membeli laptop atau tidak. Salah satu analisis yang dapat

diperoleh dari pohon keputusan tersebut adalah jika seorang pelanggan berusia muda dan seorang pelajar, maka ada kemungkinan akan membeli laptop (Bhatia, 2019).



Sumber : (Bhatia, 2019)

Gambar 2. Struktur Decision Tree

Logistic Model Tree (LMT)

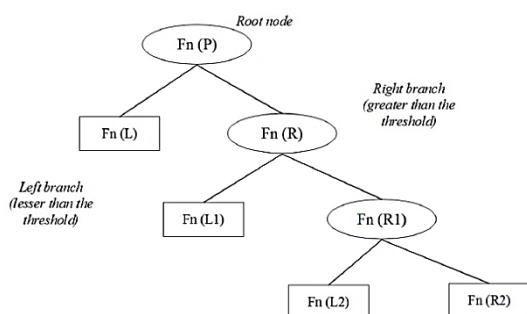
Logistic Model Tree (LMT) merupakan algoritma data mining yang menyatukan dua algoritma klasifikasi yaitu *Decision Tree* (DT) dan *Logistic Regression* (LR). Dalam praktiknya, LMT memberikan hasil yang lebih akurat dibandingkan dengan algoritma sejenisnya, seperti C4.5, CART dan LE (Colkesen & Kavzoglu, 2017).

Tidak seperti pohon keputusan biasa, node daun pada LMT memiliki fungsi *regresi logistik* (ft) dari atribut yang terkait, bukan hanya label kelas. Fungsi regresi (ft) memperhitungkan subset dari semua atribut yang ada dalam data (dengan asumsi bahwa atribut nominal telah dibinarisasi untuk tujuan regresi). Jika $F_j(x)$ adalah fungsi regresi yang dibangun pada setiap node, maka probabilitas keanggotaan kelas dimodelkan sebagai berikut:

$$Pr(G = j|X = x) = \frac{e^{F_j(x)}}{\sum_{k=1}^J e^{F_k(x)}}$$

Berdasarkan nilai ambang batas, simpul daun akan dibagi menjadi dua simpul anak di mana cabang ke kanan akan berisi nilai atribut yang lebih besar dari ambang batas dan cabang ke kiri akan berisi nilai atribut lebih rendah daripada ambang batas (Nachiappan et al., 2016). Struktur

dari *Logistic Model Tree* dapat dilihat pada Gambar 3.



Sumber : (Nachiappan et al., 2016)

Gambar 3 Struktur *Logistic Model Tree*

HASIL DAN PEMBAHASAN

Pada bagian ini dilakukan pembahasan mengenai proses pemodelan data mining mulai dari proses pemahaman data hingga penggunaan hasil pemodelan data mining. Metode penelitian yang digunakan dalam penelitian ini adalah metode KDD. Berikut ini adalah penjelasan masing-masing tahapan yang telah dilakukan dalam metode KDD:

Mengembangkan Pemahaman

Rata-rata Indeks Prestasi Kumulatif (IPK) mahasiswa merupakan salah satu hal yang menjadi penilaian dalam proses akreditasi, baik akreditasi institusi maupun program studi. STT XYZ telah berstatus terakreditasi BAN-PT namun nilai akreditasi tersebut belum maksimal. Oleh karena itu perlu dilakukan suatu analisa terhadap data-data mahasiswa yang tersedia, baik data yang berhubungan dengan akademik maupun non-akademik untuk mengetahui faktor apa saja yang mempengaruhi IPK akhir mahasiswa dan juga untuk mempresiksi IPK akhir mahasiswa yang saat ini sedang aktif. IPK akhir mahasiswa dapat direpresentasikan dengan predikat kelulusan yang diberikan kepada mahasiswa yang telah dinyatakan lulus. Analisa tersebut dapat dilakukan dengan data mining. Harapannya dengan dilakukan analisa ini, IPK akhir mahasiswa aktif yang sedang duduk pada tahun ketiga perkuliahan dapat diprediksi sehingga hasil prediksi tersebut dapat digunakan oleh

pihak institusi untuk membuat kebijakan terkait pelayanan institusi dan proses belajar mengajar yang dapat meningkatkan nilai mahasiswa dan juga mendukung peningkatan nilai akreditasi program studi maupun institusi.

