Technologia : Jurnal Ilmiah
p-ISSN: 2086-6917, e-ISSN: 2656-8047
Homepage: https://ojs.uniska-bjm.ac.id/index.php/JIT/index
Vol. 16, No. 1, Januari 2025

KLASIFIKASI MASA STUDI MAHASISWA DENGAN PENDEKATAN NEURAL NETWORK BACKPROPAGATION BERDASARKAN JALUR MASUK

Muhammad Rais Wathani[1], Eka Setya Wijaya[2*], Mochamad Fajar Al-Amin[3], Zaenuddin[4]
[1,4] Program Studi Teknik Informatika, Fakultas Teknologi Informasi, UNISKA MAB, Indonesia.
[2*,3] Program Studi Teknologi Informasi, Fakultas Teknik, Universitas Lambung Mangkurat, Indonesia.

Informasi Artikel:
Dikirim: 22/11/2024; Diterima: 21/12/2024; Diterbitkan: 06/01/2025
Doi : http://dx.doi.org/10.31602/tji.v16i1.17051

Abstrak

Pada perguruan tinggi, data mengenai mahasiswa dan tingkat kelulusan mereka dapat menjadi sumber informasi yang berharga jika dianalisis lebih mendalam. Melalui analisis data mahasiswa, seperti statistik kelulusan tahunan, profil akademik, prestasi, serta durasi studi, universitas dapat memperoleh wawasan yang signifikan. Pengolahan data yang melimpah ini memungkinkan pengungkapan informasi tersembunyi yang berdampak besar bagi pengembangan institusi.
Penelitian ini mengkaji klasifikasi masa studi mahasiswa dengan menggunakan model Neural Network Backpropagation yang memberikan hasil terbaik dengan akurasi data latih sebesar 74,15% dan akurasi data uji sebesar 71,72%. Model ini juga menghasilkan nilai presisi sebesar 0,7422, dengan nilai recall 0,7422 dan F-Measure 0,7422. Berdasarkan hasil analisis, sebaran masa studi mahasiswa periode 2011-2022 menunjukkan, pada Jalur Masuk A, 74 mahasiswa menyelesaikan studi dalam waktu cepat, 978 dalam waktu normal, 445 terlambat, dan 69 sangat terlambat. Pada jalur masuk B, tercatat 29 mahasiswa menyelesaikan studi dengan cepat, 433 normal, 146 terlambat, dan 60 sangat terlambat. Terakhir, pada jalur masuk C, sebanyak 34 mahasiswa lulus dengan cepat, 448 normal, 294 terlambat, dan 151 sangat terlambat.

Keywords: Akurasi, *Data Mining*, Klasifikasi, Masa Studi, *Neural Network Backpropagation*

This is an open-access article under a Creative Commons Attribution 4.0 International (CC-BY 4.0) License. Copyright ©2025 by authors.

Pendahulua

Data mining adalah pendekatan analisis dan eksplorasi data berjumlah massif untuk menemukan pola-pola (pattern) yang belum terlihat sebelum dengan menggunakan pendekatan ini. Para peneliti sering memanfaatkan konsep ini dalam menganalisis informasi.[1].Di lembaga pendidikan tinggi, informasi yang bermanfaat dapat dihasilkan melalui data mengenai mahasiswa dan jumlah kelulusan mereka yang bisa digali lebih dalam lagi, Informasi yang berharga bagi univer-sitas dapat diperoleh melalui analisis data mahasiswa, termasuk statistik kelulusan tahunan, profil mahasiswa, prestasi akademik, dan durasi studi mereka. Melalui pengolahan data yang melimpah mengenai mahasiswa dan jumlah kelulusan, informasi yang sebelumnya tidak terungkap dapat diungkapkan, memberikan manfaat yang signifikan bagi universitas[2]. Masalah yang dihadapi oleh universitas secara umum dan fakultas secara khusus adalah variasi tingkat kelulusan di antara mahasiswa, di mana tantangan utamanya adalah tingginya jumlah mahasiswa yang tidak menyelesaikan studi tepat waktu.[3]. Pada penelitian ini dilakukan klasifikasi terhadap data lulusan mahasiswa Pada Universitas XYZ tahun 2011 sampai

