
IMPLEMENTASI ALGORITMA ARTIFICIAL NEURAL NETWORK PADA PROSES SELEKSI PILIH PEMINATAN

Nassyfa Alfirda Riani¹, Rachmadita Andreswari², Rokhman Fauzi³

Sistem Informasi, Telkom University

e-mail: ¹ nassyfa@student.telkomuniversity.ac.id

Abstract: Changes in the 2020 curriculum regulation policy at the Telkom University Information Systems Study Program with the previous policy affected the selection process for student specialization. Based on the latest policy, the Information Systems Study Program provides five areas of specialization, consisting of two skill groups (Cybernetics and Enterprise and Industrial Systems), which students must choose in the seventh semester. Students often have difficulty in choosing specialization groups. They tend to avoid difficult specialization groups, and it's not uncommon for some students to choose specializations based on recommendations from others without considering their expertise and potential. Therefore, this study will model the selection process for selected students' specializations by adjusting the latest policies by applying machine learning using the Classification method of the Artificial Neural Network (ANN) algorithm. Researchers chose this algorithm because it can classify and understand very complex patterns. In this case, data processing will use attributes, such: Student Identification Number (NIM), the average value of prerequisite courses in the field of specialization, the first and second choices of the student specialization groups, and the quota in the specialization group that comes from the number of lecturers who teaches in the specialization group. The implementation of ANN, in this case, is excellent because it produces an AUC of 0.9950 and an accuracy of 0.9863.

Keywords: students specialization; classification; ANN.

Abstrak: Perubahan kebijakan peraturan kurikulum 2020 pada Program Studi Sistem Informasi Universitas Telkom dengan kebijakan sebelumnya berpengaruh terhadap proses seleksi peminatan mahasiswa. Berdasarkan kebijakan terbaru, Prodi Sistem Informasi menyediakan lima bidang peminatan yang tergabung dalam dua kelompok keahlian (*Cybernetics* dan *Enterprise and Industrial System*) yang nantinya harus dipilih oleh mahasiswa di semester tujuh. Mahasiswa seringkali kesulitan dalam menentukan bidang peminatan yang akan dipilih, karena cenderung akan menghindari bidang peminatan yang dianggap sulit dan tak jarang pula terdapat mahasiswa yang menentukan peminatan atas rekomendasi orang lain tanpa mempertimbangkan keahlian dan potensi yang mereka miliki. Oleh karena itu, penelitian ini akan membuat model proses seleksi pemilihan peminatan mahasiswa dengan menyesuaikan kebijakan terbaru dengan mengimplementasikan machine learning menggunakan metode Klasifikasi dari algoritma Artificial Neural Network (ANN). Peneliti memilih algoritma ini karena dapat mengklasifikasi dan memahami pola yang sangat kompleks. Pada kasus ini, pada proses pengolahan data peneliti akan menggunakan atribut NIM, rata-rata nilai matakuliah prasyarat bidang peminatan, pilihan bidang peminatan pertama dan kedua mahasiswa, status keikutsertaan mahasiswa dalam keprofesian, serta kuota bidang peminatan yang berasal dari jumlah dosen yang mengajar pada bidang peminatan tersebut. Pengimplementasian ANN pada kasus ini sangat baik karena menghasilkan AUC sebesar 0.9950 dan akurasi sebesar 0.9863.

Kata kunci: bidang peminatan; klasifikasi; ANN.

PENDAHULUAN

Peminatan merupakan suatu sistem penjurusan yang akan dipilih oleh mahasiswa program studi Sistem Informasi Universitas Telkom sesuai dengan keinginan dan keahlian mahasiswa. Peminatan berfungsi untuk memfokuskan mahasiswa pada satu ilmu keahlian khusus sesuai minat masing-masing, selain itu peminatan juga berfungsi sebagai penentu topik tugas akhir yang akan dipilih oleh mahasiswa. Berdasarkan kebijakan terbaru, Prodi Sistem Informasi menyediakan lima bidang peminatan diantaranya *Enterprise Data Engineering (EDE)*, *Enterprise Intelligent System Development (EISD)*, *System Architecture and Governance (SAG)*, *Enterprise Resource Planning (ERP)*, dan *Enterprise Infrastructure Management (EIM)* dibawah dua kelompok keahlian (*Cybernetics* dan *Enterprise dan Industrial System*) yang nantinya akan dipilih oleh mahasiswa pada semester tujuh. Pada saat memilih bidang peminatan, seringkali mahasiswa mengalami kesulitan dalam menentukan bidang peminatan sehingga menyebabkan tidak meratanya jumlah mahasiswa disalah satu kelompok keahlian. Ketidakmerataan ini dapat mempengaruhi ketersediaan dosen sebagai pembimbing tugas akhir mahasiswa. Sehingga diperlukan adanya suatu sistem yang dapat membantu mahasiswa menemukan bidang peminatan sesuai dengan kemampuan dan nilai akademis mereka. Penerapan kebijakan peraturan kurikulum 2020 pada Program Studi Sistem Informasi berpengaruh terhadap proses seleksi peminatan mahasiswa karena adanya perbedaan jumlah bidang peminatan dan matakuliah prasyarat peminatan.

