

## Regresi Logistik Multinomial untuk Prediksi Kategori Kelulusan Mahasiswa

Rafika Syahrani<sup>(1)\*</sup>, Suhartono<sup>(2)</sup>, Syahiduz Zaman<sup>(3)</sup>

Teknik Informatika, Fakultas Sains dan Teknologi, UIN Maulana Malik Ibrahim, Malang  
e-mail : 17650003@student.uin-malang.ac.id, {suhartono,syahid}@ti.uin-malang.ac.id.

\* Penulis korespondensi.

Artikel ini diajukan 22 Desember 2022, direvisi 8 April 2023, diterima 10 April 2023, dan dipublikasikan 26 Mei 2023.

### Abstract

Students must meet certain goals to earn a degree but can extend their time at university or drop out (DO). The problem of dropping out of students has become an important issue for tertiary institutions to ensure the success or graduation of students and reduce dropouts. DO can affect the accreditation of the tertiary institution. The quality of higher education institutions in Indonesia is measured based on accreditation from the National Accreditation Board for Higher Education or BAN-PT. One of the main standards measured is the Quality of Students and Graduates. The quality of educational accreditation is measured by the percentage of student graduation and the university's strategy to retain students. To predict student graduation based on graduation time categories, researchers collected academic data from students in 2012-2018 at the Informatics Engineering Study Program, State Islamic University of Maulana Malik Ibrahim Malang. The variables used as predictors are gender, type of entry pathway, and grade point average from semesters one to six. The resulting model was evaluated to obtain an accuracy value of 85.5%, a precision of 78.5%, a recall of 93.9%, and a micro f1-score of 89.8%. An accuracy value of 85.5% indicates that the system can classify properly using the logistic regression model.

**Keywords:** Categories, Graduation, Prediction, Logistic Regression, Machine Learning

### Abstrak

Mahasiswa harus memenuhi tujuan tertentu untuk mendapatkan gelar, tetapi dapat memperpanjang waktu di universitas atau *dropout* (DO). Masalah DO mahasiswa telah menjadi isu penting bagi pengelola perguruan tinggi untuk menjamin keberhasilan atau kelulusan mahasiswa dan mengurangi DO. DO dapat mempengaruhi akreditasi perguruan tinggi tersebut. Kualitas perguruan tinggi di Indonesia diukur berdasarkan akreditasi yang dilaksanakan oleh Badan Akreditasi Nasional Perguruan Tinggi atau BAN PT. Salah satu standar utama yang diukur adalah Kualitas Mahasiswa dan Lulusan. Kualitas akreditasi pendidikan diukur dari persentase kelulusan mahasiswa dan strategi perguruan tinggi untuk mempertahankan mahasiswa. Untuk memprediksi kelulusan mahasiswa berdasarkan kategori waktu kelulusan, peneliti mengumpulkan data akademik mahasiswa tahun 2012-2018 pada Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang. Variabel yang digunakan sebagai prediktor adalah jenis kelamin, jenis jalur masuk, dan indeks prestasi semester satu sampai enam. Model yang dihasilkan diuji dengan evaluasi, sehingga didapatkan nilai *accuracy* sebesar 85,5%, *precision* sebesar 78,5%, *recall* sebesar 93,9%, dan *micro f1-score* sebesar 89,8%. Nilai *accuracy* sebesar 85,5% menyatakan sistem dapat melakukan klasifikasi dengan baik menggunakan model regresi logistik.

**Kata Kunci:** Kategori, Kelulusan, Prediksi, Regresi Logistik, Pembelajaran Mesin

## 1. PENDAHULUAN

Mahasiswa harus memenuhi tujuan tertentu untuk mendapatkan gelar, dan dalam banyak kasus dapat memperpanjang waktu mereka di perguruan tinggi atau mereka *dropout* (Dalipi et al., 2018). *Dropout* adalah tindakan yang dilakukan oleh perguruan tinggi untuk mengeluarkan mahasiswa karena perilaku yang tidak baik dan masa studi kuliah yang sudah melewati batas (Lu et al., 2018). Umumnya mahasiswa S1 menyelesaikan studinya dalam waktu yang ditentukan, yaitu delapan semester atau 4 tahun. Masalah yang timbul pada mahasiswa *dropout*



adalah batas masa studi mahasiswa yang ditempuh mahasiswa maksimal sampai 14 semester atau 7 tahun (Agwil et al., 2020).

*Dropout* merupakan masalah yang memengaruhi semua perguruan tinggi di seluruh dunia, seperti, penurunan pendaftaran, pendapatan berkurang untuk perguruan tinggi, kemalangan moneter untuk negara yang mendukung penyelidikan, dan menetapkan masalah sosial untuk studi sarjana, keluarga mereka, dan masyarakat pada umumnya (Asha et al., 2020). Mahasiswa *dropout* terdiri jenis mahasiswa yang dikeluarkan, putus kuliah, dan mengundurkan diri.

