

**PEMBANGUNAN MODEL KLASIFIKASI SMOTE - KNN
(*SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE –
K-NEAREST NEIGHBOR*) DALAM MEMPREDIKSI WAKTU
KELULUSAN SISWA *BOOTCAMP* BINAR ACADEMY**

SKRIPSI

Disusun dan Diajukan untuk Memenuhi Salah Satu Syarat Memperoleh Gelar
Sarjana Pendidikan Program Studi Pendidikan Ilmu Komputer



oleh:

Andika Putra Kamula

1805836

**Program Studi Pendidikan Ilmu Komputer
Fakultas Pendidikan Matematika Dan Ilmu Pengetahuan Alam
Departemen Pendidikan Ilmu Komputer
Universitas Pendidikan Indonesia
2022**

**PEMBANGUNAN MODEL KLASIFIKASI SMOTE - KNN
(*SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE –
K-NEAREST NEIGHBOR*) DALAM MEMPREDIKSI WAKTU
KELULUSAN SISWA *BOOTCAMP* BINAR ACADEMY**

oleh

Andika Putra Kamula

Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana pada Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam

© Andika Putra Kamula 2022

Universitas Pendidikan Indonesia

Agustus 2022

Hak cipta dilindungi undang-undang

skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian,
dengan dicetak ulang, difoto kopi, atau cara lainnya tanpa ijin dari penulis

Andika Putra Kamula, 2022

**PEMBANGUNAN MODEL KLASIFIKASI SMOTE - KNN
(*SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE –
K-NEAREST NEIGHBOR*) DALAM MEMPREDIKSI WAKTU KELULUSAN SISWA *BOOTCAMP* BINAR
ACADEMY**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

PEMBANGUNAN MODEL KLASIFIKASI SMOTE - KNN
(*SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE –
K-NEAREST NEIGHBOR*) DALAM MEMPREDIKSI WAKTU
KELULUSAN SISWA *BOOTCAMP* BINAR ACADEMY

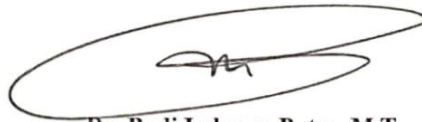
oleh

Andika Putra Kamula

1805836

disetujui dan disahkan oleh:

Pembimbing I



Dr. Budi Laksono Putro, M.T.

NIP. 197607102010121002

Pembimbing II



Erna Piantari, S. Kom., M.T.

NIP. 920171219890224201

Mengetahui,

Ketua Program Studi Pendidikan Ilmu Komputer



Dr. Wahyudin, M.T.

NIP. 197304242008121001

Andika Putra Kamula, 2022

PEMBANGUNAN MODEL KLASIFIKASI SMOTE - KNN
(*SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE –
K-NEAREST NEIGHBOR*) DALAM MEMPREDIKSI WAKTU KELULUSAN SISWA *BOOTCAMP* BINAR
ACADEMY

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

LEMBAR PERYATAAN

Dengan ini saya menyatakan bahwa skripsi dengan judul “Pembangunan Model Klasifikasi SMOTE – KNN (*Synthetic Minority Over-Sampling Technique – K-Nearest Neighbor*) Dalam Memprediksi Waktu Kelulusan Siswa *Bootcamp Binar Academy*” ini beserta seluruh isinya adalah benar-benar hasil karya saya sendiri. Saya tidak melakukan penjiplakan ataupun pengutipan dengan cara-cara yang tidak sesuai dengan etika ilmu yang berlaku dalam masyarakat keilmuan. Atas pernyataan ini, saya siap menanggung risiko/sanksi apabila di kemudian hari ditemukan adanya pelanggaran etika keilmuan atau ada klaim dari pihak lain terhadap keaslian karya saya ini.

