

Perbandingan Algoritma K-Nearest Neighbor (K-Nn) Dan Id3 Untuk Prediksi Kelulusan Mahasiswa Fakultas Teknik Universitas Muhammadiyah Jember

Comparison Of K-Nearest Neighbor (K-Nn) And Id3 Algorithm For Graduation Prediction Of Students Faculty Of Engineering, Muhammadiyah Jember University

Iradatul Maulidah¹, Taufiq Timur W^{2*}, Qurrota A'yun³

¹Mahasiswa Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
Email: iradmaulidah@gmail.com

²Fakultas Teknik, Universitas Muhammadiyah Jember* Koresponden Author
Email: taufiqtimur@unmuhjember.ac.id

³Fakultas Teknik, Universitas Muhammadiyah Jember
Email: qurrota.ayun@unmuhjember.ac.id

Abstrak

Mahasiswa adalah elit masyarakat, dan mereka memiliki karakteristik intelektual yang kompleks. Menurut BAN-PT 2011 (Badan Akreditasi Nasional Perguruan Tinggi), salah satu aspek evaluasi akreditasi ialah mahasiswa dan lulusan. Dengan demikian, tingkat kelulusan dan jumlah siswa akan mempengaruhi proses sertifikasi yang dilakukan oleh pemerintah. Salah satu cara untuk mencapai tingkat kualitas tertinggi dalam sistem pendidikan tinggi adalah dengan menggali pengetahuan dari data di bidang pendidikan sebagai atribut pembelajaran utama yang mempengaruhi kinerja siswa. Dalam penelitian ini metode klasifikasinya adalah perbandingan antara algoritma *K-Nearest Neighbor (K-NN)* dan ID3. Hasil yang didapat dari penelitian ini adalah pada algoritma *K-Nearest Neighbor (K-NN)* dengan *Accuracy* sebesar 91.66%, *Sensitivity* sebesar 96.66% dan *Precision* sebesar 87.87%. Pada algoritma ID3 (*Iterative Dichotomiser 3*) didapatkan hasil *Accuracy* sebesar 93.33%, *Sensitivity* sebesar 93.93% dan *Precision* sebesar 93.93%.

Kata Kunci: Mahasiswa, *K-Nearest Neighbor*, ID3

Abstract

*Students are the elite of society, and they have complex intellectual characteristics. According to BAN-PT 2011 (National Accreditation Board for Higher Education), one aspect of accreditation evaluation is students and graduates. Thus, the graduation rate and the number of students will affect the certification process carried out by the government. One way to achieve the highest level of quality in the higher education system is to extract knowledge from data in the field of education as the main learning attribute that affects student performance. In this study, the classification method is a comparison between the *K-Nearest Neighbor (K-NN)* and ID3 algorithms. The results obtained from this study are the *K-Nearest Neighbor (K-NN)* algorithm with an *Accuracy* of 91.66%, *Sensitivity* of 96.66% and *Precision* of 87.87%. In the ID3 algorithm (*Iterative Dichotomiser 3*), the results obtained are 93.33% *Accuracy*, 93.93% *Sensitivity* and 93.93% *Precision*.*

Keywords: Student, *K-Nearest Neighbor*, ID3

1. PENDAHULUAN

Perguruan tinggi adalah suatu pendidikan yang lebih tinggi dari jenjang pendidikan menengah di dalam susunan pendidikan sekolah. Pendidikan tinggi bertujuan untuk melatih mahasiswa menjadi anggota masyarakat dengan keterampilan akademik dan profesional yang dapat mengimplementasikan, mengembangkan dan menciptakan ilmu pengetahuan, teknologi dan seni. Pendidikan tinggi memegang peranan penting dalam membentuk sumber daya manusia yang berkualitas dan siap menghadapi era persaingan global.

Kemudian, untuk memenuhi standar kemahiran lulusan bagi mahasiswa program sarjana (S1), beban yang dipersyaratkan minimal harus 144-160 satuan kredit semester (sks). Masa studi untuk mahasiswa sarjana (S1) biasanya 8 semester atau 4 tahun, tergantung program studinya. Namun kenyataannya tidak sedikit mahasiswa yang melebihi standar kelulusan yang diharapkan selama masa studinya.