Masalah mahasiswa *dropout* telah mendapat perhatian yang luas dan telah menjadi isu penting bagi pengelola perguruan tinggi untuk menjamin keberhasilan atau kelulusan mahasiswa dan mengurangi angka mahasiswa *dropout* (Yaacob et al., 2020). Pada tahun 2019 angka mahasiswa *dropout* mencapai sebesar 602.208 mahasiswa dari jumlah mahasiswa yang terdaftar, data mahasiswa yang *dropout* didapatkan dari statisik pendidikan tinggi tahun 2020 (Handini et al., 2020).

Mencegah meningkatnya angka mahasiswa *dropout* sangat penting di berbagai lembaga pendidikan. Karakteristik *dropout* menjadi perhatian utama. Konsekuensinya termasuk kerugian finansial, tingkat kelulusan yang lebih rendah, dan reputasi perguruan tinggi yang buruk (Alturki et al., 2022). Kualitas perguruan tinggi di Indonesia diukur berdasarkan akreditasi yang dilaksanakan oleh Badan Akreditasi Nasional Perguruan Tinggi atau BAN PT. Kualitas tersebut diukur berdasarkan 9 standar utama, salah satunya adalah mahasiswa dan lulusan. Akreditasi pada perguruan tinggi sering dilihat dari keberhasilan mahasiswa menyelesaikan masa studinya. Meningkatnya jumlah mahasiswa *dropout* dapat mempengaruhi akreditasi perguruan tinggi tersebut.

Kualitas pendidikan diukur dari persentase kelulusan mahasiswa dan strategi perguruan tinggi untuk mempertahankan mahasiswa. Jika sebuah institusi kehilangan siswa karena alasan apapun, tingkat retensinya menurun. Identifikasi awal mahasiswa yang berisiko gagal sangat penting untuk keberhasilan strategi retensi apapun. Penting untuk mengidentifikasi mahasiswa ini sejak dini, dengan memberikan perhatian untuk mencapai retensi siswa. Kunci untuk mengurangi angka *dropout* adalah identifikasi dini mahasiswa berisiko, selain mempertahankan intervensi yang intensif dan berkelanjutan (Perez et al., 2018).

Penelitian ini membahas tipe pertama, yaitu memprediksi keberhasilan studi mahasiswa berdasarkan Kategori kelulusan. Sistem dibangun menggunakan teknik klasifikasi *machine learning* regresi logistik multinomial untuk memprediksi potensi lama kategori kelulusan mahasiswa pada tahap awal. Metode ini memiliki dua kegunaan utama, pertama untuk melakukan prediksi keanggotan kelompok di mana variabel dependen (respon) memiliki sifat politomi (lebih dari dua kategori). Kedua, metode regresi logistik dapat mengidentifikasi faktor kesuksesan yang paling berpengaruh (Hoffait & Schyns, 2017).

Al-Balushi & Islam (2020) mengungkapkan bahwa usia mahasiswa, jenis kelamin, perguruan tinggi, dan wilayah tempat tinggal merupakan prediktor penting dari nilai mahasiswa. Pengaruh variasi perguruan tinggi untuk usia mahasiswa, jenis kelamin mahasiswa, dan wilayah tempat tinggal lebih lanjut menyiratkan bahwa ada perbedaan yang cukup besar dalam nilai mahasiswa di antara perguruan tinggi (Urrutia-Aguilar et al., 2016). Mahasiswa yang proses penerimaannya melalui ujian memiliki motivasi diri lebih tinggi dibandingkan dengan mahasiswa yang diterima melalui metode tanpa ujian atau transfer otomatis.

Hashim et al. (2020) melakukan penelitian dengan menggunakan data mahasiswa selama enam semester menggunakan beberapa algoritma *Supervised Machine Learning*, seperti *decision tree*, Naïve Bayes, dan regresi logistik, *Support Vector Machine* (SVM), K-Nearest Neighbor, *Sequential Minimal Optimization* dan *neural network*. Hasil penelitian menunjukkan bahwa pengklasifikasi regresi logistik adalah yang paling akurat dalam memprediksi nilai akhir aktual mahasiswa dengan akurasi 68,7% untuk status lulus dan 88,8% untuk status gagal.



Umer et al. (2017) melakukan prediksi prestasi akademik mahasiswa dalam *Massive Open Online Courses* (MOOCs). Studi tersebut mengevaluasi empat teknik klasifikasi *machine learning*, yaitu regresi logistik, Naïve Bayes, *random forest*, dan K-Nearest Neighbor (KNN) di mana hasil studi menunjukkan bahwa metode regresi logistik dan Naïve Bayes memiliki tingkat akurasi yang tinggi serta mengungguli *random forest* dan KNN. Ashraf et al. (2018) melakukan penelitian yang menyimpulkan bahwa setiap pendekatan *data mining* menggunakan suatu algoritma memberikan hasil yang bervariasi sesuai dengan *dataset* dan variabel yang digunakan saat memprediksi. Berdasarkan hasil pengamatan, *decision tree*, regresi logistik, dan pendekatan *neural network* mampu memberikan hasil akurasi yang sangat baik mengenai prediksi masa mendatang serta membantu peningkatan sistem pendidikan.