Bandung, Agustus 2022

Yang membuat pernyataan,



Andika Putra Kamula

NIM. 1805836

KATA PENGANTAR

Puji syukur ke hadirat Allah SWT yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Pembangunan Model Klasifikasi SMOTE – KNN (*Synthetic Minority Over-Sampling Technique – K-Nearest Neighbor*) Dalam Memprediksi Waktu Kelulusan Siswa *Bootcamp* Binar Academy” dengan baik namun tidak terlepas dari kekurangan.

Skripsi ini ditulis untuk memenuhi sebagian dari syarat memperoleh gelar sarjana Pendidikan Ilmu Komputer Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam Universitas Pendidikan Indonesia.

Dalam proses penyusunan skripsi ini penulis mengalami berbagai kendala, namun atas pertolongan dan Ridha Allah SWT dan bantuan, bimbingan, serta kerja sama dari berbagai pihak kendala tersebut dapat diatasi. Penulis mengucapkan terima kasih kepada pihak yang membantu kelancaran penulisan, termasuk pihak yang telah membagi pengetahuannya sehingga penulis dapat menyelesaikan skripsi ini.

Penulis menyadari skripsi yang ditulis masih jauh dari kata sempurna. Oleh karena itu, kritik dan saran yang membangun penulis nantikan demi perbaikan di masa mendatang. Semoga skripsi ini dapat bermanfaat dan memberikan pembelajaran yang baik untuk penulis dan pembaca khususnya menjadi sumber ilmu pengetahuan.

Bandung, Agustus 2022

Yang membuat pernyataan,



Andika Putra Kamula

NIM. 1805836

UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi ini. Pada penelitian dan penyusunan skripsi ini, penulis banyak mendapat dukungan dan bantuan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih kepada:

1. Orang tua penulis yaitu mama tercinta ibu Nurlaela yang telah mendidik, memberikan doa, dan memberikan dukungan moril dan materil kepada penulis.
2. Bapak Dr. Wahyudin, M.T., selaku Ketua Program Studi Pendidikan Ilmu Komputer FPMIPA Universitas Pendidikan Indonesia.
3. Bapak Erlangga, M.T. selaku pembimbing akademik yang telah membimbing perihal akademis selama perkuliahan.
4. Bapak Dr. Budi Laksono Putro, M.T., selaku dosen pembimbing I yang telah bersedia meluangkan waktu, tenaga, dan pikiran untuk membimbing penulis dalam proses penelitian dan penyusunan skripsi.
5. Ibu Erna Piantari, S. Kom., M.T. selaku dosen pembimbing II yang telah bersedia meluangkan waktu, tenaga, dan pikiran untuk membimbing penulis dalam proses penelitian dan penyusunan skripsi.
6. Bapak dan Ibu Dosen Pendidikan Ilmu Komputer yang telah memberikan ilmu dan pengalamannya yang bermanfaat selama perkuliahan sebagai bekal pengetahuan.
7. Seluruh *employee* Binar Academy terkhusus tim Academic Affairs, Kak Siti, Kak Wiwit, dan Kak Gio, serta rekan-rekan MSIB *bactch* 1 Binar Academy AA Intern, Firdaus, Tanti, Kiky, dan Silvi yang telah membantu penulis dalam melaksanakan penelitian skripsi.
8. Muhammad Arsyal Kyvariwijaya sahabat selama berkuliah di Bandung yang selalu memberikan semangat dan dukungan selama perkuliahan.
9. Rekan-rekan Telur Gulung Sri, Tika, Widi, Kania, dan almh. Alfina.

10. Rekan-rekan pemandu MOKA-KU UPI 2019, KKN SMP Negeri 29 Geger Kalong, PPLS SMK Bina Wisata Lembang.
11. Rekan-rekan organisasi DPM Kemakom FPMIPA dan DPM Rema UPI.
12. Seluruh rekan seperjuangan Pendidikan Ilmu Komputer 2018.
13. Semua pihak yang telah membantu penulis dalam menyelesaikan skripsi ini yang tidak dapat penulis sebutkan satu persatu.