Khusus mengenai evaluasi standar mahasiswa dan lulusan, komponen evaluasinya adalah sistem penerimaan mahasiswa baru dan lulusan (rata-rata pembelajaran dan IPK) (Muarif, 2013). Oleh karena itu, tingkat kelulusan dan jumlah siswa akan berdampak pada proses sertifikasi yang dilakukan oleh pemerintah. Salah satu masalah utama perguruan tinggi adalah meningkatkan kualitas pendidikan mahasiswa dan kualitas keputusan manajemen institusi. Dengan demikian, tingkat kelulusan dan jumlah siswa akan mempengaruhi proses sertifikasi yang dilakukan oleh pemerintah. Salah satu cara untuk mencapai tingkat kualitas tertinggi dalam sistem pendidikan tinggi adalah dengan menggali pengetahuan dari data di bidang pendidikan sebagai atribut pembelajaran utama yang mempengaruhi kinerja siswa.

Pada penelitian sebelumnya yang dilakukan terhadap penggunaan Data Mining dalam memprediksi suatu kasus dan memiliki kesamaan pada penelitian yang akan diajarkan

di antaranya adalah pada penelitian yang dilakukan oleh Ardiyansyah, Panny Agustia Rahayuningsih, Reza Maulana (2018) dengan judul “ Analisis Perbandingan Algoritma Klasifikasi Data Mining untuk Blogger dengan Rapid Miner” hasil tertinggi dari nilai akurasi yang didapat adalah sebesar 85.00% untuk algoritma *K-Nearest Neighbor* (KNN), tetapi algoritma ini merupakan algoritma dengan performa kurang baik. Kemudian penelitian yang dilakukan oleh Abdul Rohman (2011) dengan judul “Model algoritma *K-Nearest Neighbor* (KNN) untuk prediksi kelulusan mahasiswa” memperoleh hasil dengan nilai akurasi 85,15%. Sedangkan pada penelitian yang dilakukan oleh Danny Himawan (2014) “Aplikasi Data Mining Menggunakan Algoritma ID3 untuk Mengklasifikasi Kelulusan Mahasiswa Pada Universitas Dian Nuswantoro Semarang” mendapatkan kesimpulan, semakin sedikit data training yang digunakan maka hasil klasifikasinya menunjukkan ketidakakuratan yang tinggi, sebaliknya jika data training yang digunakan semakin banyak hasilnya akan semakin akurat.

Untuk itu dalam penelitian ini akan dilakukan prediksi kelulusan dengan membandingkan algoritma *K-Nearest Neighbor* (*K-NN*) dan ID3.

2. TIANJAUAN PUSTAKA

A. Universitas

Universitas Muhammadiyah Jember ialah salah satu Universitas Persyarikatan Muhammadiyah. Didirikan sesuai dengan Piagam Muhammadiyah untuk Pendirian Perguruan tinggi. Universitas Muhammadiyah Jember saat ini memiliki 8 fakultas di bawah Kementerian Pendidikan Kebudayaan, meliputi 18 program studi sarjana (S1) dan 2 diploma (D-III), 1 program pascasarjana (S2) dan 1 fakultas dengan 1 program studi dibawah Kementrian Agama (Pendidikan Agama Islam), dan satu akademi Pariwisata dengan memiliki program studi Perhotelan D3.

B. Kelulusan Mahasiswa

Menurut BANPT 2011 (Badan Akreditasi Nasional Perguruan Tinggi) tentang standar dan prosedur akreditasi perguruan tinggi dalam buku kedua, salah satu aspek evaluasi akreditasi ialah mahasiswa dan lulusan. Khusus mengenai evaluasi standar mahasiswa dan lulusan, komponen evaluasinya adalah sistem penerimaan mahasiswa baru dan lulusan (rata-rata pembelajaran dan IPK) (Muarif, 2013). Oleh karena itu, tingkat kelulusan dan jumlah siswa akan berdampak pada proses sertifikasi yang dilakukan oleh pemerintah. Salah satu masalah utama perguruan tinggi adalah meningkatkan kualitas pendidikan mahasiswa dan kualitas keputusan manajemen institusi. Dengan demikian, tingkat kelulusan dan jumlah siswa akan mempengaruhi proses sertifikasi yang dilakukan oleh pemerintah. Salah satu cara untuk mencapai tingkat kualitas tertinggi dalam sistem pendidikan tinggi adalah dengan menggali pengetahuan dari data di bidang pendidikan sebagai atribut pembelajaran utama yang mempengaruhi kinerja siswa.