Oleh karena itu, penelitian ini dibuat untuk membuat model seleksi peminatan sesuai dengan kebijakan terbaru menggunakan bantuan machine learning. *Machine learning* merupakan subbidang

pada ilmu *Artificial Intelligence (AI)* yang berfokus untuk menciptakan suatu sistem yang dapat belajar sendiri untuk membuat suatu keputusan yang baik dengan cara melatih komputer sedemikian rupa agar dapat memahami model objek yang dipahami manusia (Wahyono, 2018). Penggunaan teknik *machine learning*, diharapkan dapat digunakan untuk menyelesaikan permasalahan ini.

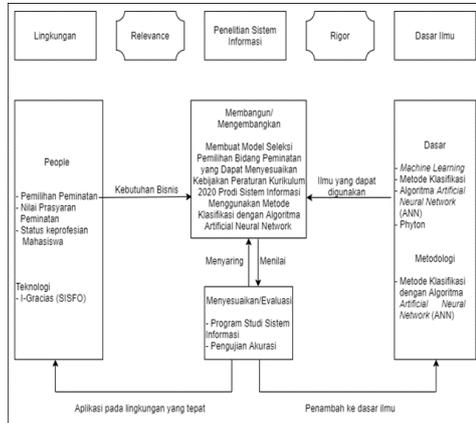
Pada machine learning terdapat beberapa metode yang dapat digunakan, salah satunya yaitu klasifikasi. Klasifikasi merupakan salah satu jenis analisis data yang dapat digunakan untuk membantu dalam prediksi label kelas sampel yang harus diklasifikasikan. Saat melakukan klasifikasi diperlukan bantuan algoritma untuk membuat suatu model yang baik, salah satu algoritma yang dapat digunakan untuk proses klasifikasi yaitu algoritma *Artificial Neural Network (ANN)*.

ANN adalah salah satu metode yang paling umum untuk melakukan klasifikasi. ANN dapat memperkirakan berbagai model statistik, dan dapat secara fleksibel menggambarkan model linier maupun non linier. ANN dapat digunakan untuk masalah yang sama dengan masalah statistika multivariate, seperti regresi berganda, analisis diskriminan, dan analisis kluster (Rahman, Darmawidjaja, & Alamsah, 2017). Manfaat lain dari ANN yaitu dapat diproses secara implisit. Oleh karena itu, jika beberapa sel dalam jaringan dihapus atau memiliki fungsi yang salah, maka masih ada peluang untuk mendapatkan jawaban yang benar (Hassanipour, et al., 2019). Oleh karena itu berdasarkan alasan-alasan diatas, algoritma ANN cocok untuk diterapkan dalam penelitian ini.

METODE

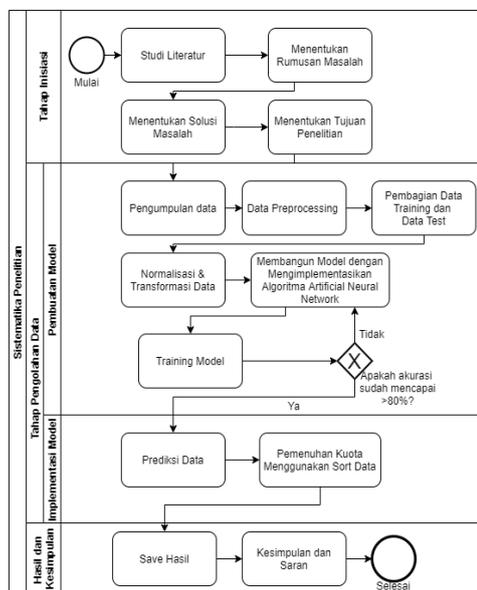
Penelitian ini menggunakan kerangka pemikiran *design science in IS research* untuk membantu memahami, melaksanakan, dan mengevaluasi

penelitian melalui kerangka kerja konseptual yang ringkas dan panduan yang jelas (Hevner, 2004). Kerangka penyelesaian pada penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Model Konseptual Hevner

Penelitian ini dikerjakan menggunakan metode klasifikasi dengan pendekatan *machine learning* serta perhitungan akurasi algoritma menggunakan *confusion matrix multiclass*. Tahapan yang dilakukan pada penelitian ini terdapat pada Gambar 2.