Semoga Allah SWT memberikan balasan yang berlipat atas semua kebaikan dari berbagai pihak yang telah membantu penulis dalam menyelesaikan skripsi ini. Amiin.

Bandung, Agustus 2022

Yang membuat pernyataan,



Andika Putra Kamula

NIM. 1805836

**PEMBANGUNAN MODEL KLASIFIKASI SMOTE - KNN
(*SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE* –
K-NEAREST NEIGHBOR) DALAM MEMPREDIKSI WAKTU
KELULUSAN SISWA *BOOTCAMP* BINAR ACADEMY**

oleh

**Andika Putra Kamula – dikaputra@upi.edu
1805836**

ABSTRAK

Kelulusan tepat waktu yang rendah mempengaruhi kualitas lulusan Binar Academy, maka perlu dilakukan pemantauan oleh tim *Academy Affairs* (AA) terhadap perjalanan *bootcamp* siswa. Salah satu cara adalah dengan prediksi. *Educational Data Mining* memiliki banyak teknik yang dapat digunakan dalam memproses data menjadi informasi, salah satunya adalah algoritma klasifikasi yaitu *K-Nearest Neighbor* (KNN). KNN merupakan salah satu teknik yang sering digunakan dari berbagai teknik data mining dengan salah satu alasannya yaitu mudah dipahami. Namun KNN masih memiliki beberapa kekurangan seperti kemampuan prediksi yang dipengaruhi oleh penentuan nilai K, metode penentuan jarak, serta data yang tidak seimbang. Pada penelitian yang dilakukan ini, ditemukan bahwa akurasi tertinggi 86.36% dengan K=1 menggunakan Eulidean Distanse setelah datanya dilakukan metode oversampling dengan metode SMOTE. Kata Kunci: *Educational Data Mining*, KNN, Prediksi, SMOTE, Binar Academy

**PEMBANGUNAN MODEL KLASIFIKASI SMOTE - KNN
(SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE –
K-NEAREST NEIGHBOR) DALAM MEMPREDIKSI WAKTU
KELULUSAN SISWA BOOTCAMP BINAR ACADEMY**

oleh

Andika Putra Kamula – dikaputra@upi.edu

1805836

ABSTRACT

Low on-time graduation affects the quality of Binar Academy graduates. Therefore, it is necessary for the Academy Affairs (AA) team to monitor students' bootcamp journey, which one of the methods is by prediction. Educational Data Mining has many techniques that can be used to process data into information, one of which is a classification algorithm, namely K-Nearest Neighbor (KNN). KNN is one of the various data mining techniques that is often used with one reason being that it is easy to understand. However, KNN still has several shortcomings, such as the predictive capability that is influenced by the determination of the value of K, the method of determining the distance, and the unbalanced data. This study found that the highest accuracy was 86.36% with K=1 by using the Eulidean Distanse after the data was oversampled through the SMOTE method.

Keywords: Educational Data Mining, KNN, Prediction, SMOTE, Binar Academy

DAFTAR ISI

LEMBAR PERYATAAN.....	iv
KATA PENGANTAR	v
UCAPAN TERIMA KASIH.....	vi
ABSTRAK.....	viii
ABSTRACT.....	ix
DAFTAR ISI.....	x
DAFTAR GAMBAR	xii
DAFTAR TABEL.....	xiii
DAFTAR RUMUS	xiv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Tujuan Penelitian	4
1.4. Manfaat Penelitian	4
1.5. Batasan Masalah	4
1.6. Sistematika Pelaporan Skripsi.....	4
BAB II KAJIAN PUSTAKA.....	7
2.1. Peta Literatur.....	7
2.2. Data Mining	8
2.3. Algoritma Klasifikasi.....	11
2.3.1. Algoritma K-Nearest Neighbor (KNN)	11
2.3.2. Tahapan Algoritma <i>K-Nearest Neighbor</i> (KNN).....	12
2.3.3. Perhitungan Jarak Tetangga.....	13
2.4. Synthetic Minority Oversampling Technique (SMOTE).....	13
2.5. Binar Academy	15
2.6. Smart Learning Environment Establishment Guideline (SLEEG)	16