2022. Dalam menentukan durasi studi mahasiswa, terdapat berbagai faktor yang berperan. Beberapa di antaranya meliputi faktor akademis, situasi keluarga, dan juga aspek psikologis[4]. Data Mining adalah proses ekstraksi informasi dari himpunan data yang memiliki dimensi yang besar, dengan memanfaatkan algoritma serta metode pengambilan informasi dari bidang statistik, Pembelajaran Mesin, dan manajemen basis data [5]. Neural Network Merupakan suatu teknik komputasi yang meniru struktur jaringan saraf dalam tubuh manusia. Teknik ini memanfaatkan operasi dasar yang bersifat non-linear yang disebut sebagai neuron, diatur dalam suatu jaringan yang saling terhubung, menyerupai hubungan jaringan saraf dalam tubuh manusia [6]. Maka dari itu, klasifikasi masa studi mahasiswa dapat dibuat dengan memanfaatkan model dari machine learning diantaranya menggunakan Algoritma Neural Network Backpropagation (BNN) dan dikelompokan berdasarkan jalur masuk. Klasifikasi dapat dibagi menjadi empat kelas masa studi yaitu, “cepat” untuk mahasiswa Sarjana yang lulus 3,5 tahun, “normal” untuk yang lulus 4 – 4,5 tahun, “lambat” untuk mahasiswa yang lulus 5 – 5,5 tahun, “sangat lambat” untuk mahasiswa yang lulus 6 tahun atau lebih.

Salah satu pendekatan penting dalam Data Mining adalah klasifikasi, yang digunakan untuk membangun model prediktif. Model ini mampu memberikan label pada data yang belum memiliki kelas, dengan tujuan utama memprediksi kelas yang relevan bagi data tersebut. Klasifikasi termasuk dalam pembelajaran berbimbing (supervised learning), di mana kelas data bersifat diskrit [7].

Salah satu teknik yang sering digunakan dalam proses klasifikasi adalah Neural Network. Pendekatan komputasi ini meniru cara kerja sistem saraf manusia, di mana neuron-neuron saling terhubung dalam suatu jaringan yang bersifat non-linier. Jaringan Syaraf Tiruan ini diciptakan untuk memecahkan berbagai masalah, termasuk pengenalan pola dan klasifikasi dalam proses pembelajaran data [6]. Untuk meningkatkan kinerja jaringan ini, metode Backpropagation digunakan dalam pelatihan model. Metode ini bertujuan untuk mencapai keseimbangan antara pengenalan pola selama pelatihan dan memberikan respons yang akurat terhadap input baru yang serupa dengan pola yang telah dilatih [8].

Dalam menguji performa model klasifikasi, teknik Cross Validation diterapkan. Teknik ini memungkinkan pengujian model secara berulang dengan membagi dataset menjadi beberapa subset yang berperan sebagai data pelatihan dan data pengujian secara bergantian. Proses ini disebut K-Fold Cross Validation, yang memastikan bahwa setiap bagian dari dataset dapat diuji dan dilatih secara bergantian, memberikan hasil performa yang lebih akurat dan stabil. Tahapan ini diulang sebanyak K-kali dengan variasi subset pengujian yang berbeda pada setiap iterasi. [9].

Selain itu, dalam kondisi di mana terdapat ketidakseimbangan (imbalance) antara data mayoritas dan minoritas, diterapkan teknik Synthetic Minority Oversampling Technique (SMOTE). Teknik ini berfokus pada pengambilan sampel pada data minoritas dengan cara menghasilkan data sintetis yang menggandakan jumlah data minoritas sehingga sejajar dengan jumlah data mayoritas. Pendekatan SMOTE bekerja dengan mengidentifikasi k-nearest neighbors (yaitu keterdekatannya dengan data sebanyak k) untuk setiap titik data di kelas minoritas. Kemudian, data sintetis dibuat dengan merujuk pada persentase duplikasi yang diinginkan, di antara data minoritas dan k-nearest neighbors yang dipilih secara acak. [10].

Untuk mengukur performa model klasifikasi, Confusion matrix digunakan sebagai alat evaluasi dalam bentuk matriks yang sangat bermanfaat dalam mengukur akurasi klasifikasi terhadap kelas dengan menggunakan algoritma tertentu[11]. Accuracy

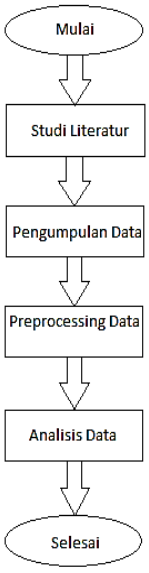
merupakan suatu metode evaluasi yang digunakan untuk menilai model klasifikasi dengan mengukur sejauh mana prediksi model mendekati nilai aktualnya[12]. Precision adalah suatu parameter evaluasi yang digunakan untuk mengidentifikasi sejauh mana positif yang diidentifikasi dengan benar[12]. Recall, digunakan sebagai pengukuran untuk mengukur seberapa baik kemampuan dalam mengenali negatif dengan benar[12]. F-Measure atau juga dikenal sebagai F1-Score, digunakan sebagai indikator untuk mengukur kualitas kinerja model klasifikasi. F-Measure adalah hasil rata-rata dari precision dan recall. Bentuk Confusion matrix dapat dilihat pada tabel 1.