Meskipun banyak penelitian tentang penggunaan metode regresi logistik untuk memprediksi keberhasilan akademik mahasiswa. Namun, masih terdapat kekurangan yang diperlukan untuk meningkatkan *performance model* prediksi meliputi identifikasi dan analisis variabel *input-output* yang mempengaruhi keberhasilan akademik mahasiswa dalam konteks yang berbeda yang telah dikembangkan oleh peneliti. Oleh karena itu, peneliti lebih lanjut diperlukan dalam membangun model menggunakan metode regresi logistik untuk memprediksi kategori kelulusan mahasiswa berdasarkan data akademik mahasiswa Program Studi Teknik Informatika UIN Malang Program S1 dari tahun 2012-2018 yang telah dilakukan pra-pemrosesan yang spesifik sehingga menghasilkan klasifikasi yang baik dengan nilai akurasi 85,5%.

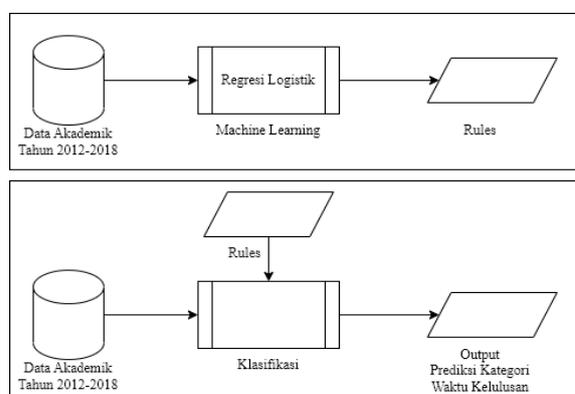
## 2. METODE PENELITIAN

### 2.1 Pengumpulan Data

Sumber data yang dibutuhkan dalam penelitian adalah data akademik mahasiswa Program Studi Teknik Informatika UIN Malang program S1 dari tahun 2012 - 2018. Data tersebut diperoleh melalui Bagian Administrasi Akademik (BAK) UIN Malang. Data induk mahasiswa merupakan data mahasiswa yang didata ketika mahasiswa pertama kali masuk perguruan tinggi setelah melakukan registrasi ulang. Sedangkan, data kelulusan merupakan data mahasiswa yang sudah menyelesaikan studinya di jurusan Teknik Informatika. *Dataset* berisi kumpulan data dari tujuh tahun akademik pertama yang diperlakukan untuk membangun model (2012-2018).

### 2.2 Desain Sistem

Desain sistem menjelaskan alur sistem dalam mengimplementasikan metode regresi logistik multinomial untuk melakukan prediksi Kategori kelulusan mahasiswa. Seperti diilustrasikan dalam Gambar 1, alur desain sistem terdiri dari masukan, *preprocessing*, proses *learning* dengan menggunakan metode regresi logistik multinomial, proses prediksi, dan evaluasi hasil prediksi, Dalam proses *learning*, pembagian data *training* dan data *testing* menggunakan perbandingan 70:30 di mana 70% untuk data *training* sedangkan 30% untuk data *testing*. Sistem akan dibangun menggunakan RStudio.



Gambar 1 Desain Sistem



### 2.3 Persiapan Data

Persiapan data diawali dengan tahapan *preprocessing* data, yaitu langkah yang dilakukan untuk membuat data mentah menjadi data berkualitas dengan input yang dibutuhkan untuk *data mining*. Tidak semua variabel data induk mahasiswa dimasukkan ke dalam *dataset*. Berdasarkan kebutuhan penelitian, variabel yang digunakan dalam *dataset* diterangkan dalam Tabel 1.

Tabel 1 Karakteristik *Dataset*

Variabel	Definisi	Indikator
<b>Dependen</b>		
IP	Penilaian yang diterapkan dikampus dengan hasil penjumlahan dari semua nilai mata kuliah dalam satu semester	Data IP selama semester 1 sampai semester 6
Jenis Kelamin	Pembeda antara laki-laki dan perempuan dilihat dari sudut pandang	L dan P
Jalur Masuk	Bagian penerimaan yang dilalui mahasiswa saat masuk ke perguruan tinggi (S1)	Prestasi dan Non Prestasi
<b>Independen</b>	<b>Definisi</b>	<b>Indikator</b>
Kategori waktu kelulusan	Lama mahasiswa menyelesaikan masa studinya	Cepat Tepat Lambat

