

Perbandingan Unjuk Kerja Algoritme Klasifikasi Data Mining dalam Sistem Peringatan Dini Ketepatan Waktu Studi Mahasiswa

Performance Comparison of Data Mining Classification Algorithms for Early Warning System of Students Graduation Timeliness

Ari Fadli^{*}, Mulki Indana Zulfa, Yogi Ramadhani

Program Studi Teknik Elektro, Fakultas Teknik, Universitas Jenderal Soedirman
Jl. Mayjen Sungkono KM 05 Blater Purbalingga 53371

Cara sitasi: A. Fadli, M. I. Zulfa, and Y. Ramadhani, "Perbandingan Unjuk Kerja Algoritme Klasifikasi Data Mining dalam Sistem Peringatan Dini Ketepatan Waktu Studi Mahasiswa," *Jurnal Teknologi dan Sistem Komputer*, vol. 6, no. 4, pp. 158-163, Oct. 2018. doi: 10.14710/jtsiskom.6.4.2018.158-163, [Online].

Abstract – Observation of growing academic data can be carried using data mining methods, for example, to obtain knowledge related to the determinants of timeliness of students graduation. This study conducted a performance comparison of the classification algorithms using decision tree (DT), support vector machine (SVM), and artificial neural network (ANN). This study used students academic data from Faculty of Engineering, Universitas Jenderal Soedirman in the 2014/2015 odd semester until the 2017/2018 odd semester and the attributes that conform to the academic regulations. The analytical method used is CRISP-DM. The results showed that SVM provided the best performance in an accuracy of 90.55% and AUC of 0.959, compared to other algorithms. A Model with SVM algorithm can be implemented in an early warning system for timeliness of student graduation.

Keywords – graduation timeliness; data mining; data classification; data mining algorithms comparison

Abstrak – Observasi terhadap data akademik yang bertumbuh dapat dilakukan menggunakan metode data mining, misalnya untuk memperoleh pengetahuan terkait faktor penentu ketepatan studi mahasiswa. Penelitian ini melakukan perbandingan kinerja algoritme klasifikasi decision tree, support vector machine (SVM), dan artificial neural network (ANN). Penelitian ini menggunakan data akademik mahasiswa Fakultas Teknik Universitas Jenderal Soedirman di semester gasal 2014/2015 hingga semester gasal 2017/2018 dan atribut-atribut yang mengacu pada peraturan akademik yang berlaku. Metode analisis yang digunakan adalah CRISP-DM. Hasil penelitian menunjukkan bahwa SVM memberikan unjuk kerja terbaik, yaitu akurasi 90,55% dan AUC 0,959, dibandingkan dengan algoritme lain. Model dengan algoritme SVM dapat

digunakan dalam sistem peringatan dini ketepatan studi mahasiswa.

Kata Kunci – ketepatan waktu studi; data mining; klasifikasi data; perbandingan algoritme data mining

I. PENDAHULUAN

Akreditasi perguruan tinggi sebagai satuan penyelenggara pendidikan tinggi menyediakan metode evaluasi terhadap institusi perguruan tinggi dalam mengukur, menetapkan mutu dan kelayakan programnya. Salah satu elemen penilaian akreditasi adalah ketepatan masa studi mahasiswa merupakan salah satu elemen penilaian akreditasi dan tertuang pada matriks penilaian instrumen akreditasi program studi dan perguruan tinggi oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) [1]. Waktu tempuh studi mahasiswa rata-rata pada sebuah program pendidikan S1 adalah 4 tahun. Beberapa mahasiswa memiliki waktu tempuh studi lebih dari 4 tahun. Masa studi yang tidak terpantau dengan baik mengakibatkan minimnya jumlah mahasiswa yang lulus tepat waktu [2], [3].

Data mining untuk mengumpulkan dan menggunakan data historis untuk menemukan pola waktu tempuh studi dapat digunakan sehingga pola [4]. Kajian data mining untuk menganalisis data pendidikan dengan menggunakan educational data mining (EDM) menghasilkan basis data perguruan tinggi, berupa data pribadi maupun akademik mahasiswa. Hasil EDM ini memberikan manfaat bagi para pemangku kebijakan dalam proses pengambilan keputusan [5]-[7].