Tabel 1. *Confusion Matrix*

Predicted Condition		***True Condition***	
	Total Population	*Condition Positive*	*Condition Negative*
	Prediction Condition Positive	*True Positive* (TP)	*False Positive* (FP)
	Prediction Condition Negative	*False Negative* (FN)	*True Negative* (TN)

Metodologi

Penelitian ini bertujuan untuk mengevaluasi dan menganalisis kinerja model terbaik yang dikembangkan dalam upaya mengklasifikasikan lama masa studi mahasiswa dengan menggunakan algoritma Neural Network Backpropagation. Algoritma ini diterapkan pada data lulusan/alumni untuk mengidentifikasi pola dan memprediksi kategori masa studi. Proses penelitian ini mengikuti langkah-langkah yang sistematis dan komprehensif, seperti yang ada dalam Gambar 1.



Gambar 1. Alur Penelitian

Tahap awal dari penelitian ini adalah melakukan studi literatur yang mendalam terkait dengan masa studi mahasiswa, serta kajian penelitian sebelumnya yang relevan. Penelusuran difokuskan pada teori-teori utama yang mendasari penggunaan Neural Network Backpropagation, serta perumusan masalah dan metode efisien yang dapat digunakan untuk mencapai hasil penelitian yang optimal. Tahap ini tidak hanya bertujuan untuk memperkuat landasan teoritis, tetapi juga untuk mengidentifikasi manfaat praktis yang dapat diperoleh dalam menyelesaikan permasalahan yang ada.

Tahap kedua melibatkan pengumpulan data yang berfokus pada data lulusan/alumni Universitas XYZ selama periode 2011-2022, yang berjumlah total 4.836 data. Data ini digunakan sebagai bahan analisis utama dalam penelitian. Potongan data yang lebih terperinci ditampilkan pada Gambar 2, memberikan gambaran mengenai atribut dan karakteristik penting yang dianalisis.