Kategori cepat, tepat, dan lambat dilakukan secara manual dengan melihat data mahasiswa yang didapat dari Bagian Administrasi Akademik (BAK) UIN Maulana Malik Ibrahim Malang. Kategori cepat apabila pada data mahasiswa mahasiswa tersebut lulus pada semester 7 atau sebelum semester 8 dan telah memenuhi jumlah minimal SKS yang telah ditentukan yaitu sejumlah 151 SKS. Kategori tepat apabila pada data mahasiswa mahasiswa tersebut lulus tepat pada semester 8 dan telah memenuhi jumlah minimal SKS yang telah ditentukan yaitu sejumlah 151 SKS. Kategori lambat apabila pada data mahasiswa mahasiswa tersebut lulus lebih dari semester 8 dan telah memenuhi jumlah minimal SKS yang telah ditentukan yaitu sejumlah 151 SKS (Satsangkaryon, 2018).

### 2.4 Modeling

Sub bab ini menjelaskan implementasi metode klasifikasi regresi logistik multinomial untuk memprediksi kategori kelulusan mahasiswa. Untuk mencapai ini, model dikembangkan dari data *training* di mana nilai *input* maupun label *output* telah diketahui. Model menggeneralisasi hubungan antara *input* dan *output* kemudian menggunakannya untuk melakukan prediksi kategori kelulusan mahasiswa pada kumpulan data lain di mana hanya nilai *input* saja diketahui. Langkah-langkah dari metode yang diterapkan sebagai berikut:

- 1) Melakukan *input dataset* penelitian ke dalam sistem. Kemudian, pisahkan *dataset* menjadi dua *subset*: 70% untuk pelatihan dan 30% untuk validasi.
- 2) Komputasi regresi logistik pada data latih di mana, setiap paramater (x) menghasilkan nilai koefisien ( $\beta$ ) dan *intercept* ditambahkan ke dalam sistem. Nilai koefisien dari masing-masing parameter dikalikan dengan nilai paramater bersangkutan. Hasil setiap perkalian lalu dijumlahkan dengan menambahkan nilai *intercept*, seperti Pers. (1).

$$g(x) = \beta_0 + \beta_1x_1 + \dots + \beta_px_p \quad (1)$$

- 3) Nilai penjumlahan diterapkan ke dalam fungsi aktivasi—fungsi sigmoid atau fungsi logistik (Scott et al., 1991), yang mana mengubah data yang diberikan dalam kisaran (0, 1) ditunjukkan pada Pers. (2). Nilai *output* yang diukur dengan regresi logistik menyesuaikan dengan probabilitas, artinya nilai 1 menunjukkan peluang lebih baik untuk berada dalam kelas kategori kelulusan tertentu daripada nilai, contoh 0,7.



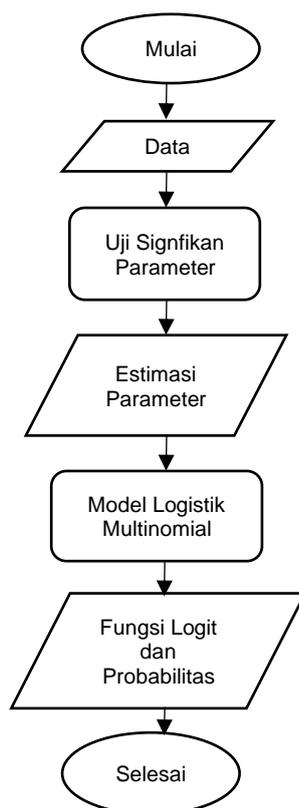
$$f(z) = \frac{1}{1+e^{-z}} \quad -\infty < z < +\infty \quad (2)$$

Di mana  $z = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p$  sehingga probabilitas bersyarat yang dimodelkan pada penelitian ini dapat dilambangkan pula dengan Pers. (3).

$$P(Y_i | x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4)}} \quad (3)$$

4) Evaluasi model menggunakan *multi-class confusion matrix*.

Diagram alir dari regresi logistik multinomial ditunjukkan pada Gambar 2.



**Gambar 2 Flowchart Regresi Logistik Multinomial**

### 2.5 Uji Signifikansi Parameter

Variabel-variabel yang diamati dalam penelitian ini dapat mempengaruhi keberhasilan akademik mahasiswa, menurut statistik deskriptif pada bab sebelumnya. Dengan nilai alpha 0,05 (Sari et al., 2017), peneliti selanjutnya meneliti statistik Wald dan perubahan statistik *Likelihood Ratio* untuk mengetahui signifikansi variabel regresi logistik. Hasil pengujian signifikansi parameter secara serentak ditunjukkan pada Tabel 2.