Membuat Dataset

Data yang digunakan dalam penelitian ini adalah data mahasiswa jenjang sarjana, yaitu mahasiswa Teknik Industri, Teknik Mesin, dan Teknik Informatika yang telah dinyatakan lulus pada yudisium tahun 2015 sampai 2019 sejumlah 1050 data dan data mahasiswa yang masih aktif, yaitu mahasiswa jenjang S1 angkatan 2017 sejumlah 100 data. Data tersebut terdiri data akademik dan non akademik mahasiswa. Data-data tersebut tersimpan dalam beberapa dokumen yang terpisah, yaitu dokumen yudisium per periode, Kartu Hasil Studi (KHS) dan transkrip nilai mahasiswa yang didapat dari sistem informasi akademik (SIMAK), dan biodata mahasiswa yang terdapat pada sistem *feeder* STT XYZ.

Atribut yang digunakan dalam dokumen yudisium adalah NIM, Nama, dan predikat kelulusan mahasiswa. Atribut yang digunakan dalam dokumen KHS mahasiswa adalah NIM, Nama, dan Kelas. Atribut yang digunakan dalam transkrip nilai mahasiswa adalah NIM, Nama, Program Studi, dan IPS semester satu sampai empat. Atribut yang digunakan dalam dokumen biodata mahasiswa adalah *id_pd* yang merupakan NIM, *nm_pd* yang merupakan Nama Mahasiswa, *jk* yang merupakan jenis kelamin, *tgl_lahir*, dan *id_wilayah* yang merupakan alamat.

Data yang sudah melalui proses seleksi tersebut kemudian diintegrasikan berdasarkan NIM dan Nama untuk menjadi dataset. Namun, atribut Nama hanya menjadi ID atau identitas baris data, tidak disertakan ke dalam dataset. Dataset terdiri dari 1050 baris data dengan 11 atribut yang merupakan atribut prediktor. Penjelasan mengenai atribut-atribut tersebut dapat dilihat pada Tabel 1.

Tabel 1. Deskripsi Atribut Dataset

No.	Atribut	Tipe Atribut	Keterangan
1	NIM	Numerik	Nomor induk mahasiswa
2	Prodi	Nominal	Program studi mahasiswa
3	Jenis Kelamin	Nominal	Jenis kelamin mahasiswa
4	Tanggal Lahir	Numerik	Tanggal lahir mahasiswa
5	Alamat	Nominal	Alamat tinggal mahasiswa
6	Kelas	Nominal	Tipe kelas
7	IPS1	Numerik	Indeks prestasi semester 1
8	IPS2	Numerik	Indeks prestasi semester 2
9	IPS3	Numerik	Indeks prestasi semester 3
10	IPS4	Numerik	Indeks prestasi semester 4
11	Predikat	Nominal	Predikat kelulusan mahasiswa

Persiapan Awal dan Pembersihan

Data yang telah diintegrasikan dan diseleksi kemudian dibersihkan untuk mendapatkan *dataset* yang lebih handal. Adapun hal yang dilakukan dalam proses pembersihan adalah menghapus data mahasiswa transfer. Data mahasiswa transfer tidak dimasukkan karena mahasiswa transfer tidak menjalani pendidikan secara penuh di STT XYZ. Setelah proses pembersihan, jumlah set data yang tersisa sejumlah 962 baris data.

Transformasi Data

Dataset yang telah melalui proses pembersihan dan penghapusan pada tahap sebelumnya kemudian dilakukan transformasi data. Transformasi data dilakukan dengan membuat atribut baru yang diturunkan dari atribut yang lain dan mengubah tipe data atribut dari nominal ke numerik.