Karamouiz dan Vretoz [8] melakukan studi tentang indikator efektivitas lembaga pendidikan berdasarkan tingkat kelulusan siswa menggunakan artificial neural network (ANN). Prediksi tingkat kelulusan mahasiswa telah dilakukan menggunakan naive bayes dalam [9]-[11], C4.5 dalam [9], k-nearest neighbors (kNN) [12], algoritme apriori dalam [13], dan ANN [14]. Dengan beragam data kelulusan mahasiswa, akurasi prediksi naive bayes mencapai 80,85% [9] dan 80,72% [10], akurasi C4.5 sebesar 85,7% [9], akurasi kNN 80% [12].

^{*}) Penulis korespondensi (Ari Fadli)
Email: arifadli@unsoed.ac.id

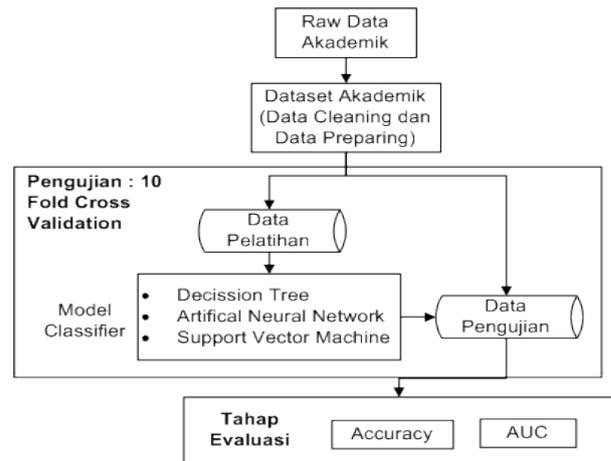
Tingkat akurasi prediksi di atas menggunakan data yang beragam. Perbandingan beragam algoritme prediksi dengan menggunakan data yang sama dilakukan dalam [15]. Algoritme decision tree, naïve bayes, ANN, support vector machine (SVM), dan logistic regression (LR) digunakan untuk memprediksi tingkat kelulusan mahasiswa berdasarkan data dalam satu tahun ajaran pertama.

Berbeda dengan [15] yang menggunakan data dalam satu tahun ajaran pertama, penelitian ini melakukan studi penerapan algoritme decision tree, ANN, dan SVM untuk prediksi tingkat kelulusan mahasiswa menggunakan data mahasiswa dalam 4 tahun terakhir. Penelitian ini menggunakan atribut IP semester tahap evaluasi, IPK akhir semester, jumlah matakuliah dan SKS yang diulang, nilai PKL, KKN, seminar, ujian pendadaran, publikasi artikel ilmiah, dan jumlah nilai E. Penelitian ini juga mengembangkan purwarupa sistem peringatan dini ketepatan waktu studi mahasiswa menggunakan algoritme-algoritme data mining ini yang dapat digunakan untuk meningkatkan rasio ketepatan waktu studi mahasiswa, melakukan pemantauan masa studi mahasiswa, dan memetakan mahasiswa yang berpotensi mengalami keterlambatan masa studi.

II. METODE PENELITIAN

Alur penelitian ini mengacu pada kerangka penelitian seperti Gambar 1. Penelitian menggunakan data akademik Fakultas Teknik Universitas Jenderal Soedirman (FT Unsoed), tahun ajaran 2014/2015 semester gasal sampai dengan 2017/2018 semester gasal. Data akademik tersebut memiliki atribut nim, nama mhs, kodenim, prodi, fakultas, tahunangkatan, kodekul, nama mata kuliah, sks, sks tatap muka, sks praktikum, nilai huruf, bobot nilai, matakuliah b/u, tahun kurikulum, nip dosen pengampu, nama dosen pengampu, kelas, tahunakademik, status KRS, jenis kelamin.