**Tabel 2 Estimasi Parameter Secara Serentak**

Pearson's chi-squared test
Data: Training\$Lama_Studi and Predict(mutinomModel)
X-squared = 136,83, df= 4, p-value <2.2e-16

Berdasarkan keterangan dari Tabel 2, diketahui bahwa hasil pengujian *Likelihood Ratio* memiliki nilai *p-value* sebesar 2.2e-16, di mana nilai ini lebih kecil dari 0,05. Hal ini mengartikan bahwa



terdapat paling sedikit satu variabel independen yang berpengaruh signifikan terhadap kategori kelulusan mahasiswa.

**Tabel 3 Estimasi Parameter Secara Parsial**

Coefficients:									
(Intercept)	JK	JM	IP1	IP2	IP3	IP4	IP5	IP6	
2.	-17.07503	0.2906574	0.1301306	-6.649710	2.659724	4.350708	3.3276037	-2.133633	3.407288
3.	19.61444	0.2878672	0.7440973	-4.489338	1.600311	0.223948	0.1354936	-5.101165	2.082269

Tabel 3 di atas menampilkan estimasi parameter secara parsial untuk kategori kelulusan mahasiswa di mana tepat diberi keterangan 2, lambat diberi keterangan 3. Berdasarkan pengujian tersebut, kategori kelulusan yang dijadikan pembanding adalah cepat dengan keterangan 1.

## 2.6 Model Logistik

Proses *training* menghasilkan dua buah fungsi logit, yaitu fungsi logit untuk kategori kelulusan tepat, dan lambat sementara kategori yang digunakan sebagai pembanding adalah kategori kelulusan cepat. Merujuk pada Pers. (1) maka fungsi regresi logistik multinomial untuk prediksi kategori kelulusan seorang mahasiswa apabila berjenis kelamin laki-laki, jalur masuk prestasi, IP semester 1 sebesar 3.55, IP semester 2 sebesar 3.15, IP semester 3 sebesar 3.38, IP semester 4 sebesar 3, IP semester 5 sebesar 3.46, IP semester 6 sebesar 3.7 adalah sebagai berikut:

- 1)  $g_2(x)$   

$$= -17.075 + 0.290x_1 + 0.130x_2 - 6.649x_3 + 2.659x_4 + 4.350x_5 + 3.327x_6 - 2.133x_7 + 3.407x_8$$

$$= -17.075 + 0.290(1) + 0.130(1) - 6.649(3.55) + 2.659(3.15) + 4.350(3.38) + 3.327(3) - 2.133(3.46) + 3.407(3.7)$$

$$= -1.974$$
- 2)  $g_3(x)$   

$$= 19.614 + 0.287x_1 + 0.744x_2 - 4.489x_3 + 1.600x_4 + 0.223x_5 + 0.135x_6 - 5.101x_7 + 2.082x_8$$

$$= 19.614 + 0.287(1) + 0.744(1) - 4.489(3.55) + 1.600(3.15) + 0.223(3.38) + 0.135(3) - 5.101(3.46) + 2.082(3.7)$$

$$= 0.962$$

Setelah mendapatkan nilai dari kedua model logit maka langkah berikutnya adalah mengukur nilai probabilitas masing-masing kategori dalam kategori kelulusan mahasiswa menggunakan Pers. (3) sehingga didapatkan nilai probabilitas sebagaimana berikut:

- 1)  $\pi_1(x) = \frac{1}{1 + \exp(-1.974) + \exp(0.962)} = 0.266$
- 2)  $\pi_2(x) = \frac{\exp(-1.974)}{1 + \exp(-1.974) + \exp(0.962)} = 0.036$
- 3)  $\pi_3(x) = \frac{\exp(0.962)}{1 + \exp(-1.974) + \exp(0.962)} = 0.696$

Dari fungsi probabilitas di atas, menunjukkan bahwa peluang mahasiswa dengan deskripsi berjenis kelamin laki-laki, jalur masuk prestasi, IP semester 1 sebesar 3,71, IP semester 2 sebesar 3,6, IP semester 3 sebesar 3,43, IP semester 4 sebesar 3,55, IP semester 5 sebesar 3,1, IP semester 6 sebesar 3,38, IP semester 7 sebesar 3,3, IP semester 8 sebesar 2,71 berpeluang paling besar mendapatkan kategori waktu kelulusan lambat di masa studinya dengan nilai peluang paling sebesar 0,696 (EI-Habil, 2012).