NIM	Nama	Kota Lahir	Tanggal Lahi	Jenis Kelamin	Telp/No H	Email	Program Studi	Fakultas	No Ijazah	nggal Yudisiu	IPK	Lama Studi	Predikat	Judul Skripsi	Pembimbing
A0.085727289197	AA	MARTAPUR	09/10/1997	Laki-Laki	085922121	akhmadridv	A	A3	10743260652	28 Juli 2023	3.05	7	Sangat Memuask	Kecernaan In <	1968041319
A0.520265832059	AB	BANJARMAS	20/07/1999	Perempuan	081256411	jihannbl20	B	A6	219785056605	05 Juli 2023	3.3	4.5	Sangat Memuask	Minat Siswa Terhadap Inst	1975032120
A0.788645800761	AC	BANJARMAS	07/11/1999	Perempuan	081244995	ummishabr	C	A1	44970006978	05 Juli 2023	3.13	4.5	Sangat Memuask	KONTRIBUSI FLEXIBILITAS	1975032120
A0.904429960142	AD	BANJARMAS	25/06/1997	Perempuan	081255744	PUTERIRN2	D	A7	329629864509	19 Januari 20	3.17	7.5	Sangat Memuask	PENGARUH PERSEPSI CASH	1975091320
A0.073023084617	AE	Barabai	19/03/1996	Laki-Laki	08125081	ikibadali@	E	A1	325796314896	17 Juni 2019	3.13	6	Sangat Memuask	Analisis Pendapatan Wanit	1958030115
A0.409633552688	AF	BANJARMAS	01/04/1998	Laki-Laki	081903421	bustanul22	F	A6	328676403591	05 Juli 2023	3.41	6.5	Sangat Memuask	Kualitas Pelayanan Admini	1967090919
A0.789948734791	AG	Simpang Tig	21/12/1996	Perempuan	081251854	rusmilawat	H	A1	80752266567	19 Januari 20	3.42	7	Sangat Memuask	Pengaruh Citra Merek, Kua	1961111019
A0.024661760600	AH	BANJARMAS	24/02/2001	Perempuan	08575487	sukmaenda	I	A1	606376683	05 Juli 2023	3.76	5	Sangat Memuask	PENGARUH INTEGRITAS, K	1959070519
A0.369486999957	AA	Mundar	21/06/2002	Perempuan	085753514	ekafatmaw	A	A9	130564224414	28 Desember	3.61	4.5	Sangat Memuask	Pengembangan Soal Berba	1975110520
A0.145144629196	AB	MARTAPUR	10/01/1998	Laki-Laki	08524693(	aaldialpian	B	A9	186460789230	03 Juli 2023	3.34	6	Sangat Memuask	PENGARUH JOGGING TER	1963090419
A0.170020687698	AC	BONTANG	14/07/1998	Perempuan	085283370	dillarizma1	C	A2	21669538437	02 Juli 2021	3.4	5	Sangat Memuask	PENGARUH LATIHAN	1956042719
A0.289883591696	AD	Banyuwang	07/05/1999	Laki-Laki	08783787	adamzaka1	D	A6	156082767660	19 Januari 20	3.58	4.5	Sangat Memuask	PENGARUH SOSIALISASI, P	1953082820
A0.127882389249	AE	Tapin	12/02/2000	Laki-Laki	081934475	aroyani5@	E	A5	55705463109	09 November	3.4	5.5	Sangat Memuask	PENGEMBANGAN MEDIA P	1960050719
A0.826325119337	AF	BANJAR	06/09/1999	Perempuan	08194586(	Herlina069	F	A8	139637899323	20 Juni 2023	3.28	7	Sangat Memuask	Pengaruh Citra Merek dan	1967042919
A0.833762881440	AG	Simpang Em	12/05/1997	Perempuan	083142446	nonita.sari	H	A3	164790130581	22 Juni 2022	3.1	6.5	Sangat Memuask	Pengaruh Media Poster Da	1963070519
A0.637381830148	AH	MARABATU	15/07/1997	Laki-Laki	082148254	darpialfatih	I	A4	328981239450	05 Juli 2023	3.04	7	Sangat Memuask	Implementasi Penyaluran I	1958081619
A0.612674933618	AA	Balikpapan	07/05/2000	Perempuan	088804118	auliashaaf	A	A6	166028963841	28 Desember	3.84	4.5	Sangat Memuask	Pelayanan Kesehatan Pasi	1982082120
A0.095853246048	AB	BANJARMAS	18/12/1998	Laki-Laki	085822465	ais.salman	B	A5	295281076347	28 Desember	3.54	6.5	Sangat Memuask	Strategi Sanggar Bahana A	1963122719
A0.958585200525	AC	JUKING PAJ	20/12/2000	Laki-Laki	085320721	rikeysaputr	C	A2	93076207956	28 Desember	3.51	4.5	Sangat Memuask	PENGGUNAAN MEDIA GAN	1973031320
A0.632696696869	AD	PALANGKA	09/11/1997	Laki-Laki	082166249	RFDUAEMF	D	A8	218937071772	05 Juli 2023	2.81	7	Memuaskan	PENGARUH <EM>VIRAL M	1978052120
A0.455491696790	AE	Banjarmasir	24/10/2000	Perempuan	081998030	191011612	E	A7	145746079362	06 Juli 2023	3.59	4	Pujian	KEWENANGAN PEMERINTA	1964021619
A0.438928319749	AF	BANJARMAS	07/07/2001	Perempuan	089711541	widyaterra	F	A3	243861027201	18 Desember	3.87	4.5	Sangat Memuask	ANALISIS EFEKTIVITAS PAJ	1970020919
A0.900814192661	AG	Tanah Laut	11/10/2002	Perempuan	08152283:	riskyoktapi	H	A7	155839264074	15 Januari 20	3.67	3.5	Pujian	KEWENANGAN PEMERINTA	1960041819
A0.126453960850	AH	Barabai	05/12/1999	Perempuan	081348766	adilahiffah	I	A2	49775795046	12 Desember	3.45	4.5	Sangat Memuask	Evaluasi Kualitas Air Sunga	1967042919
A0.922588138367	AA	Banjarmasir	07/04/2002	Laki-Laki	089988300	mfikriansha	A	A3	51291069588	15 Januari 20	3.59	3.5	Pujian	Kampung Janda: Konstruks	1961062119
A0.006519738555	AB	MARABAHA	14/04/2001	Perempuan	081931466	riaapriliana	B	A7	295421164785	28 Desember	3.49	4.5	Sangat Memuask	Realita Seni Bela Diri Kunt	1979112720
A0.927567666789	AC	SAMARIND/	08/04/2000	Perempuan	081645467	AprilasariO	C	A8	246270088062	28 Desember	3.42	5.5	Sangat Memuask	ANALISIS PENYUSUNAN RE	1975030720
A0.698928083996	AD	Desa pundu	04/04/1998	Laki-Laki	082240804	haliswma@	D	A3	149636426121	28 Desember	3.33	5.5	Sangat Memuask	PELECEHAN SEKSUAL DI AI	1965010119
A0.038312691734	AE	Bandung	04/05/2002	Perempuan	089668210	annisashat	E	A5	116693945244	22 Juni 2023	3.6	3	Pujian	HUBUNGAN LITERASI DIGI	1967062419
A0.814270946377	AF	Banjarmasir	04/11/2002	Perempuan	081258085	dewirostika	F	A2	270873324531	15 Januari 20	3.68	3.5	Pujian	Posisi Tawar Perempuan B	1979110920
A0.848833010000	AG	KANDANGA	27/05/2000	Laki-Laki	082253518	MHMMDN	H	A2	89309337264	28 Desember	3.64	5.5	Sangat Memuask	Pengaruh Promotion dan S	1957020619
1685xxxxxx	AH	ANJIR SERA	09/09/1999	Perempuan	083152011	normasari1	I	A3	75098597895	28 Desember	3.32	6.5	Sangat Memuask	ANALISIS STRATEGI PENGE	1975030720