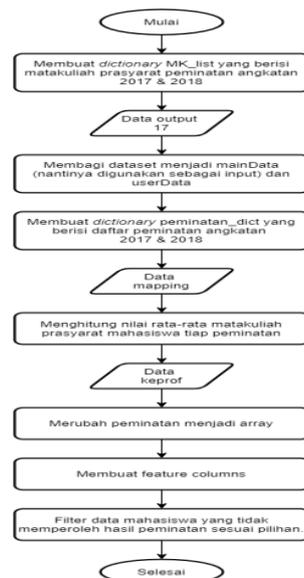


Gambar 2. Sistematisasi Penelitian

Penelitian ini dibagi menjadi 3 tahapan. Tahap pertama yaitu inisiasi

dengan melakukan studi literatur dan merumuskan masalah mengenai seleksi bidang peminatan mahasiswa sesuai dengan kebijakan terbaru yang diterapkan pada prodi S1 Sistem Informasi. Selanjutnya peneliti menentukan solusi yang dapat diimplementasikan pada studi kasus kali ini. Kemudian peneliti juga menentukan tujuan dari penelitian.

Tahap pengolahan data peneliti membagi menjadi 2 tahapan, yaitu tahap pembuatan model dan tahap implementasi model. Pada tahap pembuatan model, peneliti melakukan pengumpulan data – data yang akan digunakan sebagai dataset pada penelitian ini. Peneliti menggunakan data mahasiswa Program Studi Sistem Informasi angkatan 2017. File yang digunakan terdiri dari beberapa file yaitu data mahasiswa, data peminatan, data status keprofesian, dan data dosen. Data mahasiswa diperoleh dari SISFO Universitas Telkom, sementara data lainnya diberikan oleh dosen program studi Sistem Informasi. Selanjutnya peneliti melakukan data *preprocessing* untuk menyiapkan data input yang berkualitas. Langkah – langkah yang dilakukan peneliti untuk mendapatkan data yang berkualitas yaitu sesuai dengan Gambar 3.



Gambar 3. Flowchart Preprocessing Data

Kemudian peneliti membagi dataset menjadi data training dan data testing. Dari 361 dataset, peneliti membaginya dengan rasio 8:2. Berdasarkan rasio tersebut, maka didapatkan 288 data training dan 73 data testing dengan `batch_size = 32`. Setelah data sudah terbagi menjadi data testing dan data testing, maka selanjutnya yaitu melakukan transformasi data dengan menormalisasikan nilai rata-rata tiap peminatan menggunakan *scaler*. Tujuan normalisasi data setelah pembagian data yaitu untuk meningkatkan akurasi, selain itu juga untuk mencegah adanya data *leakage* (Erwan, 2019). *Leakage* merupakan kebocoran data akibat adanya informasi tambahan yang tidak terduga dalam data training yang nantinya dapat menghasilkan prediksi yang tidak realistis. Setelah mendapatkan data yang ideal maka selanjutnya yaitu mengimplementasikan algoritma *Artificial Neural Network* untuk mendapatkan model pada data training. Implementasi dilakukan dengan melatih data ke dalam program pembelajaran sehingga nantinya program dapat memprediksi objek dengan membentuk *layer* dan *dropout* untuk menghasilkan prediksi sesuai dengan penelitian. Penggunaan *dropout* berfungsi untuk meminimalisir terjadinya *overfitting*. Kemudian peneliti melakukan evaluasi model menggunakan *multiclass classifier metrics* dengan kurva ROC. Selanjutnya pada tahapan pengimplementasian model, peneliti memprediksi *score* peminatan mahasiswa dengan melakukan *load_model* dari hasil pembentukan model dengan algoritma sebelumnya. Hasil prediksi akan disimpan kedalam kolom `Eval_Score` dan disatukan dengan data user. Tahap selanjutnya peneliti melakukan sortir data mahasiswa untuk memenuhi kuota di tiap bidang peminatan. Tahap terakhir yang dilakukan peneliti yaitu dengan menarik kesimpulan untuk merangkum hasil yang telah dibahas pada penelitian ini, serta memberikan saran yang nantinya dapat

dijadikan sebagai bahan pertimbangan pada evaluasi penelitian selanjutnya.