Andika Putra Kamula, 2022

**PEMBANGUNAN MODEL KLASIFIKASI SMOTE - KNN
(SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE -
K-NEAREST NEIGHBOR) DALAM MEMPREDIKSI WAKTU KELULUSAN SISWA BOOTCAMP BINAR
ACADEMY**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

2.7. State of The Art.....	17
BAB III METODOLOGI PENELITIAN	20
3.1. Analyze	21
3.2. Design	21
3.3. Develop.....	22
3.4. Implement	23
3.5. Evaluate	23
BAB IV HASIL DAN PEMBAHASAN	25
4.1. Analyze	25
4.2. Design	25
4.2.1. Flowchart	26
4.2.2. Entity Relationship Diagram (ERD)	27
4.2.3. Storyboard.....	29
4.3. Develop.....	33
4.3.1. Data Pre-Processing.....	33
4.3.2. Pembangunan Model KNN Tanpa SMOTE	36
4.3.3. Pembangunan Model KNN dengan SMOTE.....	52
4.4. Implement	66
4.5. Evaluate	70
BAB V SIMPULAN DAN SARAN.....	73
5.1. Simpulan	73
5.2. Saran	74
DAFTAR PUSTAKA	76
LAMPIRAN.....	83

DAFTAR GAMBAR

Gambar 2. 1 Peta Literatur	7
Gambar 2. 2 Alur Umum Data Mining	9
Gambar 2. 3 Kategori Educational Data Mining (EDM)	10
Gambar 2. 4 Visualisasi KNN.....	12
Gambar 2. 5 Undersampling dan Oversampling Data	14
Gambar 2. 6 Synthetic Minority Oversampling Technique (SMOTE).....	15
Gambar 2. 7 Tahapan Umum Metodologi SLEEG	17
Gambar 3. 1 Alur Metodologi Penelitian	20
Gambar 4. 1 Flowchart Sistem.....	26
Gambar 4. 2 Alur Database dengan ERD.....	27
Gambar 4. 3 Halaman Login User.....	29
Gambar 4. 4 Halaman Dashboard User.....	29
Gambar 4. 5 Halaman Read Variabel	30
Gambar 4. 6 Halaman Read Sampel	30
Gambar 4. 7 Halaman Penhujian Data Testing.....	31
Gambar 4. 8 Halaman Read Hasil Pengujian.....	32
Gambar 4. 9 Perbandingan Nilai Siswa Chapter 1-11	34
Gambar 4. 10 Penilaian Siswa Terhadap Materi dan Fasilitator.....	35
Gambar 4. 11 Perbandingan Keterangan Lulus Tepat Waktu dan Tidak Tepat Waktu...	36
Gambar 4. 12 Cara Import Dataset pada Rapid Miner	37
Gambar 4. 13 Cara Merubah Role Label Pada Rapid Miner.....	37
Gambar 4. 14 Cara Memilih Lokasi Menyimpan Dataset.....	38
Gambar 4. 15 Keterangan Import Dataset Berhasil.....	38
Gambar 4. 16 Mengecek Missing Data Pada Rapid Miner	38
Gambar 4. 17 Menu Kerja – Proses Desain Alur Algoritma	39
Gambar 4. 18 Lokasi Dataset Pada Rapid Miner	39
Gambar 4. 19 Lokasi Operators pada Rapid Miner.....	39
Gambar 4. 20 Cara Melakukan Split Data – Merubah Rasio	40
Gambar 4. 21 Lokasi Merubah Parameters Pada Algoritma	40
Gambar 4. 22 Lokasi Tombol Play Pada Rapid Miner	40
Gambar 4. 23 Hasil Desain Pada Rapid Miner Tanpa SMOTE.....	41