Pada proses transformasi dibuat sebuah atribut baru, yaitu usia yang diturunkan dari atribut tanggal lahir. Selain itu, dilakukan

pengubahan tipe data dari nominal ke numerik pada atribut Prodi, Jenis Kelamin, Alamat, dan Kelas. Atribut nominal diinisialisasi dengan angka sesuai dengan banyaknya nilai pada atribut tersebut. Inisialisasi nilai atribut Prodi, Jenis Kelamin, Alamat, dan Kelas dapat dilihat pada Tabel 2, Tabel 3, Tabel 4, dan Tabel 5. Contoh dataset hasil transformasi data dapat dilihat pada Tabel 6.

Tabel 2. Transformasi Atribut Program Studi

Nilai Atribut	Inisial
Teknik Informatika	1
Teknik Industri	2
Teknik Mesin	3

Tabel 3. Transformasi Atribut Jenis Kelamin

Nilai Atribut	Inisial
Laki-laki	1
Perempuan	2

Tabel 4. Transformasi Atribut Alamat

Nilai Atribut	Inisial
Purwakarta	1
Selain Purwakarta	2

Tabel 5. Transformasi Atribut Kelas

Nilai Atribut	Inisial
Pagi	1
Malam	2

Contoh dataset yang telah siap untuk pemodelan data mining dapat dilihat pada Tabel 6.

Memilih Model Data Mining

Model data mining yang digunakan pada penelitian ini adalah model klasifikasi. Adapun tujuan data miningnya adalah untuk prediksi.

Tabel 6. Contoh Dataset Hasil Transformasi Data

Prodi	Jenis Kelamin	Usia	Alamat	Kelas	IPS1	IPS2	IPS3	IPS4	Predikat
1	1	18	2	1	2,67	3,1	2,67	3,23	Sangat Memuaskan
1	1	18	1	1	3,19	3,1	2,91	3,21	Dengan Pujian
1	1	18	2	1	3,05	2,55	3,32	3,37	Sangat Memuaskan
1	1	20	2	2	2,52	2,9	3,3	3,11	Sangat Memuaskan
1	1	19	1	2	3,05	2,6	2,68	2,79	Sangat Memuaskan
1	1	21	2	2	2,76	3,5	2,73	2,68	Sangat Memuaskan
1	1	21	2	2	3,52	3,3	2,09	2,63	Memuaskan
1	2	21	1	2	3,38	3,4	3,5	3,42	Sangat Memuaskan

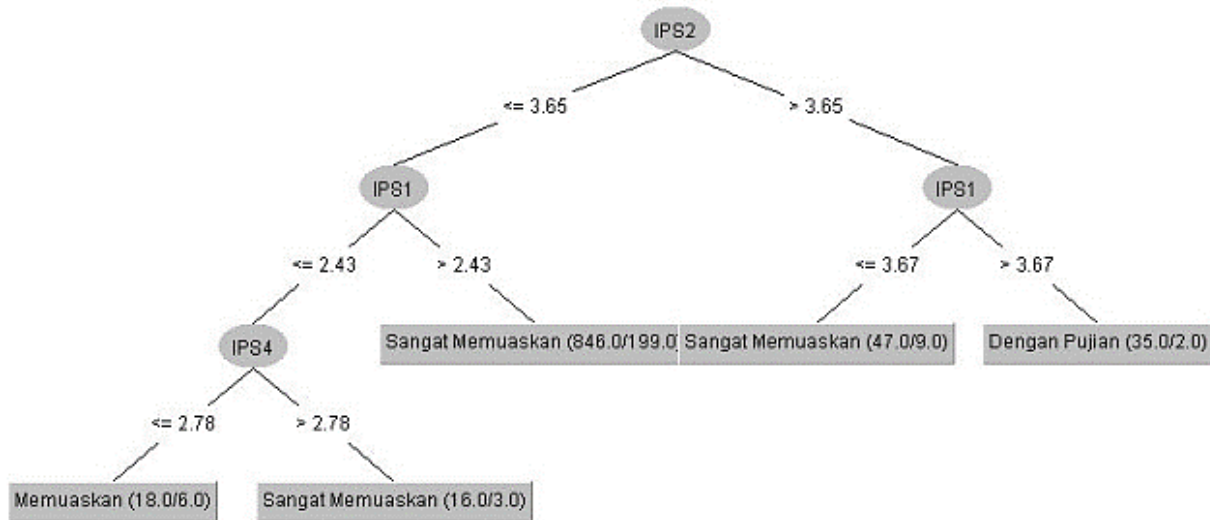
Prodi	Jenis Kelamin	Usia	Alamat	Kelas	IPS1	IPS2	IPS3	IPS4	Predikat
1	1	19	2	2	3,38	3,8	2,63	3,11	Sangat Memuaskan
1	1	19	1	2	2,76	3,3	2,92	3,09	Sangat Memuaskan
1	2	26	1	2	3,71	3,75	3,5	3,47	Dengan Pujian
1	2	17	1	1	3,76	3,6	3,64	3,76	Sangat Memuaskan
1	2	18	1	2	2,81	3,3	2,59	3,11	Sangat Memuaskan
1	1	18	2	1	2,9	2,85	2,36	2,84	Sangat Memuaskan