Untuk mengukur kinerja algoritme berdasarkan akurasi, penelitian ini menggunakan model analisis CRISP-DM sebagai standar model proses untuk data mining [16]. Tahap pemahaman bisnis mengacu pada standard evaluasi yang diberikan kepada mahasiswa berdasarkan peraturan akademik yang berlaku di FT Unsoed. Evaluasi dua tahun pertama mahasiswa diperbolehkan melanjutkan studi apabila telah memperoleh SKS ≥ 40 dengan IPK ≥ 2.00 . Evaluasi dua tahun kedua mahasiswa diperbolehkan melanjutkan studi apabila telah memperoleh lebih dari SKS ≥ 80 dengan IPK ≥ 2.00 . Evaluasi keberhasilan studi pada akhir program studi sarjana mahasiswa dinyatakan telah menyelesaikan studi program sarjana apabila telah mengumpulkan 144-160 SKS, sesuai dengan ketentuan fakultas masing-masing dengan persyaratan berikut: a) indeks prestasi kumulatif sekurang-kurangnya 2.00 dan tidak ada nilai E, b) telah berhasil melaksanakan studi akhir yaitu PKL, KKN, skripsi, seminar, ujian pendadaran, dan c) telah mengunggah artikel ilmiah ke e-SIA.



Gambar 1. Alur penelitian

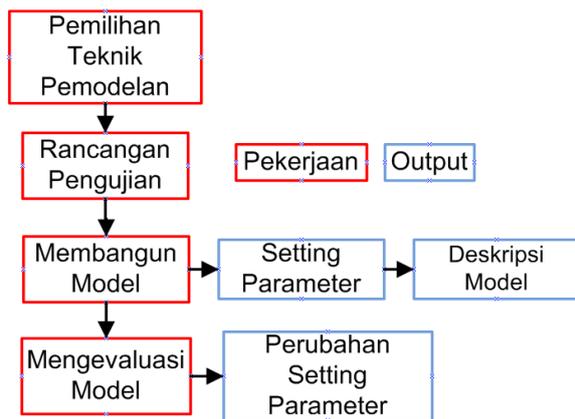
Tabel 1. Atribut data mentah

No	Nama Atribut	Tipe Data
1	No	Integer
2	NIM	Varchar
3	Nama	Varchar
4	Kode Nim	Varchar
5	Prodi	Varchar
6	Fakultas	Varchar
7	Tahun Angkatan	Varchar
8	Kodekul	Varchar
9	Nama Mata Kuliah	Varchar
10	SKS	Integer
11	Bobot Nilai	Integer
12	B/U	Varchar
13	Tahun Akademik	Varchar
14	Ujian Pendadaran	Boolean
15	Artikel Ilmiah	Boolean
16	Jumlah MK Ulang	Integer
17	Jumlah SKS MK Ulang	Integer

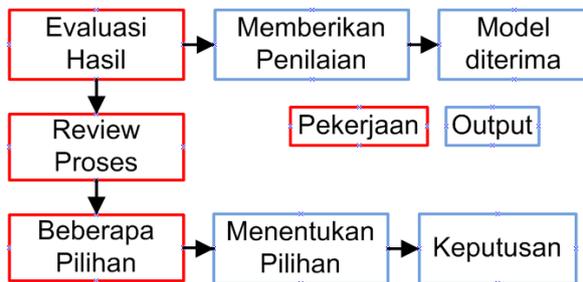
Tabel 2. Atribut dataset

No	Nama Atribut	Tipe Data
1	NIM	Integer
2	IPK1	Double
3	IPK2	Double
4	IPK3	Double
5	SKS1	Double
6	SKS2	Double
7	SKS3	Double
8	PKL	Integer
9	KKN	Integer
10	SEMINAR	Integer
11	Ujian Pendadaran	Integer
12	Artikel Ilmiah	Integer
13	Jumlah MK Ulang	Integer
14	Jumlah SKS MK Ulang	Integer

Tahap pemahaman data dilakukan dengan pengumpulan data akademik 2014/2015 semester gasal sampai dengan 2017/2018 semester gasal. Penanganan missing value sebagai kondisi data tidak lengkap



Gambar 2. Tahapan dalam pemodelan



Gambar 3 Tahapan dalam evaluasi model

dilakukan dengan menghapus data kosong. Hasil pemahaman data adalah berupa atribut dari basis data akademik (Tabel 1). Tahap persiapan data menghasilkan dataset yang menjadi masukan algoritme di tahap pemodelan (Tabel 2).