C. Data Mining

Data mining ialah serangkaian proses untuk memperoleh pengetahuan atau pola dari kumpulan data (Ian H. Witten, 2011). Sedangkan menurut Aggarwal (2015), data mining adalah studi mengumpulkan, membersihkan, mengolah, menganalisis dan memperoleh pengetahuan yang berguna dari data. Tujuan dari data mining itu sendiri adalah untuk menemukan hubungan atau pola yang dapat memberikan sinyal yang berguna.

D. Klasifikasi

Salah satu tugas yang dapat dilakukan data mining adalah klasifikasi. Klasifikasi pertama kali diterapkan pada bidang tumbuhan untuk mengklasifikasikan spesies tertentu, seperti yang dilakukan Carolus von Linne (juga dikenal sebagai Carolus Linnaeus), ia adalah orang pertama yang mengklasifikasikan spesies berdasarkan karakteristik fisik. Dia kemudian disebut bapak klasifikasi.

E. K-Nearest Neighbor

Menurut Gorunescu (Badu, 2016), *K-Nearest Neighbor* adalah metode untuk mengklasifikasikan objek baru berdasarkan tetangga terdekat (K). Tetangga dekat atau jauh biasanya dihitung berdasarkan jarak Euclidean, yang diwakili oleh persamaan berikut (Kartika et al., 2017):

$$d(x_i, x_j) = \sqrt{\sum_{n=1}^p (x_{ip} - x_{jp})^2}$$

Keterangan:

x_{ip}	= Data training
x_{jp}	= Data testing
$d(x_i, x_j)$	= jarak euclidean. .

F. Decision Tree

Decision Tree adalah metode klasifikasi dan prediksi yang terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang mewakili aturan. Para peneliti telah mengembangkan berbagai algoritma pohon keputusan untuk meningkatkan kinerja dan kemampuan menangani berbagai jenis data. Beberapa algoritma yang penting tersebut, yaitu Chid, Cart, ID3, C4.5, C5.0, dan Hunt's Algorithm (Hariyanto, 2016). Oleh karena itu pada penelitian ini akan menggunakan algoritma ID3 sebagai pembentukan pohon keputusan nantinya.

G. ID3 (Iterative Dichotomiser 3)

Algoritma ID3 atau Iterative Divider 3 (ID3) adalah sebuah metode untuk membuat pohon keputusan

Langkah kerja algoritma ID3 dapat digambarkan sebagai berikut:

1. Pertama menyiapkan data training
2. Kemudian menghitung nilai Entropy menggunakan rumus :

$$Entropy(s) = -(P_+ \log_2 P_+) - (P_- \log_2 P_-)$$

Keterangan :

S = Data yang digunakan untuk training

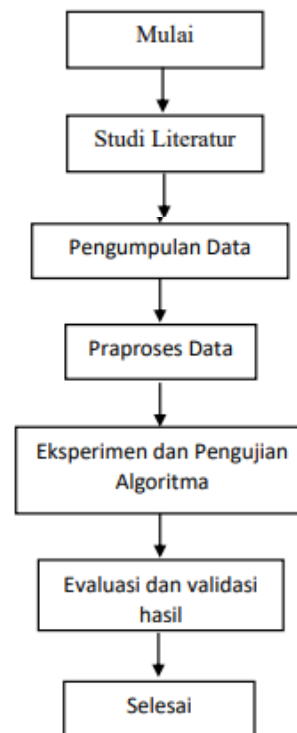
p_+ = Jumlah yang bersolusi positif pada data kriteria tertentu.

p_- = Jumlah yang bersolusi negative (tidak mendukung) pada data sample kriteria tertentu.

1. Setelah nilai Entropy didapat kemudian mencari Information Gain dari setiap atribut untuk mendapatkan nilai yang paling tinggi.
2. Rumus information Gain yaitu :

$$Gain(S,A) = Entropy(S) - \sum_{v \in \text{nilai}(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

S = data sample yang digunakan untuk training
 A = atribut
 V = suatu nilai yang mungkin untuk atribut A
 $\text{Nilai}(A)$ = himpunan yang mungkin untuk atribut A
 $|S_v|$ = Jumlah sample untuk nilai V
 $|S|$ = Jumlah nilai sample data
 $Entropy(S_v)$ = Entropy untuk sample – sample yang memiliki nilai V
3. Nilai dari Information Gain tertinggi akan menjadi simpul yang pertama dan akan menempati paling atas.
4. Proses perhitungan Information Gain diulangi sampai semua data yang masuk dalam kelas yang sama. Sehingga akan membentuk pohon keputusan.