HASIL DAN PEMBAHASAN

Beberapa tahapan yang dilakukan pada tahap data *preprocessing* berfungsi untuk mendapatkan atribut yang digunakan pada penelitian ini. Daftar atribut yang digunakan pada penelitian ini terdapat pada Gambar 4. serta contoh data yang digunakan terdapat pada Gambar 5. dan Gambar 6.

Atribut	Deskripsi	Nilai
nim	Nomer Induk Mahasiswa (NIM) yang digunakan sebagai ID.	Unique ID
mk_pil_0	Atribut yang berisikan rata-rata nilai matakuliah prasyarat untuk peminatan yang memiliki ID 0.	Nilai dengan rentang 0 - 4
mk_pil_1	Atribut yang berisikan rata-rata nilai matakuliah prasyarat untuk peminatan yang memiliki ID 1.	Nilai dengan rentang 0 - 4
mk_pil_2	Atribut yang berisikan rata-rata nilai matakuliah prasyarat untuk peminatan yang memiliki ID 2.	Nilai dengan rentang 0 - 4
mk_pil_3	Atribut yang berisikan rata-rata nilai matakuliah prasyarat untuk peminatan yang memiliki ID 3.	Nilai dengan rentang 0 - 4
mk_pil_4	Atribut yang berisikan rata-rata nilai matakuliah prasyarat untuk peminatan yang memiliki ID 4.	Nilai dengan rentang 0 - 4
mk_pil_5	Atribut yang berisikan rata-rata nilai matakuliah prasyarat untuk peminatan yang memiliki ID 5.	Nilai dengan rentang 0 - 4
mk_pil_6	Atribut yang berisikan rata-rata nilai matakuliah prasyarat untuk peminatan yang memiliki ID 6.	Nilai dengan rentang 0 - 4
keprofesian	Menjelaskan status keikutsertaan mahasiswa pada suatu keprofesian.	Daspro, Sisjar, SAG, ERP, EISD
id_peminatan_1	Merupakan atribut yang berisi pilihan pertama peminatan yang dipilih mahasiswa.	0, 1, 2, 3, 4, 5, 6
id_peminatan_2	Merupakan atribut yang berisi pilihan kedua peminatan yang dipilih mahasiswa.	0, 1, 2, 3, 4, 5, 6
peminatan_x	Atribut yang berisi ID hasil peminatan yang didapat mahasiswa.	0, 1, 2, 3, 4, 5, 6
peminatan	Berisi kuota per peminatan yang didapat dari jumlah dosen aktif pada setiap peminatan.	0, 90, 120, 50, 150, 0, 60

Gambar 4. Daftar Atribut yang Digunakan

PadaSetelah melakukan tahap pembagian dataset, normalisasi dan transformasi data maka didapatkan data input seperti yang terlihat pada Gambar 5. dan Gambar 6.

mk_pil_0	mk_pil_1	mk_pil_2	mk_pil_3	mk_pil_4	mk_pil_5	mk_pil_6	keprofesian	id_peminatan_1	id_peminatan_2	peminatan_x	peminatan
3.375	3.000	3.1875	3.375	3.250	3.625	3.375	7	5			
3.250	2.500	3.1250	3.000	3.250	3.375	2.875	7	6			
2.500	3.125	2.3125	3.125	3.125	3.375	3.125	7	6			
3.625	3.625	3.1875	3.125	3.125	3.500	3.375	0	0			
3.750	3.375	3.2500	3.375	3.500	3.500	3.250	4	4			
...			
3.375	2.375	3.3125	2.750	3.000	3.625	3.625	6	6			
3.625	3.750	3.2500	3.375	3.625	3.500	3.375	7	6			
2.875	3.625	3.5000	3.375	3.125	3.250	3.500	7	0			
3.000	2.875	2.5625	2.750	2.500	3.250	3.125	7	4			
3.750	3.875	3.6875	3.250	3.500	3.500	3.625	7	1			