## 2.7 Evaluasi Model

Langkah selanjutnya adalah evaluasi terhadap model yang telah didapatkan dalam proses *learning* metode regresi logistik multinomial untuk prediksi kategori kelulusan mahasiswa. Evaluasi model diukur menggunakan *multiclass confusion matrix* dikarenakan penelitian ini termasuk ke dalam *multi-class classification*. Evaluasi model meliputi pengukuran akurasi, presisi, *recall*, dan *micro f1-score*. Akurasi klasifikasi adalah persentase data *testing* yang diklasifikasi secara benar oleh model. Apabila akurasi klasifikasi dianggap dapat diterima maka model dapat diterapkan untuk mengklasifikasi set data masa mendatang yang label kelas tidak diketahui (Agarwal, 2013). Presisi atau *precision* menyatakan proporsi unit yang diprediksi positif oleh model yang juga positif pada data sebenarnya. Presisi dapat diartikan sebagai kecocokan antara permintaan informasi dengan jawaban terhadap permintaan tersebut (Mayadewi & Rosely, 2015). *Recall* merupakan hasil perhitungan terhadap semua data uji yang positif yang telah diprediksi benar sebagai positif pada klasifikasi. *Recall* juga dikenal sebagai *True Positive Rate* (TPR), sensitivitas, dan probabilitas deteksi (Grandini et al., 2020). Seluruh kelas observasi digabung untuk mengukur *micro-averaged* presisi dan *micro-averaged recall* sehingga diperoleh rata-rata harmonik *micro-F1* (Zhang et al., 2015). Pers. (4) sampai (7) merupakan rumus dari akurasi, presisi, *recall*, dan *micro-F1*.

$$accuracy = \frac{TP}{Total\ data\ testing} \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$MicroF1 = 2 \times \left( \frac{Recall \times Precision}{Recall + Precision} \right) \quad (7)$$

## 3. HASIL DAN PEMBAHASAN

Metode regresi logistik multinomial diterapkan pada data *training* untuk proses *learning* sehingga menghasilkan dua buah model logit kategori kelulusan mahasiswa dengan satu kategori lainnya dijadikan pembanding. Peneliti membagi data menjadi dua bagian yang digunakan untuk proses *training* dan *testing* model dengan menggunakan perbandingan 70:30 di mana artinya 70% atau sebanyak 210 data digunakan sebagai data *training*, sedangkan 30% atau sebanyak 90 data digunakan sebagai data *testing*. Tabel 4 menunjukkan lima sampel dari data *training* sedangkan Tabel 5 menunjukkan lima sampel dari data *testing*.

Setelah mendapatkan model, maka pengujian dilanjutkan dengan membandingkan hasil prediksi model terhadap kategori kelulusan dalam 90 data *testing*. Hal ini bertujuan untuk mengukur kemampuan sistem prediksi Kategori kelulusan mahasiswa yang telah dibangun dalam hal presisi, *recall*, akurasi, dan *f1-score*. Hasil prediksi Kategori kelulusan mahasiswa tahun 2012-2018 menggunakan metode regresi logistik multinomial ditunjukkan pada Tabel 6.

Tabel 4 Sampel Data *Training*

No.	X_train							Y_train	
	Jenis kelamin	Jalur masuk	IP 1	IP 2	IP3	IP 4	IP5	IP 6	Kategori
1	2	2	3,4	3,26	3,59	3,57	3,61	3,61	2
2	2	2	3,38	3,31	3,26	3,17	3,23	3,5	3
3	1	2	2,88	2,29	2,81	2,73	3,16	2,23	3
4	2	2	3,57	3,59	3,72	3,69	3,46	3,73	2
5	2	2	3,45	3,07	3,83	3,85	3,76	3,79	2



Tabel 5 Sampel Data *Testing*

No.	X_train								Y_train
	Jenis kelamin	Jalur masuk	IP 1	IP 2	IP3	IP 4	IP5	IP 6	Kategori
1	2	1	3,55	3,31	3,42	3,59	3,05	3,34	3
2	1	2	3,45	3,4	3,23	3,02	3,46	2,95	1
3	1	2	3,62	3,56	3,48	3,56	3,15	3,31	3
4	2	1	3,76	3,61	3,74	3,79	3,52	3,86	2
5	2	2	3,38	2,98	3,1	3,36	3	3,52	3

Tabel 6 Hasil Prediksi Mahasiswa Tahun 2012-2018

No.	X_pred	Y_pred	Keterangan
1.	3	3	Sesuai
2.	1	3	Tidak Sesuai
3.	3	3	Sesuai
4.	2	2	Sesuai
5.	3	3	Sesuai

Visualisasi dari *multiclass confusion matrix* berdasarkan hasil prediksi model pada Tabel 6 ditunjukkan oleh Tabel 7.

Tabel 7 *Multiclass Confusion Matrix* Kategori Kelulusan

		Prediksi		
		1	2	3
Aktual	1	1	2	3
	2	1	16	2
	3	4	6	60

Analisis perhitungan nilai *True Postive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) pada *multiclass confusion matrix* dilakukan dengan menghitung per kelas kategori kelulusan. Kalkulasi perolehan nilai TP, TN, FP, dan FN dalam semua kelas kategori kelulusan tertera dalam Tabel 8.