Gambar 1 Kerangka Penelitian
 Sumber: Hasil penelitian sendiri.

3. METODOLOGI PENELITIAN

Kerangka penelitian yang akan digunakan ditunjukkan pada gambar di bawah ini.

Tabel 1 Data Uji Mahasiswa

No	Sex	Age	IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	IPS 6	IPS 7	IPS 8	Pembayaran	Status
1	1	24	3,44	3,43	3,43	3,39	3,4	3,36	3,45	3,48	1	1
2	1	24	3,47	3,29	3,15	3,15	3,18	3,24	3,27	3,31	1	1
3	1	24	3,82	3,83	3,74	3,69	3,66	3,63	3,53	3,54	0	0
4	1	27	3,51	3,49	3,32	3,34	3,36	3,3	3,32	3,33	1	1
5	0	23	3,78	3,84	3,69	3,65	3,63	3,62	3,59	3,5	0	0
6	1	24	3,73	3,74	3,67	3,67	3,67	3,71	3,71	3,71	1	1
7	1	25	3,25	3,12	3,16	3,22	3,24	3,25	3,26	3,15	0	0
8	1	23	3,39	3,42	3,29	3,27	3,01	2,93	2,96	2,83	0	0
9	0	24	3,94	3,89	3,89	3,8	3,76	3,65	3,76	3,76	1	1
10	1	23	3,52	3,12	3,01	2,96	2,99	3,07	3,1	3	0	0
11	1	30	2,15	0	2,14	2,47	2,44	2,65	2,65	2,85	0	0
12	1	23	3,19	3,25	3,23	3,17	3,18	3,17	3,2	3,13	0	0
13	1	23	2,81	0	2,98	2,96	2,8	2,64	2,53	2,53	0	0

14	0	23	3,7	3,73	3,73	3,71	3,65	3,6	3,65	3,67	1	1
15	1	23	3,19	3,19	3,31	3,26	3,3	3,26	3,31	3,21	1	1
16	0	24	3,78	3,65	3,67	3,63	3,6	3,46	3,62	3,64	1	1
17	1	23	3,7	3,72	3,73	3,55	3,54	3,56	3,64	3,66	1	1
18	0	23	3,92	3,89	3,88	3,86	3,86	3,79	3,85	3,85	1	1
19	1	23	3,89	3,84	3,88	3,78	3,77	3,71	3,76	3,77	1	1
20	0	23	3,65	3,65	3,49	3,5	3,49	3,39	3,55	3,55	0	0

Sumber: Hasil penelitian sendiri.

A. Pengumpulan data

Tahapan ini proses pengumpulan data yang didapat dari Universitas Muhammadiyah Jember, yaitu data mahasiswa angkatan 2016. Keseluruhan data yang didapat mempunyai beberapa atribut yaitu Program Studi, Nim, Nama, Umur, Jenis Kelamin dan Index Prestasi Semester 1 sampai dengan semester 8, Pembayaran.

B. Praproses Data

Pada tahapan ini penulis melakukan praproses data sebelum data diolah. Data yang diperoleh yaitu sebanyak 413 data, namun tidak semua data dapat digunakan dan tidak semua atribut digunakan karena harus melalui beberapa tahap pengolahan data. Berikut adalah beberapa teknik yang dilakukan agar mendapat data yang berkualitas, sebagai berikut:

1. Melakukan inisialisasi pada atribut jenis kelamin dengan keterangan laki – laki menjadi (1) dan perempuan menjadi (0).
2. Melakukan inisialisasi pada atribut pembayaran dengan 16 kali (4 tahun standart lulus tepat waktu) pembayaran (1) dan kurang atau lebih dari 16 kali pembayaran.
3. Dalam penelitian ini atribut yang tidak relevan seperti Program Studi, Nim dan nama akan dihapuskan, data set akan memiliki jumlah atribut yang lebih sedikit tetapi bersifat informative.

C. Eksperimen dan Pengujian Algoritma.

Pada proses ini dilakukan eksperimen dengan menggunakan metode klasifikasi data mining K-Nearest Neighbor dan ID3 terhadap data mahasiswa yang terkait dengan kelulusan.