Gambar 5. Contoh Data Input (1)

peminatan_0	peminatan_1	peminatan_2	peminatan_3	peminatan_4	peminatan_5	peminatan_6
0	0	0	0	0	1	0
0	0	0	0	0	0	1
0	0	0	0	0	0	1
1	0	0	0	0	0	0
0	0	0	0	1	0	0
...
0	0	0	0	0	0	1
0	0	0	0	0	0	1
1	0	0	0	0	0	0
0	1	0	0	0	0	0
0	1	0	0	0	0	0

Gambar 6. Contoh DataInput (2)

Gambar 5. merupakan contoh data dari variabel *input*, sedangkan Gambar 6. adalah contoh dari variabel *output*. Pada tahap selanjutnya peneliti melakukan pembagian data set dengan rasio 8 : 2 lalu melakukan normalisasi pada nilai rata – rata mahasiswa menggunakan metode *Min Max Normalization*. *Min Max Normalization* merupakan cara standarisasi data dengan memetakan sebuah nilai kedalam rentang baru 0 dan 1. Setelah melakukan normalisasi pada data rata-rata mahasiswa maka didapatkan data baru seperti Gambar 7.

mk_pil_0	mk_pil_1	mk_pil_2	mk_pil_3	mk_pil_4	mk_pil_5	mk_pil_6
0.821429	0.76	0.571429	0.708333	0.750000	0.809524	0.764706
0.892857	0.84	0.761905	0.875000	0.857143	0.857143	0.764706
0.857143	0.92	0.833333	0.875000	0.785714	0.904762	0.823529
0.857143	0.64	0.571429	0.791667	0.750000	0.809524	0.647059
0.678571	0.04	0.500000	0.333333	0.678571	0.619048	0.529412

Gambar 7. Nilai Rata-Rata Setelah Normalisasi

Selanjutnya peneliti membuat model menggunakan ANN dengan algoritma *backpropagation*. Algoritma *backpropagation* biasa digunakan dalam metode klasifikasi dengan pendekatan *supervised learning* yang dan terdapat banyak lapisan (*multi layer network*) yang terdiri dari *input layer*, *hidden layer*, dan *output layer*. Pada penelitian ini peneliti menggunakan 3 hidden layer (neuron =

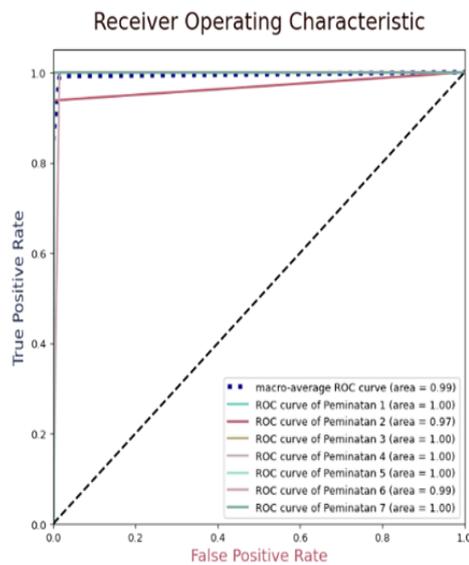
128, 64, 42 dengan activation Relu), output (neuron = 7 dengan activation Sigmoid) dan dropout sebesar 0.2 pada setiap data input. Penggunaan dropout berfungsi untuk meminimalisir terjadinya overfitting. Setelah menginputkan model maka peneliti melakukan fitting model sebanyak 100 kali epoch. Epoch adalah ketika semua dataset melakukan satu kali proses *forward* and *backward* melewati semua *node neural network*. Selain itu peneliti juga menggunakan *optimizer adam*.

Setelah mendapatkan model maka peneliti melakukan perhitungan performansi algoritma. Dikarenakan penelitian ini memiliki 7 kelas, maka peneliti menggunakan *multiclass classifier metrics* dengan kurva ROC. Setelah mengimplementasikan algoritma untuk membuat model maka didapatkan hasil seperti tabel di bawah ini.

Tabel 1. Hasil Prediksi Confusion Matrix Multiclass

Klasifikasi	Prediksi						
	0	1	2	3	4	5	6
Aktual	0	8	0	0	0	0	0
	1	0	15	0	0	0	1
	2	0	0	14	0	0	0
	3	0	0	0	9	0	0
	4	0	0	0	0	7	0
	5	0	0	0	0	0	7
	6	0	0	0	0	0	0

Berdasarkan hasil pada Tabel 1., maka pada implementasi model menggunakan ANN dengan menginputkan 3 hidden layer (neuron = 128, 64, 42 dan activation = ReLU), dropout sebesar 0.2 pada setiap hidden layer, output (neuron = 7, dan activation = sigmoid) menghasilkan AUC rata-rata sebesar 0.9887 dan *accuracy_score* sebesar 0.9452. Peneliti juga memvisualisasikan hasil performansi ke dalam Kurva ROC, dan hasilnya tertera pada Gambar 8.