Andika Putra Kamula, 2022

**PEMBANGUNAN MODEL KLASIFIKASI SMOTE - KNN
(SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE -
K-NEAREST NEIGHBOR) DALAM MEMPREDIKSI WAKTU KELULUSAN SISWA BOOTCAMP BINAR
ACADEMY**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

Gambar 4. 24 Tingkat Accuracy Euclidean dan Manhattan Tanpa SMOTE.....	51
Gambar 4. 25 Tingkat Precision Euclidean dan Manhattan Tanpa SMOTE	51
Gambar 4. 26 Tingkat Recall Euclidean dan Manhattan Tanpa SMOTE	52
Gambar 4. 27 Dataset Sebelum Implementasi SMOTE.....	53
Gambar 4. 28 Penambahan Operator SMOTE Pada Desain	53
Gambar 4. 29 Dataset Setelah Implementasi SMOTE.....	54
Gambar 4. 30 Desain Keseluruhan Dengan Metode SMOTE	54
Gambar 4. 31 Tingkat Accuracy Euclidean dan Manhattan Dengan SMOTE	65
Gambar 4. 32 Tingkat Precision Euclidean dan Manhattan Dengan SMOTE	65
Gambar 4. 33 Tingkat Recall Euclidean dan Manhattan Dengan SMOTE.....	66
Gambar 4. 34 Tambah Variabel Pada Sistem	67
Gambar 4. 35 Halaman Variabel.....	67
Gambar 4. 36 Tambah Sampel Pada Sistem.....	68
Gambar 4. 37 Halaman Sampel	68
Gambar 4. 38 Halaman Testing	68
Gambar 4. 39 Hasil Perhitungan – Hasil Perhitungan Eulidean	69
Gambar 4. 40 Hasil Perhitungan – Pengurutan Jarak Terkecil	69
Gambar 4. 41 Grafik Jarak K.....	69
Gambar 4. 42 Accuracy Model Euclidean Distance Tanpa SMOTE dan SMOTE.....	70
Gambar 4. 43 Accuracy Model Manhattan Distance Tanpa SMOTE dan SMOTE.....	70
Gambar 4. 44 Precision Model Euclidean Distance Tanpa SMOTE dan SMOTE.....	71
Gambar 4. 45 Precision Model Manhattan Distance Tanpa SMOTE dan SMOTE.....	71
Gambar 4. 46 Recall Model Euclidean Distance Tanpa SMOTE dan SMOTE	72
Gambar 4. 47 Recall Model Manhattan Distance Tanpa SMOTE dan SMOTE	72
Gambar 4. 48 Perbandingan Data Aktual dan Prediksi.....	73

DAFTAR TABEL

Tabel 3. 1 Confusion Matrix.....	24
Tabel 4. 1 Atribut Data Hasil Wawancara.....	25
Tabel 4. 2 Hasil Prediksi Tanpa SMOTE Euclidean K= 1	41
Tabel 4. 3 Hasil Prediksi Tanpa SMOTE Euclidean K= 3	42
Tabel 4. 4 Hasil Prediksi Tanpa SMOTE Euclidean K= 5	43
Tabel 4. 5 Hasil Prediksi Tanpa SMOTE Euclidean K= 7	44
Tabel 4. 6 Hasil Prediksi Tanpa SMOTE Euclidean K= 9.....	45
Tabel 4. 7 Hasil Prediksi Tanpa SMOTE Manhattan K= 1	46
Tabel 4. 8 Hasil Prediksi Tanpa SMOTE Manhattan K= 3.....	47
Tabel 4. 9 Hasil Prediksi Tanpa SMOTE Manhattan K= 5	48
Tabel 4. 10 Hasil Prediksi Tanpa SMOTE Manhattan K= 7	49
Tabel 4. 11 Hasil Prediksi Tanpa SMOTE Manhattan K= 9.....	50
Tabel 4. 12 Hasil Prediksi Dengan SMOTE Euclidean K= 1	55
Tabel 4. 13 Hasil Prediksi Dengan SMOTE Euclidean K= 3.....	56
Tabel 4. 14 Hasil Prediksi Dengan SMOTE Euclidean K= 5.....	57
Tabel 4. 15 Hasil Prediksi Dengan SMOTE Euclidean K= 7.....	58
Tabel 4. 16 Hasil Prediksi Dengan SMOTE Euclidean K= 9.....	59
Tabel 4. 17 Hasil Prediksi Dengan SMOTE Manhattan K= 1.....	60
Tabel 4. 18 Hasil Prediksi Dengan SMOTE Manhattan K= 3.....	61
Tabel 4. 19 Hasil Prediksi Dengan SMOTE Manhattan K= 5.....	62
Tabel 4. 20 Hasil Prediksi Dengan SMOTE Manhattan K= 7.....	63
Tabel 4. 21 Hasil Prediksi Dengan SMOTE Manhattan K= 9.....	64
Tabel 4. 22 Cofussion Matrix – Hasil Prediksi	73