Pemodelan sistem meliputi 4 langkah (Gambar 2), yaitu pemilihan teknik pemodelan, rancangan pengujian, pembangunan dan evaluasi model. Pada rancangan pengujian ditentukan terkait jumlah data yang digunakan pada tahap pelatihan dan pengujian serta penentuan nilai parameter yang ada pada setiap model yang akan digunakan. Model dibangun berdasarkan algoritme data mining yaitu decision tree, ANN, dan SVM dengan memanfaatkan perangkat Rapid Miner. Evaluasi model dilakukan dengan menggunakan metode validasi 10-Fold cross validation. Tahap evaluasi dilakukan setelah tahap pemodelan (Gambar 3) untuk menginterpretasikan terhadap hasil data mining pada proses pemodelan fase sebelumnya. Evaluasi ini dilakukan untuk menyesuaikan model. Dalam penelitian ini, tahap penyebaran, pemantauan dan pemeliharaan sistem tidak dilakukan.

III. HASIL DAN PEMBAHASAN

Implementasi model dilakukan menggunakan perangkat Rapid Miner. Model klasifikasi yang dirancang adalah model SVM (Gambar 4), ANN (Gambar 5), dan model decision tree (Gambar 6). Nilai akurasi algoritme terlihat dari matrik confusion dalam

Tabel 3. Model matriks *confusion* algoritme SVM

Kelas Prediksi	True Ya	True Tidak	Class Precision
Prediksi Ya	51	5	91.07 %
Prediksi Tidak	8	75	90.36 %
Class Recall	86.44%	93.75%	

Tabel 4. Model matriks *confusion* algoritme ANN

Kelas Prediksi	True Ya	True Tidak	Class Precision
Prediksi Ya	51	8	86.44
Prediksi Tidak	8	72	90.00
Class Recall	86.44	90.00	

Tabel 5. Model matriks *confusion* algoritme *decision tree*

Kelas Prediksi	True Ya	True Tidak	Class Precision
Prediksi Ya	50	8	82.61
Prediksi Tidak	9	72	88.89
Class Recall	84.75	90.00	

Tabel 6. Hasil komparasi algoritme data mining

Parameter	Decision Tree	ANN	SVM
Accuracy	87.64%	88.35%	90.55%
AUC	0.886	0.934	0.959

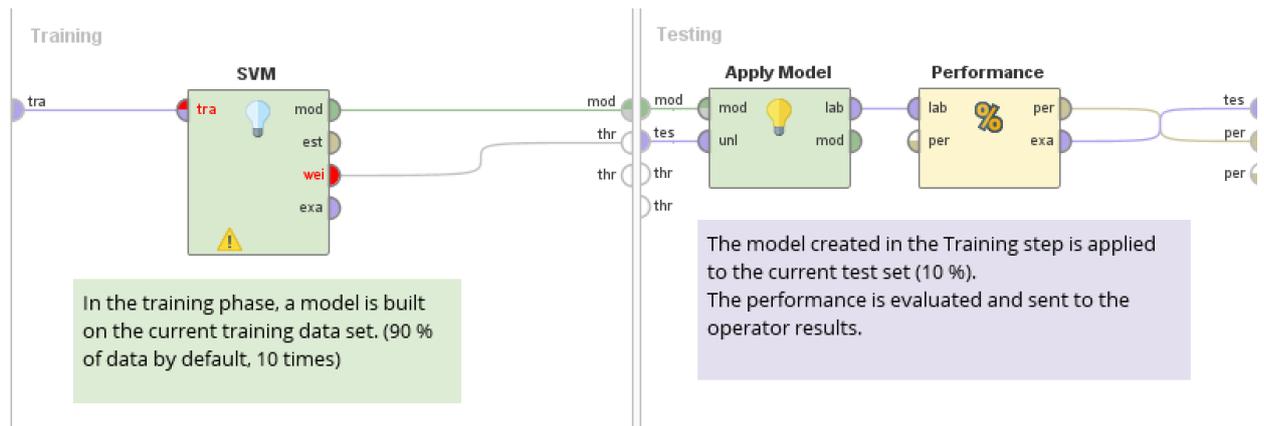
Tabel 3 (SVM), Tabel 4 (ANN), dan Tabel 5 (decision tree).