Gambar 8. Visualisasi Kurva ROC

Merujuk kepada standar kategori klasifikasi, maka performa model yang digunakan masuk kedalam klasifikasi sangat baik karena grafik mendekati angka 1. Setelah mendapatkan akurasi yang baik maka tahap selanjutnya yaitu pengimplementasian model untuk memprediksi mahasiswa Angkatan 2018.

Setelah mendapatkan model yang baik, selanjutnya peneliti memprediksi *score* peminatan mahasiswa dengan melakukan *load_model* dari hasil pembentukan algoritma sebelumnya. Hasil *score* prediksi akan digunakan sebagai acuan apakah mahasiswa tersebut dapat diterima pada suatu bidang peminatan.

Setelah mendapatkan nilai prediksi maka tahap selanjutnya peneliti menginputkan daftar dosen pada setiap peminatan untuk dijadikan sebagai kuota perpeminatan. Dikarenakan dosen aktif memiliki kapasitas maksimal untuk menjadi pembimbing 1 atau 2 dari 20 mahasiswa, maka 1 dosen dapat menghasilkan 10 kuota untuk tiap peminatan. Selanjutnya peneliti melakukan sortir data untuk menempatkan mahasiswa ke setiap bidang peminatan sesuai kuota yang tersedia. Berdasarkan hasil pensortiran maka,

diketahui dari total 367 data mahasiswa, 80 mahasiswa masuk ke dalam peminatan EDE, 120 mahasiswa masuk ke dalam peminatan EISD, 38 mahasiswa masuk ke dalam peminatan ERP, 77 mahasiswa masuk ke dalam peminatan SAG, dan 41 mahasiswa masuk ke dalam peminatan EIM.

SIMPULAN

Berdasarkan hasil dan pembahasan diatas, kesimpulan yang dapat diambil dari penelitian ini adalah penggunaan algoritma ANN dengan menginputkan 3 hidden layer (neuron = 128, 64, 42 dan activation = ReLU), dropout sebesar 0.2 pada setiap hidden layer, output (neuron = 7, dan activation = sigmoid) menghasilkan klasifikasi yang sangat baik dengan mendapatkan AUC rata-rata sebesar 0.9887 dan *accuracy_score* sebesar 0.9452. Sistem ini dapat digunakan untuk prediksi bidang peminatan mahasiswa, sehingga dapat membantu menyelesaikan permasalahan mahasiswa untuk mendapatkan bidang peminatan yang sesuai dengan kemampuan dan keahlian mereka. Untuk penelitian selanjutnya diharapkan dapat mempertimbangkan urutan matakuliah prasyarat sebagai pertimbangan penentu bidang peminatan dan status dosen untuk meminimalisir tidak meratanya jumlah pembimbing dikemudian hari.

DAFTAR PUSTAKA

- Erwan. (2019, Juli 2). *Data normalization before or after train-test split?* Retrieved from StackExchange: <https://datascience.stackexchange.com/questions/54908/datanormalization-before-or-after-train-test-split>
- Hassanipour, S., Ghaem, H., Arab-Zozani, M., Seif, M., Fararouei, M., Abdzadeh, E., . . . Paydar, S. (2019). Comparison of artificial neural network and logistic

-
- regression models for prediction of outcomes in trauma patients: A systematic review and meta-analysis. doi: 10.1016/j.injury.2019.01.007. *Injury*.
- Hevner, A. R. (2004). DESIGN SCIENCE IN INFORMATION SYSTEMS RESEARCH. *Management Information Systems Research Center, University of Minnesota*. <https://doi.org/10.2307/25148625>
- Rahman, M., Darmawidjadja, M., & Alamsah, D. (2017). KLASIFIKASI UNTUK DIAGNOSA DIABETES MENGGUNAKAN METODE BAYESIAN REGULARIZATION NEURAL NETWORK (RBNN). *Jurnal Informatika*, <http://dx.doi.org/10.26555/jifo.v11i1.a5452>.
- Wahyono, T. (2018). *Python for Machine Learning (Dasar-dasar Pemrograman Python untuk Machine Learning dan Kecerdasan Buatan)*. Yogyakarta: Gava Media.