Memilih Algoritma Data Mining

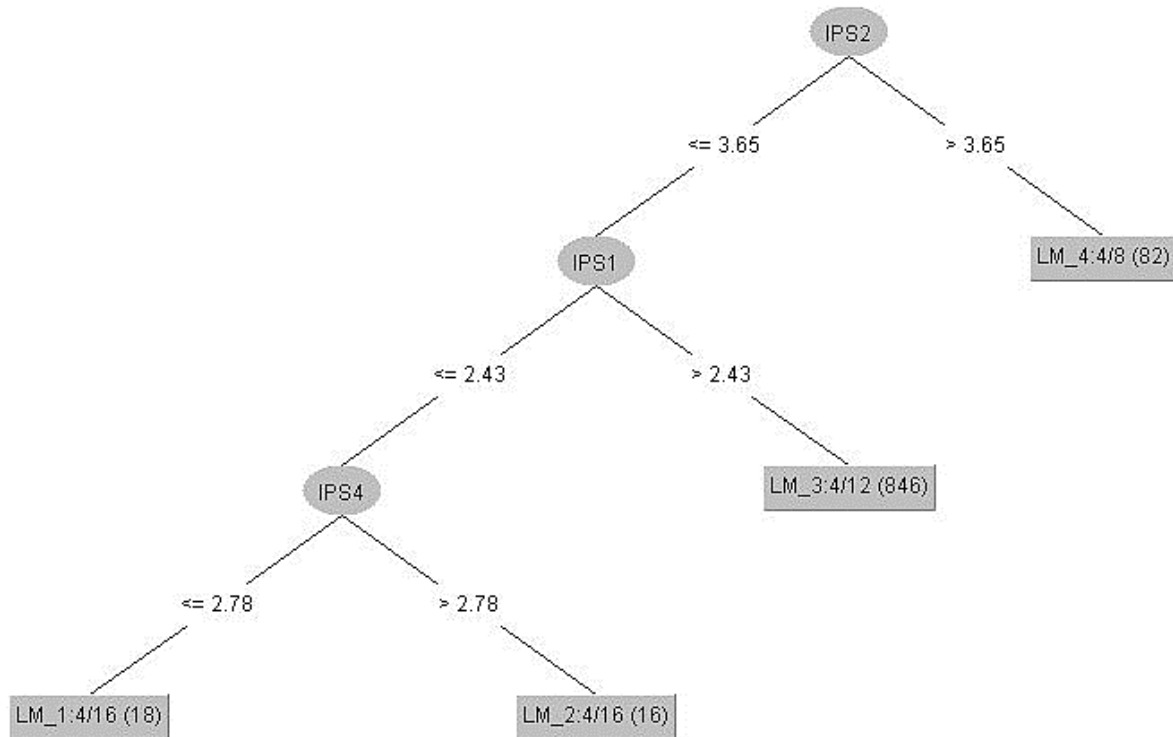
Algoritma data mining yang digunakan dalam penelitian ini adalah algoritma J48 dan *Logistic Model Tree* (LMT). Decision Tree J48 merupakan implementasi dari algoritma C4.5 pada aplikasi WEKA. LMT merupakan algoritma data mining yang menyatukan dua algoritma klasifikasi yaitu *Decision Tree* (DT) dan *Logistic Regression* (LR). Pohon keputusannya dibangun dengan menggunakan algoritma C4.5, sedangkan kelasnya ditentukan oleh regresi linear. LMT dinilai mampu memberikan hasil yang lebih akurat dibandingkan dengan algoritma klasifikasi sejenisnya, seperti C4.5, CART dan *Linear Regression*.