Gambar 2. Dataset Lulusan

Data yang telah dikumpulkan akan melalui tahap Preprocessing untuk mempersiapkan data tersebut agar siap dianalisis lebih lanjut. Proses Preprocessing ini terdiri dari beberapa tahapan penting, yaitu:

1. Seleksi Data (*Selection Data*)

   Tahap pertama ini bertujuan untuk memilih atribut-atribut yang relevan dari *dataset* lulusan Universitas XYZ pada periode 2011-2022. Dari total 18 atribut yang tersedia, dilakukan proses seleksi guna menyaring atribut yang paling signifikan. Setelah seleksi dilakukan, tersisa 7 atribut utama yang akan digunakan dalam analisis lebih lanjut. Rincian atribut yang dipilih dapat dilihat pada Tabel 2.

Tabel 2. Atribut Penelitian

Variabel/Atribut	Penjelasan
NIM	Nomor Induk Mahasiswa, pada NIM juga dapat terlihat angkatan masuk mahasiswa
Nama	Nama mahasiswa
Jenis Kelamin	Jenis Kelamin Mahasiwa(Laki-laki/Perempuan)
Program Studi	Program Studi yang diambil oleh mahasiswa
IPK	Indeks prestasi kumulatif, yaitu nilai rata-rata yang menunjukkan prestasi akademis mahasiswa pada suatu periode tertentu. IPK dihitung dengan membagi total nilai yang diperoleh oleh total SKS (Satuan Kredit Semester) yang diambil.
Status Mengulang	Mengulang atau tidaknya mahasiswa pada sebuah mata kuliah
Jalur Masuk	Jalur Masuk mahasiswa baru yang terdiri dari tiga jalur yaitu, Jalur Masuk A, Jalur Masuk B, dan Jalur Masuk C

2. Pembersihan Data (*Cleaning Data*)
   Tahap pembersihan data dilakukan untuk memastikan bahwa dataset bebas dari variabel yang tidak relevan dan menghilangkan missing values serta nilai-nilai null yang dapat mengganggu analisis. Selain itu, kesalahan penulisan atau ketidakkonsistenan data diperbaiki dalam tahap ini untuk menjaga integritas data sebelum digunakan dalam proses selanjutnya.
3. Konstruksi Data *(Construct Data)*
   Setelah data dibersihkan, langkah berikutnya adalah membangun data yang lebih konstruktif. Pada tahap ini, dilakukan pembuatan atribut turunan dari atribut yang sudah ada. Selain itu, satu atribut tambahan ditambahkan untuk dijadikan class target, yang akan digunakan dalam proses klasifikasi. Atribut class target ini dapat dilihat lebih jelas pada Tabel 3

Tabel 3. *Class Target*

Masa Studi
Cepat
Normal
Terlambat
Sangat Terlambat

4. Transformasi Data *(Transformation Data)*
   Transformasi data merupakan proses dimana data diubah menjadi format numerik agar dapat diolah secara efektif menggunakan metode *Backpropagation Neural Network*. Pada tahap ini, data atribut seperti jenis kelamin dan program studi diubah menjadi bentuk numerik. Untuk data jenis kelamin, nilai 0 dan 1 digunakan sebagai representasi, sedangkan data program studi dan jalur masuk mahasiswa dikonversi dari bentuk nominal menjadi numerik, seperti yang ditunjukkan pada Tabel 4.