Tabel 8 Kalkulasi *Multiclass Confusion Matrix*

Kategori Kelulusan	Confusion Matrix			
	TP	TN	FP	FN
Cepat	1	84	5	0
Tepat	16	65	6	3
Lambat	60	18	10	2
<b>Total</b>	<b>77</b>	<b>167</b>	<b>21</b>	<b>5</b>

Berdasarkan Tabel 8 maka nilai presisi, *recall*, akurasi, dan *micro f1-score* dapat diukur menggunakan Pers. (4) sampai Pers. (7). Hasil pengukuran sebagai berikut.

$$Accuracy = \frac{77}{90} = 0.855 = 85.5\%$$

$$Precision = \frac{77}{77 + 21} = 0.785 = 78.5\%$$

$$Recall = \frac{77}{77 + 5} = 0.939 = 93.9\%$$



$$Micro F1 = 2 \frac{(0.785 \times 0.939)}{(0.785 + 0.855)} = 0.898 = 89.8\%$$

Berdasarkan hasil evaluasi sistem dalam melakukan prediksi kategori kelulusan mahasiswa menggunakan metode regresi logistik multinomial menunjukkan nilai *accuracy* sebesar 85,5%, *precision* sebesar 78,5%, nilai *recall* sebesar 93,9%, dan nilai *micro F1* sebesar 89,8%. Nilai *accuracy* 85,5% menunjukkan sistem dapat melakukan klasifikasi dengan baik.

#### 4. KESIMPULAN

Peneliti telah membangun model prediksi dengan menggunakan regresi logistik multinomial berdasarkan *dataset* studi prediksi kategori kelulusan mahasiswa terkumpul sebanyak 300 data dengan variabel berupa jenis kelamin, jenis jalur masuk, indeks prestasi semester satu sampai enam sedangkan tiga tingkat klasifikasi kategori kelulusan program Sarjana (S1) dijadikan hasil. Model prediksi telah dilakukan uji dan evaluasi sehingga didapatkan nilai *accuracy* sebesar 85,5%, *precision* sebesar 78,5%, *recall* sebesar 93,9%, dan *micro f1-score* sebesar 89,8%. Kesimpulan berdasarkan nilai *accuracy* sebesar 85,5% menyatakan model prediksi dapat melakukan klasifikasi dengan baik sehingga penggunaan metode regresi logistik multinomial dapat diterima dan dapat digunakan untuk memprediksi kategori kelulusan mahasiswa. Saran untuk penelitian selanjutnya dapat menggunakan tambahan metode optimasi yang lain untuk menghasilkan nilai akurasi yang lebih optimal dan membuat model prediksi tersedia bagi administrator untuk membiarkan mereka membuat keputusan tentang penerapan dan kegunaannya secara dinamik sesuai dengan dunia nyata.

#### DAFTAR PUSTAKA

- Agarwal, S. (2013). Data Mining: Data Mining Concepts and Techniques. *2013 International Conference on Machine Intelligence and Research Advancement*, 203–207. <https://doi.org/10.1109/ICMIRA.2013.45>
- Agwil, W., Fransiska, H., & Hidayati, N. (2020). Analisis Ketepatan Waktu Lulus Mahasiswa dengan Menggunakan Bagging CART. *FIBONACCI: Jurnal Pendidikan Matematika Dan Matematika*, 6(2), 155. <https://doi.org/10.24853/fbc.6.2.155-166>
- Al-Balushi, M. S., & Islam, M. M. (2020). Predicting Academic Performance of Students of Sultan Qaboos University, Oman, Using Multilevel Modeling Approach. *Far East Journal of Theoretical Statistics*, 58(1), 59–76. <https://doi.org/10.17654/TS058010059>
- Alturki, S., Hulpuş, I., & Stuckenschmidt, H. (2022). Predicting Academic Outcomes: A Survey from 2007 Till 2018. *Technology, Knowledge and Learning*, 27(1), 275–307. <https://doi.org/10.1007/s10758-020-09476-0>
- Asha, P., Vandana, E., Bhavana, E., & Shankar, K. R. (2020). Predicting University Dropout through Data Analysis. *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*, 852–856. <https://doi.org/10.1109/ICOEI48184.2020.9142882>
- Ashraf, A., Anwer, S., & Khan, M. G. (2018). A Comparative Study of Predicting Student's Performance by use of Data Mining Techniques. *American Scientific Research Journal for Engineering, Technology, and Sciences*, 44(1), 122–136. [https://www.asrjetsjournal.org/index.php/American\\_Scientific\\_Journal/article/view/4170](https://www.asrjetsjournal.org/index.php/American_Scientific_Journal/article/view/4170)
- Dalipi, F., Imran, A. S., & Kastrati, Z. (2018). MOOC dropout prediction using machine learning techniques: Review and research challenges. *2018 IEEE Global Engineering Education Conference (EDUCON)*, 1007–1014. <https://doi.org/10.1109/EDUCON.2018.8363340>
- El-Habil, A. M. (2012). An Application on Multinomial Logistic Regression Model. *Pakistan Journal of Statistics and Operation Research*, 8(2), 271. <https://doi.org/10.18187/pjsor.v8i2.234>
- Grandini, M., Bagli, E., & Visani, G. (2020). *Metrics for Multi-Class Classification: an Overview*. <http://arxiv.org/abs/2008.05756>
- Handini, D., Hidayat, F., Putri, D. A. V., Rouf, M. R., Anjani, N. R., & Attamimi, A. N. R. (2020). *Statistik pendidikan tinggi tahun 2020 (higher education statistics 2020)*. Sekretariat Direktorat Jenderal Pendidikan Tinggi. <https://repositori.kemdikbud.go.id/22653/>
- Hashim, A. S., Awadh, W. A., & Hamoud, A. K. (2020). Student Performance Prediction Model based on Supervised Machine Learning Algorithms. *IOP Conference Series: Materials*