Menggunakan Metode Data Mining

Pemodelan data mining dilakukan dengan menggunakan Weka versi 3.9.3. Objek minimal per daun yang digunakan adalah 15 objek. Algoritma J48 dan LMT menghasilkan pohon keputusan yang dapat membantu untuk memprediksi predikat kelulusan mahasiswa. Pohon keputusan hasil pemodelan data mining dengan menggunakan algoritma J48 dapat dilihat pada Gambar 4. Gambar 5 merupakan hasil dengan menggunakan algoritma LMT. Adapun fungsi regresi setiap node daun pada pohon keputusan tersebut dapat dilihat pada Tabel 7.



Gambar 4. Pohon Keputusan Hasil Pemodelan Data Mining dengan Algoritma J48



Gambar 5. Pohon Keputusan Hasil Pemodelan Data Mining dengan Algoritma LMT

Tabel 7. Fungsi Regresi Setiap Node

Node	Regression Function
LM1	Sangat Memuaskan : $-14.46 + [\text{Prodi}] * -0.98 + [\text{Jenis Kelamin}] * 0.11 + [\text{Usia}] * 0.26 + [\text{Alamat}] * 1.37 + [\text{IPS1}] * -0.32 + [\text{IPS2}] * 0.5 + [\text{IPS4}] * 3.44$ Dengan Pujian : $4.48 + [\text{Prodi}] * 1.38 + [\text{Usia}] * -0.1 + [\text{Alamat}] * -0.86 + [\text{IPS1}] * 0.2 + [\text{IPS2}] * -0.32 + [\text{IPS3}] * -2.19 + [\text{IPS4}] * -0.08$ Memuaskan: $11.4 + [\text{Prodi}] * -0.32 + [\text{Jenis Kelamin}] * -0.19 + [\text{Usia}] * -0.06 + [\text{Alamat}] * -1.35 + [\text{Kelas}] * 0.09 + [\text{IPS1}] * -0.35 + [\text{IPS2}] * -1.41 + [\text{IPS4}] * -1.33$
LM2	Sangat Memuaskan : $21.75 + [\text{Prodi}] * -0.98 + [\text{Jenis Kelamin}] * 0.11 + [\text{Usia}] * 0.01 + [\text{Alamat}] * 0.13 + [\text{IPS1}] * -0.89 + [\text{IPS3}] * -3.54 + [\text{IPS4}] * -2.28$ Dengan Pujian : $-2.14 + [\text{Prodi}] * 0.82 + [\text{Usia}] * -0.1 + [\text{Alamat}] * -0.18 + [\text{IPS1}] * 0.8 + [\text{IPS2}] * -0.99 + [\text{IPS3}] * 0.93 + [\text{IPS4}] * -0.08$ Memuaskan : $-33.15 + [\text{Jenis Kelamin}] * -0.19 + [\text{Usia}] * -0.06 + [\text{Alamat}] * -1.35 + [\text{Kelas}] * 0.09 + [\text{IPS1}] * -0.35 + [\text{IPS2}] * -0.29 + [\text{IPS3}] * 2.13 + [\text{IPS4}] * 9.22$
LM3	Sangat Memuaskan : $-0.72 + [\text{Prodi}] * -0.04 + [\text{Jenis Kelamin}] * 0.16 + [\text{Usia}] * 0.01 +$