Matriks confusion dari algoritme SVM (Tabel 3) menunjukkan bahwa dari 139 data menghasilkan presisi kelas 91,07% untuk prediksi Ya dan 90,36% untuk prediksi Tidak dengan class recall sebesar 86,44% dan 93,75%. Akurasi algoritme ini diperoleh sebesar 90,55% ± 6,75% dan mikro 90,65%.

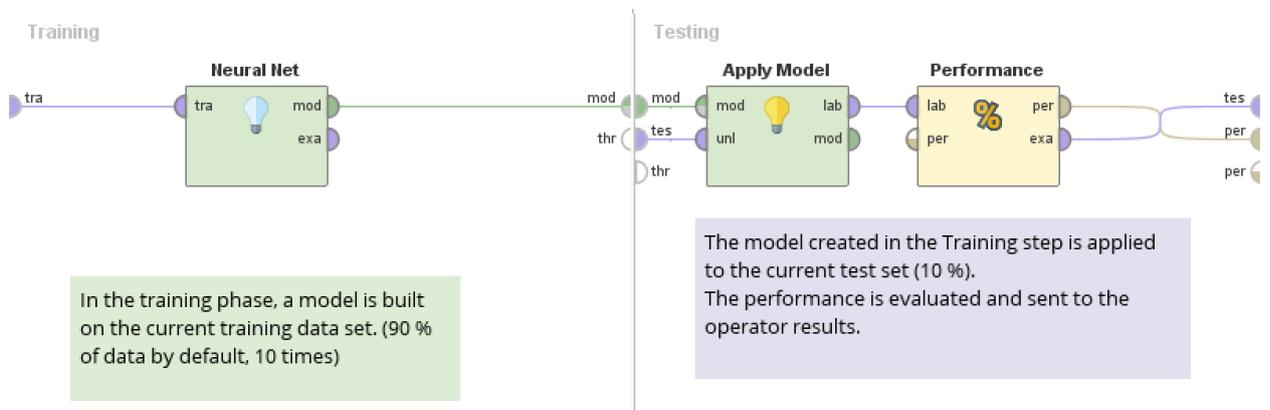
Matriks confusion dari algoritme ANN (Tabel 4) menunjukkan bahwa dari 139 data menghasilkan presisi kelas 86,44% untuk prediksi Ya dan 90% untuk prediksi Tidak dengan class recall sebesar 86,44% dan 90%. Akurasi algoritme ini diperoleh sebesar 88,35% ± 8,44% dan mikro 88,49%.

Matriks confusion dari algoritme decision tree (Tabel 5) menunjukkan bahwa dari 139 data menghasilkan presisi kelas 82,61% untuk prediksi Ya dan 88,89% untuk prediksi Tidak dengan class recall sebesar 84,75% dan 90%. Akurasi algoritme ini diperoleh sebesar 87,64% ± 8,92% dan mikro 87,77%.