D. Evaluasi dan Validasi Hasil

Setelah melakukan pengolahan data awal, data menjadi 298 mahasiswa dengan jumlah 112 mahasiswa lulus, sedangkan yang belum lulus berjumlah 186 mahasiswa. Atribut yang digunakan dalam eksperimen ini yaitu, Umur, Sex dan IP Semester 1 sampai semester 8. Kemudian untuk proses evaluasi dilakukan dengan Confusion Matrix menghitung True Positive, True Negative, False Positive dan False Negative sehingga menghasilkan *accuracy*, *Sensitivity (recall)*, *PPV (Positive Predictive Value)* atau *Precision*.

4. HASIL DAN PEMBAHASAN

A. Hasil Klasifikasi K-Nearest Neighbor (KNN)

Berikut matrix dan perhitungan *Accuracy*, *recall*, dan *precision* pada K-Nearest Neighbor (KNN) yang diperoleh dari proses klasifikasi pada Python.

Tabel 2 *confusion matrix* pada Algoritma ID3

		Nilai Sebenarnya	
		TRUE	FALSE
Nilai Prediksi	TRUE	29	4
	FALSE	1	26

Sumber: Hasil penelitian sendiri

Perhitungan *Accuracy* :

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{29+26}{29+26+4+1} =$$

$$0.9166 \times 100\% = 91.66\%$$

Perhitungan *Recall*:

$$Recall = \frac{TP}{TP+FN} = \frac{29}{29+1} = 0.96.66 \times 100\% = 96.66\%$$

Perhitungan *Precision* :

$$Precision = \frac{TP}{TP+FP} = \frac{29}{29+4} = 0.8787 \times 100\% = 87.87\%$$

B. Hasil Klasifikasi ID3

Berikut matrix dan perhitungan *Accuracy*, *recall*, dan *precision* pada ID3 (*Iterative Dichotomiser 3*) yang diperoleh dari proses klasifikasi pada Python.

Tabel 3 *confusion matrix* pada Algoritma ID3

		Nilai Sebenarnya	
		TRUE	FALSE
Nilai Prediksi	TRUE	31	2
	FALSE	2	25

Sumber: Hasil penelitian sendiri

Perhitungan *Accuracy* :

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{31+25}{31+25+2+2} = 0.9333 \times 100\% = 93.33\%$$

Perhitungan *Recall* :

$$Recall = \frac{TP}{TP+FN} = \frac{31}{31+2} = 0.9393 \times 100\% = 93.93\%$$

Perhitungan *Precision* :

$$Precision = \frac{TP}{TP+FP} = \frac{31}{31+2} = 0.9393 \times 100\% = 93.93\%$$

Pengujian dilakukan pada algoritma *K-Nearest Neighbor* (KNN) dan ID3 (*Iterative*

Dichotomiser 3) dengan *Confusion Matrix* untuk mengetahui tingkat *Sensitivity (recall)*, *PPV (positive predictive value)* atau *Precision*, dan *Accuracy*.

Tabel 4 Hasil *Confusion Matrix*

	<i>Accurac y</i>	<i>Sensitivit y</i>	<i>Precisio n</i>
K-NN	91.66%	96.66%	87.87%
ID3	93.33%	93.93%	93.93%

Sumber: Hasil penelitian sendiri.

Berdasarkan Tabel diatas dapat diketahui hasil *Accuracy*, *Sensitivity* dan *Precision* dari data set Mahasiswa Fakultas Teknik untuk perbandingan *K-Nearest Neighbor* (KNN) dan ID3 (*Iterative Dichotomiser 3*). sehingga dalam melakukan prediksi kelulusan mahasiswa, bahwa algoritma ID3 (*Iterative Dichotomiser 3*) yang memiliki nilai akurasi 93.33% lebih baik dari pada algoritma *K-Nearest Neighbor* (KNN) yang memiliki nilai akurasi 91.66%.

C. KESIMPULAN DAN SARAN

A. Kesimpulan

Kesimpulan yang di dapat dari hasil penelitian perbandingan algoritma *K-Nearest Neighbor* (KNN) dan ID3 (*Iterative Dichotomiser 3*) untuk prediksi kelulusan mahasiswa fakultas Teknik adalah :

Berdasarkan penelitian yang telah dilakukan menggunakan algoritma *K-Nearest Neighbor* (KNN) menghasilkan *Accuracy* sebesar 91.66%, *recall* sebesar 96.66% dan *precision* sebesar 87.87% sedangkan pada algoritma ID3 (*Iterative Dichotomiser 3*) menghasilkan *Accuracy* sebesar 93.33%, *recall* sebesar 93.93%, dan *precision* sebesar 93.93%. sehingga dari kesimpulan diatas maka Algoritma ID3 (*Iterative Dichotomiser 3*) yang lebih tepat untuk kasus prediksi kelulusan mahasiswa.