Tabel 4. *Transformation Data*

Program Studi	Inisialisasi
Program Studi A	[1,0,0,0,0,0,0]
Program Studi B	[0,1,0,0,0,0,0]
Program Studi C	[0,0,1,0,0,0,0]
Program Studi D	[0,0,0,1,0,0,0]
Program Studi E	[0,0,0,0,1,0,0]
Program Studi F	[0,0,0,0,0,1,0]
Program Studi G	[0,0,0,0,0,0,1]

5. Penggalian Data *(Data Mining)*
   Pada tahap ini, algoritma yang sesuai dipilih untuk menggali informasi dari data yang telah dikumpulkan dan diproses. Penelitian ini memanfaatkan metode *Backpropagation Neural Network* sebagai teknik utama untuk melakukan klasifikasi masa studi mahasiswa. Model ini dikembangkan menggunakan perangkat lunak *RapidMiner* yang dipilih karena kemampuannya dalam membangun model prediktif secara efisien.

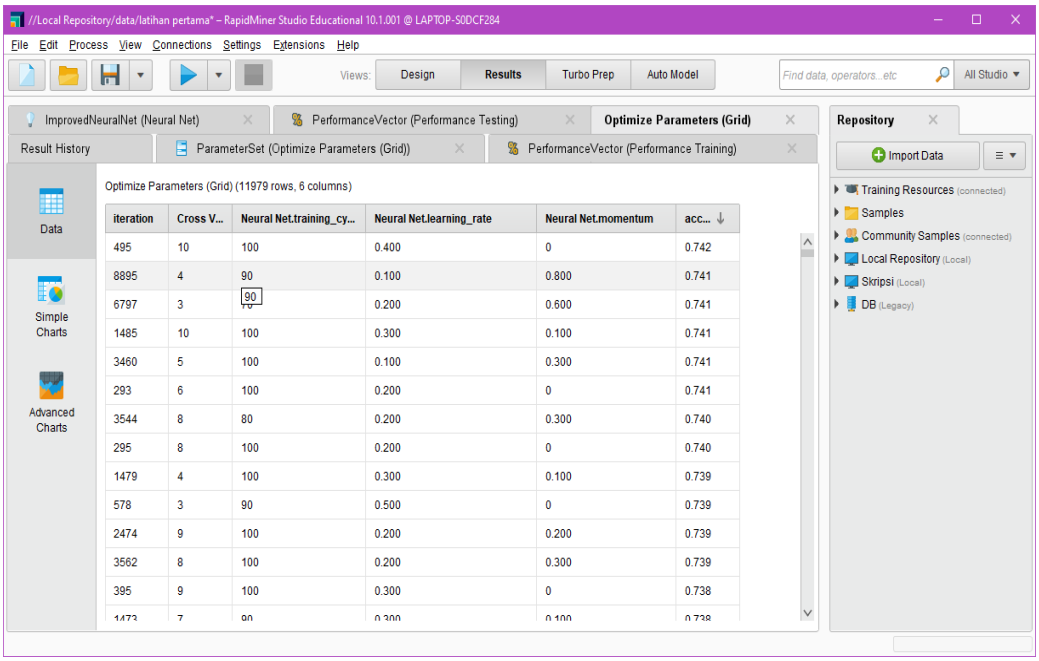
6. Evaluasi (*Evaluation*)

Tahap evaluasi dilakukan untuk menilai pola dan performa model yang dihasilkan dari proses data mining. Teknik *K-Fold Cross Validation* diterapkan untuk memastikan validitas model dengan melakukan pengujian berulang melalui pemecahan dataset menjadi beberapa subset pelatihan dan pengujian. Proses ini memungkinkan evaluasi yang lebih akurat terhadap model yang dibangun

Hasil

Pengujian model Neural Network dilakukan dengan mengubah berbagai parameter yang diteliti untuk mendapatkan hasil terbaik. Penelitian ini menguji beberapa parameter penting, seperti Training Cycle, Learning Rate, Momentum, jumlah K-fold Cross Validation, dan jumlah Neuron. Setiap parameter diuji dalam rentang tertentu untuk menemukan kombinasi yang menghasilkan akurasi optimal.

Rentang Training Cycle yang diuji berkisar antara 1 hingga 100, sementara Learning Rate diuji dalam rentang 4,9E-324 hingga 1,00. Parameter Momentum diuji dalam rentang 0 hingga 1,00, dengan jumlah K-fold Cross Validation bervariasi dari 2 hingga 10, diulang sebanyak 10 kali. Jumlah Neuron juga diuji dalam rentang 1 hingga 10. Data yang digunakan dalam pengujian ini dibagi dengan rasio 60% untuk data training dan 40% untuk data testing. Hasil akhir dari pengujian model Neural Network ini dapat dilihat pada Gambar 3, yang menunjukkan akurasi terbaik yang dihasilkan dari kombinasi parameter tersebut.