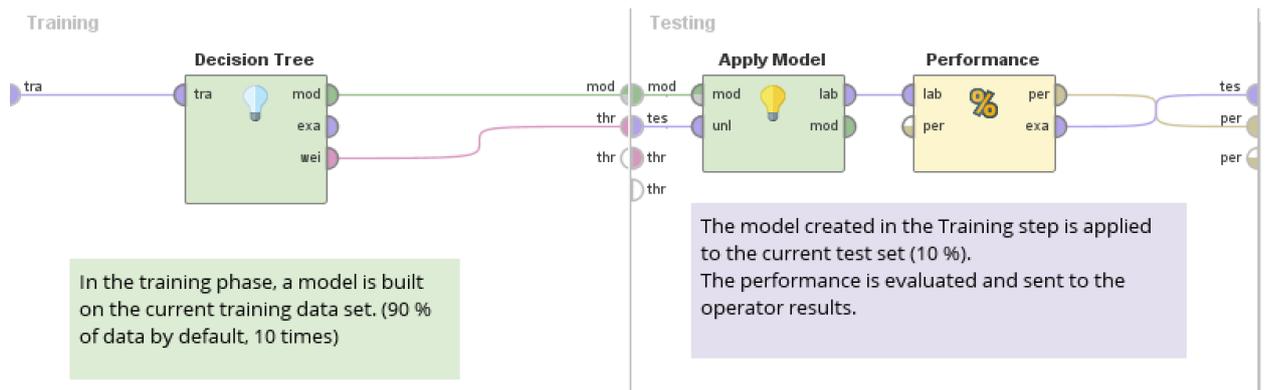
Perbandingan nilai akurasi dan perhitungan luas daerah di bawah kurva (AUC) dari algoritme data mining yang digunakan untuk membangun pengetahuan faktor penentu kecepatan waktu studi mahasiswa disajikan dalam Tabel 6. Algoritme SVM memberikan nilai terbaik, yaitu akurasi sebesar 90,55% dan nilai AUC sebesar 0,959. Akurasi terendah diperoleh dengan menggunakan algoritme decision tree, yaitu sebesar 87,64%. Hasil ini selaras dengan [14] yang



Gambar 4. Desain model klasifikasi SVM



Gambar 5. Desain model klasifikasi ANN



Gambar 6. Desain model klasifikasi *decision tree*

memperoleh akurasi terbaik menggunakan SVM dan LR sebesar 100%, sedangkan paling rendah menggunakan naive bayes sebesar 75,16% dan decision tree sebesar 80,01%. Perbedaan akurasi dari hasil penelitian ini dengan [15] disebabkan jumlah data yang berbeda, yaitu 139 data mahasiswa berbanding dengan 651 data.

Hasil tersebut menunjukkan bahwa algoritme SVM memiliki kinerja yang baik dalam membangun sebuah model yang digunakan untuk menemukan informasi berdasarkan dengan dataset yang disesuaikan dengan tiga tahap evaluasi yang ada pada aturan akademik yaitu

dua tahun pertama, dua tahun kedua dan akhir studi berdasarkan atribut yang telah ditentukan sebelumnya (Tabel 2). Informasi yang diperoleh tersebut memberikan pengetahuan untuk pemantauan evaluasi ketepatan masa studi mahasiswa di setiap fase evaluasi (Gambar 7), yang berbeda dengan [9], [15] yang hanya menggunakan dataset mahasiswa di satu tahun pertama, dengan [12] yang menggunakan dataset dua tahun pertama, dan dengan [11], [13], [14] yang menggunakan dataset mahasiswa yang telah menyelesaikan studinya. Mahasiswa dapat memperoleh ruang gerak yang lebih

Sistem Peringatan Dini Masa Studi

Hasil Evaluasi Tahun 1

Kriteria Evaluasi

1. IPK \geq 2.0
2. SKS \geq 40

Show 10 entries

NIM	Nama	Capaian	Status
H1C013001	WILDAN MUKHOLADUN WAHYUDI	Total SKS = 46 IPK = 2.74	Lolos Evaluasi Tahap 1
H1C013002	SATRIO DWI NUGROHO	Total SKS = 43 IPK = 2.14	Lolos Evaluasi Tahap 1

Gambar 7. Visualiasi presentasi laporan

dalam proses perbaikan diri karena peringatan diberikan secara berjenjang.

Penelitian ini tidak melakukan proses verifikasi, seperti dengan uji statistik, sebagaimana [15] yang melakukan verifikasi akurasi dengan uji T untuk menaikkan tingkat kepercayaan. Perbaikan kinerja masih dapat dilakukan dengan melakukan proses penalaan parameter pada proses validasi, sebagaimana [11], [12] untuk memperoleh nilai akurasi prediksi terbaik atau mempersiapkan data menggunakan metode cross-tabulation seperti [14].