Node	Regression Function
	$[\text{Alamat}] * 0.07 + [\text{IPS1}] * 0.08 + [\text{IPS2}] * -0.13 + [\text{IPS4}] * 0.52$ Dengan Pujian : $-1.68 + [\text{Prodi}] * -0.02 + [\text{Jenis Kelamin}] * -0.04 + [\text{Usia}] * -0.01 + [\text{Alamat}] * -0.18 + [\text{Kelas}] * -0.05 + [\text{IPS1}] * 0.2 + [\text{IPS3}] * 0.55 + [\text{IPS4}] * -0.08$ Memuaskan : $1.36 + [\text{Jenis Kelamin}] * -0.19 + [\text{Usia}] * -0.08 + [\text{Alamat}] * 0.21 + [\text{Kelas}] * 0.09 + [\text{IPS1}] * -0.01 + [\text{IPS2}] * -0.29 + [\text{IPS4}] * -0.09$
LM4	Sangat Memuaskan : $15.52 + [\text{Prodi}] * -0.07 + [\text{Jenis Kelamin}] * 0.11 + [\text{Usia}] * 0.03 + [\text{IPS1}] * -3.48 + [\text{IPS2}] * -1.35 + [\text{IPS4}] * 0.64$ Dengan Pujian : $-15.69 + [\text{Prodi}] * -0.29 + [\text{IPS1}] * 4.51 + [\text{IPS3}] * 0.5 + [\text{IPS4}] * -0.33$ Memuaskan : $6.85 + [\text{Jenis Kelamin}] * -1.81 + [\text{Usia}] * -0.1 + [\text{Kelas}] * -0.74 + [\text{IPS1}] * -0.35 + [\text{IPS2}] * -0.29 + [\text{IPS4}] * -0.18$

Perlu dilakukan pengujian pada model data mining yang telah dibangun untuk mengetahui kehandalan model tersebut. Pada penelitian ini, model data mining diuji

dengan menggunakan model tes *K-fold cross-validation*. Nilai K yang digunakan adalah 5. Perbandingan ringkasan hasil pengujian dari algoritma J48 dengan dengan LMT dapat dilihat pada Tabel 8. Perbandingan detail akurasi per kelas dari hasil pengujian dapat dilihat pada Tabel 9. Perbandingan *confusion matrix* hasil pengujian dapat dilihat pada Tabel 10.

Tabel 8. Perbandingan Hasil Pengujian Hasil Pemodelan Data Mining dengan Algoritma J48 dengan LMT

Information	J48	LMT
Correctly classified instances	726 (75.4688%)	707 (73.4927%)

Information	J48	LMT
Incorrectly classified instances	236 (24.5322%)	255 (26.5073%)
Kappa statistic	0.1598	0.1426
Mean absolute error	0.2609	0.2617
Root mean squared error	0.3622	0.3709
Relative absolute error	91.9349%	92.2103%
Root relative squared error	96.2611%	98.56%
Total number of instances	962	962

Tabel 9. Perbandingan Detail Akurasi Per Kelas Hasil Pengujian Hasil Pemodelan Data Mining dengan Algoritma J48 dengan LMT

Algorithm	Informations								
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
J48	0,993	0,881	0,752	0,993	0,856	0,262	0,567	0,758	Sangat Memuaskan
	0,138	0,005	0,867	0,138	0,237	0,303	0,565	0,321	Dengan Pujian
	0,056	0,002	0,667	0,056	0,103	0,178	0,614	0,145	Memuaskan
	0,755	0,643	0,768	0,755	0,678	0,263	0,570	0,627	
LMT	0,963	0,851	0,753	0,963	0,845	0,199	0,622	0,800	Sangat Memuaskan
	0,138	0,031	0,520	0,138	0,218	0,191	0,606	0,321	Dengan Pujian
	0,083	0,010	0,400	0,083	0,138	0,156	0,646	0,202	Memuaskan
	0,735	0,627	0,680	0,735	0,669	0,194	0,621	0,661	

Tabel 10. Confusion Matrix Hasil Pengujian Hasil Pemodelan Data Mining dengan Algoritma J48 dan LMT (a = Sangat Memuaskan, b = Dengan Pujian, c = Memuaskan)

	J48			LMT		
	a	b	c	a	b	c
a	696	4	1	675	19	7
b	162	26	1	161	26	2
c	68	0	4	61	5	6

Algoritma *Decision Tree* J48 berhasil mengklasifikasikan 726 data dengan benar sehingga menghasilkan tingkat akurasi sebesar 75,5% sedangkan LMT berhasil mengkasifikasikan 707 data dari 962 data dengan benar sehingga menghasilkan tingkat akurasi sebesar 73,5%. Berdasarkan hal tersebut dapat disimpulkan bahwa J48 mampu memprediksi predikat kelulusan

mahasiswa lebih baik dibandingkan dengan algoritma LMT.