DAFTAR RUMUS

Rumus 2. 1 Perhitungan Jarak Euclidean	13
Rumus 2. 2 Perhitungan Jarak Manhattan	13
Rumus 2. 3 Perhitungan Jarak Minkowski	13
Rumus 3. 1 Mengitung Nilai Accuracy Model	24
Rumus 3. 2 Mengitung Nilai Precision Model	24
Rumus 3. 3 Mengitung Nilai Recall Model.....	24

Andika Putra Kamula, 2022

**PEMBANGUNAN MODEL KLASIFIKASI SMOTE - KNN
(SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE -
K-NEAREST NEIGHBOR) DALAM MEMPREDIKSI WAKTU KELULUSAN SISWA BOOTCAMP BINAR
ACADEMY**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

DAFTAR PUSTAKA

- Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, 37(April 2018), 13–49. doi: 10.1016/j.tele.2019.01.007
- Aldowah, H., Al-samarraie, H., & Mohamad, W. (2019). Telematics and Informatics Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, 37(January), 13–49. doi: 10.1016/j.tele.2019.01.007
- Algarni, A. (2016). *Data Mining in Education*. 7(6).
- Bayer, J., Byd̃, H., & Jan, G. (2012). *Predicting drop-out from social behaviour of students*. *Dm*, 103–109.
- Binar Academy. (2017). Binar Bootcamp Welcome Kit. *Binar Academy*.
- Cunningham, P., & Delany, S. J. (2007). *k -Nearest Neighbour Classifiers*. April. doi: 10.1145/3459665
- Dabhade, P., Agarwal, R., Alameen, K. P., Fathima, A. T., Sridharan, R., & Gopakumar, G. (2021). Materials Today: Proceedings Educational data mining for predicting students' academic performance using machine learning algorithms. *Materials Today: Proceedings*, xxxx. doi: 10.1016/j.matpr.2021.05.646
- Date, P. U. B., Type, P. U. B., & Price, E. (2002). *Data Mining and Knowledge Management in Higher Education*.
- Gao, X., & Li, G. (2020). *A KNN Model Based on Manhattan Distance to Identify the SNARE Proteins*. 8.
- Gou, J., Ma, H., Ou, W., Zeng, S., Rao, Y., & Yang, H. (2018). A generalized mean distance-based k-nearest neighbor classifie. *Expert Systems With Applications*. doi: 10.1016/j.eswa.2018.08.021
- Han, H., Wang, W., & Mao, B. (2005). *Borderline-SMOTE : A New Over-Sampling*

Method in. 878–887.