- Science and Engineering*, 928(3), 032019. <https://doi.org/10.1088/1757-899X/928/3/032019>
- Hoffait, A.-S., & Schyns, M. (2017). Early detection of university students with potential difficulties. *Decision Support Systems*, 101, 1–11. <https://doi.org/10.1016/j.dss.2017.05.003>
- Lu, O. H. T., Huang, J. C. H., Huang, A. Y. Q., & Yang, S. J. H. (2018). Applying learning analytics for improving students engagement and learning outcomes in an MOOCs enabled collaborative programming course. In O. H. T. Lu, J. C. H. Huang, A. Y. Q. Huang, & S. J. H. Yang (Eds.), *Learning Analytics* (pp. 78–92). Routledge. <https://doi.org/10.4324/9780429428500-7>
- Mayadewi, P., & Rosely, E. (2015). Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma Klasifikasi Data Mining. *Seminar Nasional Sistem Informasi Indonesia (SESINDO) 2015*, 2015. <https://is.its.ac.id/pubs/oajis/index.php/home/detail/1582/PREDIKSI-NILAI-PROYEK-AKHIR-MAHASISWA-MENGGUNAKAN-ALGORITMA-KLASIFIKASI-DATA-MINING>
- Perez, B., Castellanos, C., & Correal, D. (2018). Applying Data Mining Techniques to Predict Student Dropout: A Case Study. *2018 IEEE 1st Colombian Conference on Applications in Computational Intelligence (CoCACI)*, 1–6. <https://doi.org/10.1109/CoCACI.2018.8484847>
- Sari, A. Q., Sukestiyarno, Y., & Agoestanto, A. (2017). Batasan Prasyarat Uji Normalitas Dan Uji Homogenitas pada Model Regresi Linear. *Unnes Journal of Mathematics*, 6(2), 168–177. <https://doi.org/10.15294/UJM.V6I2.11887>
- Satsangkaryon, S. (2018). Analisis Korelasi Pemanfaatan Hasil Perkembangan Teknologi Informatika Terhadap Tingkat Kelulusan Mahasiswa pada Fakultas Ekonomi Universitas Pakuan Bogor. *JIMFE (Jurnal Ilmiah Manajemen Fakultas Ekonomi)*, 2(1), 73–87. <https://doi.org/10.34203/jimfe.v2i1.722>
- Scott, A. J., Hosmer, D. W., & Lemeshow, S. (1991). Applied Logistic Regression. *Biometrics*, 47(4), 1632. <https://doi.org/10.2307/2532419>
- Umer, R., Susnjak, T., Mathrani, A., & Suriadi, S. (2017). On predicting academic performance with process mining in learning analytics. *Journal of Research in Innovative Teaching & Learning*, 10(2), 160–176. <https://doi.org/10.1108/JRIT-09-2017-0022>
- Urrutia-Aguilar, M. E., Fuentes-García, R., Martínez, V. D. M., Beck, E., León, S. O., & Guevara-Guzmán, R. (2016). Logistic Regression Model for the Academic Performance of First-Year Medical Students in the Biomedical Area. *Creative Education*, 07(15), 2202–2211. <https://doi.org/10.4236/ce.2016.715217>
- Yaacob, W. F. W., Sobri, N. M., Nasir, S. A. M., Yaacob, W. F. W., Norshahidi, N. D., & Husin, W. Z. W. (2020). Predicting Student Drop-Out in Higher Institution Using Data Mining Techniques. *Journal of Physics: Conference Series*, 1496, 012005. <https://doi.org/10.1088/1742-6596/1496/1/012005>
- Zhang, D., Wang, J., & Zhao, X. (2015). Estimating the Uncertainty of Average F1 Scores. *Proceedings of the 2015 International Conference on The Theory of Information Retrieval*, 317–320. <https://doi.org/10.1145/2808194.2809488>