Evaluasi

Berdasarkan pohon keputusan yang dihasilkan dari tahap pemodelan data mining dengan menggunakan algoritma J48 dan LMT, dapat diketahui bahwa atribut yang berpengaruh pada prediksi predikat kelulusan mahasiswa adalah IPS semester 2, IPS semester 1, dan IPS semester 4. Hal tersebut mendeskripsikan bahwa IPK akhir mahasiswa lebih dominan dipengaruhi oleh aspek akademik mahasiswa, yaitu nilai yang didapat oleh mahasiswa selama proses pembelajaran setiap semesternya. Mahasiswa yang setiap semesternya memiliki IPS lebih dari 3,6 diprediksi akan mendapatkan predikat kelulusan dengan

pujian atau memiliki IPK lebih dari 3,5. Mahasiswa yang setiap semesternya memiliki IPS berkisar antara 2,7 sampai 3,6 diprediksi akan mendapatkan predikat kelulusan sangat memuaskan atau IPK 2,76 sampai 3,5. Mahasiswa yang pernah memiliki nilai IPS sama dengan atau dibawah 2,43 diprediksi akan mendapatkan predikat kelulusan memuaskan atau IPK antara 2-2,75.

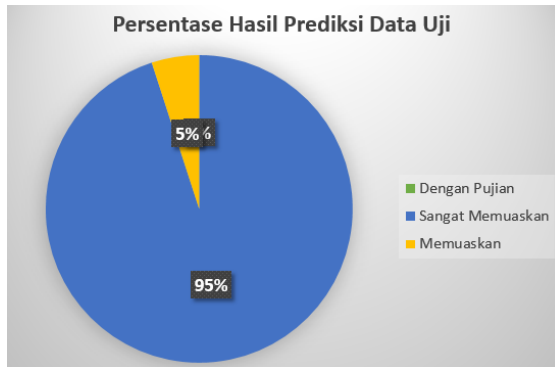
Pada studi kasus ini, algoritma J48 dengan jumlah objek minimum per daun sebanyak 15 objek dapat mengklasifikasikan suatu objek dengan lebih sederhana tanpa harus melakukan perhitungan peluang pada simpul akhirnya. Oleh karena proses klasifikasi yang sederhana, maka waktu yang dibutuhkan algoritma J48 untuk memprediksi suatu objek juga lebih singkat sehingga algoritma ini baik digunakan jika data yang akan diprediksi memiliki jumlah yang besar.

Menggunakan Pengetahuan yang Telah Ditemukan

Pada tahapan ini pengetahuan yang telah ditemukan kemudian digunakan untuk mengklasifikasikan data yang belum diketahui kelas. Data yang digunakan adalah data mahasiswa angkatan 2017 sebanyak 100 orang mahasiswa. Data tersebut digunakan sebagai data uji sedangkan data latih yang digunakan adalah data mahasiswa yang dinyatakan lulus pada yudisium tahun 2015 sampai 2019. Algoritma yang digunakan adalah algoritma J48. Contoh data uji yang digunakan dapat dilihat pada Tabel 11. Hasil klasifikasi data uji tersebut dalam bentuk grafik dapat dilihat pada Gambar 6 dimana hampir seluruh mahasiswa pada data uji tersebut diprediksi akan mendapatkan predikat kelulusan sangat memuaskan.