B. Saran

Dari hasil pengujian yang telah dilakukan dan hasil kesimpulan yang diberikan maka ada saran atau usul yang di berikan antara lain:

1. Untuk penelitian selanjutnya dapat dibandingkan dengan algoritma lain untuk mendapatkan nilai akurasi yang lebih tinggi, seperti Gaussian Naïve Bayes.
2. Untuk penelitian selanjutnya pada python disarankan data training dan data testing sudah terbagi dengan jelas atau split data tidak random.
3. Dapat menggunakan data yang berbeda atau data training lebih banyak dan menambahkan jumlah kriteria untuk prediksi kelulusan seperti nilai Toefl, asal sekolah dan jalur masuk.

DAFTAR PUSTAKA

- Peraturan pemerintah Republik Indonesia No 30 tahun 1990, pasal 1(1) tentang pendidikan tinggi
- Universitas Muhammadiyah Jember. 2017. Buku pedoman Akademik Universitas Mhammadiyah Jember. Jember. Universitas Muhammadiyah Jember
- Bustami, B. (2014). *Penerapan Algoritma Naïve bayes Untuk Mengklasifikasi Data Nasabah Asuransi*. Jurnal Informatika, 8(1)
- Saxena, K., Khan, Z., & Singh, S. (2014). Diagnosis of Diabetes Mellitus using K Nearest Neighbor Algorithm. *International Journal of Computer Science Trends and Technology (IJCST)*.
- Abu Tair Mohammed M., El-Halees Alaa M. Mining, Educational Data to Improve Students' Performance: A Case Study, 2012, International Journal of Information and Communication Technology Research, Volume 2 No. 2, pp 140-146.
- Muarif, Khoirul. (2013).Komparasi Pemodelan Data Menggunakan C4.5 Dan C4.5 Berbasis Particle Swarm Optimization Untuk Memprediksi Kelulusan Mahasiswa. Semarang :Universitas Dian Nuswantoro.a
- Aggarwal, C. C. (2015). Data Mining. New York: Springer.
- Muzakir, A., & Wulandari, R. A. (2016). Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree. *Scientific Journal of Informatics*, 3(1),19–26. [wihttps://doi.org/10.15294/sji.v3i1.4610](https://doi.org/10.15294/sji.v3i1.4610)
- Rohman, Abdul. 2011. Model algoritma K-Nearest Neighbor (KNN) untuk prediksi kelulusan mahasiswa. Fakultas Teknik Universitas Pandanaran Semarang.
- Himawan, Danny. 2014. Aplikasi Data Mining Menggunakan Algoritma ID3 untuk Mengklasifikasi Kelulusan Mahasiswa Pada Universitas Dian Nuswantoro Semarang.
- Ardiyansyah, dkk. 2018. Analisis Perbandingan Algoritma Klasifikasi Data Mining untuk Blogger dengan Rapid Miner.
- M, Ghada, dkk. 2015. *An Experimental Study of Classification Algorithms for Terrorism Prediction*.
- Aguswira, Budiutama and Susan Dian, Purnamasari (2019) *Sistem Informasi Georafis Pemetaan Perguruan Tinggi Di Kota Palembang*. Diploma thesis, Universitas Bina Darma.
- Witten, Ian H, Frank, Eibe, & Hal, M.A. 2011. Data Mining: Pratical Machine Learning Tools and Techniques, Third Edition. Burlington: Morgan Kaufmann Publishers.
- Adinugroho, S. & Sari, Y., 2018. Implementasi Data Mining Menggunakan WEKA. Malang: UB Press.
- Badu, Z.S., 2016. Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Dana Desa.” (November).
- Kartika, Irjaya, J., Santoso, E. dan Sutrisno, 2017. Penentuan siswa berprestasi menggunakan metode K-Nearest Neighbor dan Weighted Product (Studi Kasus : SMP Negeri 3 Mejayan).” 1(5).
- Haryanto,2016. Pemograman web. Jakarta:Informatika

Han, J, Kamber, M, & Pei, J. 2012. *Data Mining: Concept and Techniques, Third Edition*. Waltham: Morgan Kaufmann Publishers.