Gambar 3. Hasil Pengujuan Model

Berdasarkan Gambar di atas, akurasi terbaik dari model Neural Network ini tercapai pada akurasi data training sebesar 74,15%, sedangkan akurasi data testing mencapai 71,72%. Hasil ini diperoleh dengan parameter cross validation sebanyak 10, pada iterasi ke-495, dengan training cycle sebanyak 100, learning rate sebesar 0,400, dan momentum 0.

Tahap selanjutnya adalah melakukan evaluasi terhadap model yang telah dibangun, dengan menghitung matriks performa untuk menilai kinerja model secara lebih

mendalam. Matriks performa yang dihitung meliputi *precision*, *recall*, dan *f-measure*. Pada tahap ini, data training digunakan untuk mengevaluasi performa model yang telah dikembangkan. Data prediksi yang dihasilkan oleh model dibandingkan dengan label asli untuk menghitung nilai *precision*, *recall*, dan *f1-score*. Perhitungan nilai-nilai ini dapat dilihat pada perhitungan berikut

Precision

$$: \frac{\sum \text{True positive}}{\sum \text{True positive} + \sum \text{False negative}}$$

$$: \frac{2289}{2289+795}$$

$$: 0{,}7422178988326$$

Recall

$$: \frac{\sum \text{True positive}}{\sum \text{False positive} + \sum \text{True positive}}$$

$$: \frac{2289}{795_2289}$$

$$: 0{,}7422178988326$$

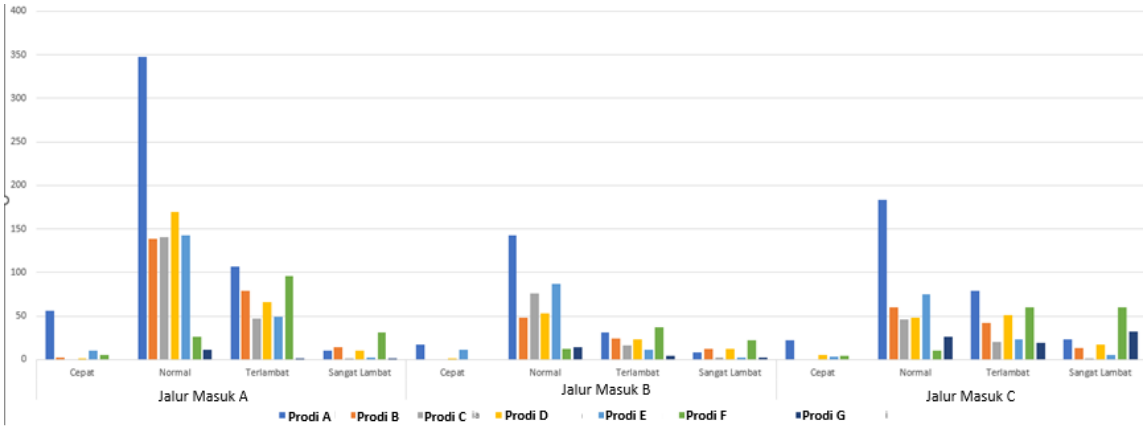
F-Measure

$$: \frac{2 \times precision \times recall}{precision + recall}$$

$$: \frac{2 \times 0{,}7422178988326 \times 0{,}7422178988326}{0{,}7422178988326 + 0{,}7422178988326}$$

$$: 0{,}7422178909205$$

Tahap terakhir dalam penelitian ini adalah mengimplementasikan model Neural Network yang telah dibangun untuk mengklasifikasi masa studi mahasiswa dari data *testing*. Hasil implementasi model tersebut dapat dilihat pada Gambar 4, yang menggambarkan distribusi prediksi masa studi berdasarkan jalur masuk dan program studi.



Gambar 4. Grafik Hasil Klasifikasi Masa Studi Mahasiswa Yang Masih Aktif Berdasarkan Jalur Masuk Dari Model Yang Dibangun

Dari grafik yang ditampilkan, dapat disimpulkan bahwa secara keseluruhan, mahasiswa yang masuk melalui Jalur A cenderung menyelesaikan masa studi lebih cepat

dibandingkan dengan mahasiswa yang masuk melalui Jalur B dan Jalur C. Sebaliknya, Jalur C menunjukkan bahwa mahasiswa pada jalur ini memiliki durasi masa studi yang lebih lama dibandingkan dua jalur lainnya.

Program Studi A, yang memiliki jumlah mahasiswa terbesar, juga menunjukkan distribusi tertinggi untuk kategori masa studi "cepat," dengan sebaran 56 mahasiswa pada Jalur A, 17 mahasiswa pada Jalur B, dan 22 mahasiswa pada Jalur C. Sementara itu, untuk kategori masa studi yang paling lama, Program Studi F mencatat sebaran tertinggi, dengan 31 mahasiswa pada Jalur A, 22 mahasiswa pada Jalur B, dan 60 mahasiswa pada Jalur C.