- Hidayati, N., & Hermawan, A. (2021). *K-Nearest Neighbor (K-NN) algorithm with Euclidean and Manhattan in classification of student graduation.* 2(2), 86–91.
- Huang, C. N., Tsai, P., Hsu, C., & Pan, R. (2006). *Exploring Cognitive Difference in instructional outcomes using Text mining technology.* 2116–2120.
- Jha, M. (2019a). *Smart Intelligent Computing and Applications* (Vol. 104). doi: 10.1007/978-981-13-1921-1
- Jha, M. (2019b). *Smart Intelligent Computing and Applications* (Vol. 104). doi: 10.1007/978-981-13-1921-1
- Kesuma, R., & Akbar, H. (2020). *Algoritma K-Nearest Neighbor dengan Euclidean Distance dan Manhattan Distance untuk Klasifikasi Transportasi Bus.* 12(2), 104–111.
- Kularbphettong, K., & Tongsir, C. (2014). *Mining Educational Data to Support Students ' Major Selection.* 8(1), 21–23.
- Luque, A., Carrasco, A., Martín, A., & De, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, 91, 216–231. doi: 10.1016/j.patcog.2019.02.023
- Maciejewski, T., & Stefanowski, J. (2011). *Local Neighbourhood Extension of SMOTE for Mining Imbalanced Data.*
- Malini, J., & Kalpana, Y. (2021). Materials Today : Proceedings Investigation of factors affecting student performance evaluation using education materials data mining technique. *Materials Today: Proceedings*, xxxx. doi: 10.1016/j.matpr.2021.05.026
- Mesquita, D. P. P., Gomes, J. P. P., Souza, A. H., & Nobre, J. S. (2017). Euclidean distance estimation in incomplete datasets. *Neurocomputing*, 0, 1–8. doi: 10.1016/j.neucom.2016.12.081
- Mohammed, R., Rawashdeh, J., & Abdullah, M. (2020). *Machine Learning with*

Oversampling and Undersampling Techniques: Overview Study and Experimental Results. May.

Muliono, R., Lubis, J. H., & Khairina, N. (2020). Analysis K-Nearest Neighbor Algorithm for Improving Prediction Student Graduation Time. *Sinkron*, 4(2), 42. doi: 10.33395/sinkron.v4i2.10480

Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). *Supervised Machine Learning Algorithms: Classification and Comparison.* July. doi: 10.14445/22312803/IJCTT-V48P126

Pena-Ayala, A. (2013). *Expert Systems with Applications Educational data mining: A survey and a data mining-based analysis of recent works.* September. doi: 10.1016/j.eswa.2013.08.042

Qiong, G. U., Xian-ming, W., Zhao, W. U., Bing, N., & Chun-sheng, X. I. N. (2016). *An Improved SMOTE Algorithm Based on Genetic Algorithm for Imbalanced Data Classification.* 14(2).

Rahayuningsih, P. A., & Maulana, R. (2018). *Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapid Miner.* VI(1), 20–28.

Romero, C. (2010). *Educational Data Mining: A Review of the State of the Art.* 40(6), 601–618.

Rosmansyah, Y., Putro, B. L., Putri, A., & Utomo, N. B. (2022). *A simple model of smart learning environment.* doi: 10.1080/10494820.2021.2020295

Salim, P. (2020). *Time Series Prediction on College Graduation Using KNN Algorithm.*

Sembiring, S., Embong, A., Mohamad, M. A., & Furqan, M. (2009). *by An Application of Data Mining Techniques.* 390–394.

Setiawan, A., Matematika, D., Sains, F., Satya, U. K., & Tengah, J. (2022). *Perbandingan Penggunaan Jarak Manhattan , Jarak Euclid , dan Jarak*

Minkowski dalam Klasifikasi Menggunakan Metode KNN pada Data Iris. 5(1), 28–37.

Setiyorini, T., Asmono, R. T., Informatika, T., & Informatika, T. (2018). *KOMPARASI METODE DECISION TREE , NAIVE BAYES DAN K-NEAREST NEIGHBOR PADA KLASIFIKASI KINERJA SISWA.* 15(2), 85–92.

Siringoringo, R. (2018). *KLASIFIKASI DATA TIDAK SEIMBANG MENGGUNAKAN ALGORITMA SMOTE DAN k-NEAREST NEIGHBOR.* 3(1), 44–49.

Vimalraj, S., & Porkodi R. (2018). A Review on Handling Imbalanced Data. *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, 1–11.