Tabel 11. Contoh Data Uji

NIM	Prodi	Jenis Kelamin	Usia	Alamat	Kelas	IPS1	IPS2	IPS3	IPS4
171351002	1	1	21	2	2	3,23	3,86	3,73	3,89
171351006	1	1	24	2	2	3,45	3,45	2,91	3,58
171351007	1	1	19	1	1	2,59	1,55	1,09	0,43
171351008	1	2	19	2	2	3,18	3,18	2,73	3,21
171351009	1	1	18	2	2	3,09	2,95	2,95	2,53
171151043	2	1	18	1	1	3	3,15	3,3	3,48
171151004	2	1	18	1	1	3	2,8	2,2	1,33
171151006	2	1	18	1	1	3,1	2,35	1,74	0,83
171151007	2	1	20	2	2	2,35	2,55	2,06	2,52
171151008	2	2	20	2	1	3	3,25	3,13	3,13
171251001	3	1	20	2	2	3,15	2,7	2,05	2,25
171251005	3	1	18	1	2	3,45	3,7	2,73	3
171251007	3	1	19	2	2	3,1	3,45	3,55	2,82
171251008	3	1	21	2	2	3,55	3,1	2,9	2,3
171251009	3	1	18	1	2	3,5	3,15	3,2	1,85



Gambar 6. Persentase Hasil Prediksi Data Uji

SIMPULAN

Berdasarkan penjelasan yang telah disampaikan sebelumnya dapat disimpulkan bahwa algoritma J48 dan LMT menghasilkan pohon keputusan dapat digunakan untuk mengklasifikasikan predikat kelulusan dan juga memprediksi predikat kelulusan mahasiswa yang masih berada pada tingkat tiga. Pemodelan data mining dengan menggunakan algoritma J48 menghasilkan akurasi sebesar 75,5% dengan data yang digunakan sebanyak 962 data, sedangkan LMT menghasilkan akurasi sebesar 73,5%. Pada kedua algoritma, atribut yang berpengaruh pada predikat kelulusan mahasiswa adalah IPS semester 2, IPS semester 1, dan IPS semester 4. Algoritma J48 mampu mengklasifikasikan predikat kelulusan mahasiswa dengan lebih baik dan melakukan proses klasifikasi dengan lebih sederhana juga membutuhkan waktu yang lebih singkat.

Penelitian ini masih dapat dikembangkan dengan menggunakan data mahasiswa seluruh jenjang studi sehingga dapat melakukan prediksi predikat kelulusan secara menyeluruh. Hasil prediksi tersebut hasilnya nanti dapat digunakan untuk membantu pihak kampus dalam mengambil kebijakan secara menyeluruh juga.

DAFTAR PUSTAKA

Agustiani, S., Mustopa, A., Saryoko, A.,

Gata, W., & Wildah, S. K. (2020). PENERAPAN ALGORITMA J48 UNTUK DETEKSI PENYAKIT TIROID. *Paradigma-Jurnal Komputer dan Informatika*, 22(2).

Alaoui, S. S., & Dkk. (2018). Classification Algorithms in Data Mining. *International Journal of Tomography and Simulation*, 34–44.

Bhatia, P. (2019). *Data Mining and Data Warehousing: Principles and Practical Techniques*. Cambridge University Pres.

Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D. T., ... & Ma, J. (2017). A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena*, 151, 147–160.

Colkesen, I., & Kavzoglu, T. (2017). The Use of Logistic Model Tree (LMT) for Pixel-and-Object-Based Classifications using High-Resolution WorldView-2 Imagery. *Geocarto International*, 32(1), 71–86.

Kaunang, F. J. (2019). Penerapan Algoritma J48 Decision Tree Untuk Analisis Tingkat Kemiskinan di Indonesia. *CogITO Smart Journal*, 4(2), 348. <https://doi.org/10.31154/cogito.v4i2.141.348-357>

Maulana, M. F., & Defriani, M. (2020). Logistic Model Tree and Decision Tree J48 Algorithms for Predicting the Length of Study Period. *PIKSEL: Penelitian Ilmu Komputer Sistem Embedded and Logic*, 8(1), 39–48.

Nachiappan, M. R., Sugumaran, V., & Elangovan, M. (2016). Performance of Logistic Model Tree Classifier using Statistical

Features for Fault Diagnosis of Single Point Cutting Tool. *Indian Journal of Science and Technology*, 9(47), 1–8.

Rokach, L., & Maimon, O. (2015). *Data Mining with Decision Trees: Theory and Applications* 2nd Ed. In *Singapore*. World Scientific Publishing Co.