Pembahasan

Salah satu penyebab akurasi yang kurang optimal pada penelitian ini bisa berasal dari ketidakseimbangan data dalam kelas-kelas yang dianalisis. Model mungkin kesulitan untuk memprediksi kelas dengan proporsi yang lebih kecil (kelas minoritas). Walaupun teknik SMOTE (Synthetic Minority Oversampling Technique) sudah diterapkan, namun ini mungkin belum cukup untuk sepenuhnya mengatasi ketidakseimbangan data secara optimal.

Nilai akurasi yang tidak terlalu tinggi bisa juga disebabkan oleh pemilihan jumlah neuron yang kurang optimal pada lapisan tersembunyi. Model yang terlalu sederhana (dengan terlalu sedikit neuron) tidak mampu menangkap kompleksitas dari pola data yang dianalisis. Sebaliknya, terlalu banyak neuron juga dapat menyebabkan overfitting, di mana model bekerja sangat baik pada data training tetapi tidak dapat memprediksi dengan baik pada data testing.

Kesimpulan

Neural Network Backpropagation dapat mengklasifikasikan data lulusan untuk mendapatkan masa lama studi dengan data sebanyak 4.836, dibagi menjadi 1934 data testing dan data training sebanyak 2.902, dan mendapatkan hasil nilai akurasi training sebesar 74,15% dan tingkat akurasi testing sebesar 71,72%. Didapatkan perhitungan untuk nilai precision sebesar 0,7422178988326, untuk nilai recall sebesar 0,7422178988326 dan untuk F-Measure sebesar 0,7422178909205.

Nilai akurasi model Neural Network Backpropagation dari pengujian data testing masih dapat ditingkatkan dengan mempertimbangkan berbagai kemungkinan seperti, menyeimbangkan data, pemilihan dan penambahan parameter lain, atau juga eksplorasi metode machine learning lainnya untuk mencapai akurasi yang lebih optimal dan prediksi yang lebih akurat.

Referensi

[1] Han, J., & Kamber, M. (2006). Data Mining: Concepts and Techniques. SunFransisco: Morgan Kaufmann Publishers, 2006.

[2] Johan Oscar Ong, ‘Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President Uni-versity,’ Jurnal Ilmiah Teknik Industri, vol. 12, no.1, hal. 10-13, 2013.

[3] I. P. Astuti, “Prediksi Ketepatan Waktu Kelulusan Dengan Algoritma Data Mining C4.5,” Fountain of Informatics Journal, vol. 2, no. 2, hal. 5, 2017.

[4] S. S.-I. journal of disability studies “Faktor-faktor yang mempengaruhi keberhasilan studi mahasiswa penyandang disabilitas,” ijds.ub.ac.id, 2014.

[5] R. S. Taruna, S. H.-I. J. of C. Science “Enhanced Naïve Bayes algorithm for intrusion detection in data mining,” Citeseer 2013

[6] A. Sespajayadi, N. Rushadah, Indrabayu, E. Warny, I. I.-I. J., “Prediksi Nilai Tukar Valuta Asing Menggunakan Metode Genetic Algorithm-Neural Network (Euro Terhadap US Dollar),” jurnal.akba.ac.id, 2013

[7] F. Lumbanraja, W. M.-P. S., “Implementasi Metode Random Forest untuk Prediksi Posisi Metilasi pada Sekuens Pro-tein,” jurnal.fmipa.unila.ac.id, 2019.

[8] Siang, J.J. Jaringan Syaraf Tiruan dan Pemerograman menggunakan Matlab. ANDI, Yogyakarta, 2005.

[9] P. Klasifikasi, G. Musik, M. Metode, S. Untuk, and M. Persyaratan, “Implementasi Algoritma Genetika Untuk Seleksi Fitur.”,Banjarbaru, Indonesia, bab 2, bag. 7, hal. 16-17, 2022.

[10] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic minority over-sampling tech-nique,” Journal of Artificial Intelligence Research, vol. 16, hal. 321–357, 2002.

[11] T. S.-J. T. I. dan Sistem, “Komparasi dan Analisis Kinerja Model Algoritma SVM dan PSO-SVM (Studi Kasus Klasifi-kasi Jalur Minat SMA),” journal.maranatha.edu, 2016.

[12] M. Bekkar, H. Djemaa, T. A.-J. I. E. Appl “Evaluation measures for models assessment over imbalanced data sets,” eva.fing.edu.uy, vol. 3, hal. 10, 2013

.