IV. KESIMPULAN

Hasil perbandingan unjuk kerja algoritme decision tree, ANN dan SVM yang menggunakan data akademik mahasiswa aktif di FT Unsoed menunjukkan bahwa algoritme SVM memberikan nilai terbaik, yaitu accuracy sebesar 90,55% dan AUC sebesar 0,959. Model dengan algoritme SVM memiliki kinerja yang baik yang dapat digunakan untuk memberikan pengetahuan dari basis data mahasiswa dan diimplementasikan menjadi sebuah sistem peringatan dini ketepatan studi mahasiswa sehingga dapat digunakan oleh para pemangku kebijakan dalam melakukan pemantauan masa studi mahasiswa dan memetakan mahasiswa yang berpotensi mengalami keterlambatan masa studi.

UCAPAN TERIMA KASIH

Tim peneliti mengucapkan terima kasih kepada LPPM Unsoed atas dukungan dana pada penelitian skim riset dosen pemula dan Fakultas Teknik Unsoed atas dukungan penggunaan data akademik yang digunakan sebagai bahan mentah dalam penelitian ini.

DAFTAR PUSTAKA

- [1] BAN-PT, *Buku VI. Matriks Penilaian Instrumen Akreditasi Program Studi*. Jakarta: Badan Akreditasi Nasional Perguruan Tinggi, 2008.
- [2] D. Kabakchieva, "Predicting Student Performance by Using Data Mining," *Cybernetics and Information Technologies*, vol. 13, no. 1, pp. 61–72, 2013.
- [3] E. Osmanbegovic and M. Suljic, "Data Mining Approach for Predicting Student Performance," *Economic Review – Journal of Economics and Business*, vol. X, issue 1, pp. 3–12, 2012.
- [4] B. Santoso, *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu, 2007.
- [5] A. El-Halees, "Mining Students Data To Analyze Learning Behavior: a Case Study Educational Systems," in *Proc. of the 2008 International Arab Conference of Information Technology (ACIT2008)*, 15-18 Dec 2008, University of Sfax, Tunisia.
- [6] S. Ahmed, R. Paul, A. Sayed, and L. Hoque, "Knowledge Discovery from Academic Data using Association Rule Mining," in *Proc. of 17th International Conference on Computer and Information Technology (ICIT)*, 22-23 Dec 2014, Dhaka, Bangladesh, pp. 22–23.
- [7] M. I. Al-Twijri and A. Y. Noaman, "A New Data Mining Model Adopted for Higher Institutions," *Procedia Computer Science*, vol. 65, pp. 836–844, 2015.
- [8] S. T. Karamouzis and A. Vrettos, "Sensitivity Analysis of Neural Network Parameters for Identifying the Factors for College Student Success," in *Proc. of the 2019 World Congress on*

- Computer Science and Information Engineering*, 2 Apr 2009, Los Angeles, USA.
- [9] M. S. Suhartinah and Ernastuti, "Graduation Prediction of Gunadarma University Students Using Algorithm Naive Bayes and C4.5 Algorithm," Skripsi, Universitas Gunadarma, 2010.
- [10] S. Salmu and A. Solichin, "Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naive Bayes: Studi Kasus UIN Syarif Hidayatullah Jakarta," in *Prosiding Seminar Nasional Multidisiplin Ilmu*, 22 Apr 2017, Jakarta, Indonesia.
- [11] Y. S. Samponu dan K. Kusriani, "Optimasi Algoritma Naive Bayes Menggunakan Metode Cross Validation Untuk Meningkatkan Akurasi Prediksi Tingkat Kelulusan Tepat Waktu," *Jurnal ELTIKOM*, vol. 1, no. 2, pp. 56–63, 2017.
- [12] M. A. Banjarsari, H. I. Budiman, dan A. Farmadi, "Penerapan K-Optimal pada Algoritma KNN untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer FMIPA UNLAM berdasarkan IP sampai dengan Semester 4," *Kumpulan Jurnal Ilmu Komputer*, vol. 02, no. 02, pp. 50–64, 2015.
- [13] G. I. Marthasari, "Implementasi Teknik Data Mining untuk Evaluasi Kinerja Mahasiswa Berdasarkan Data Akademik," *Fountain of Informatics Journal*, vol. 2, no. 2, pp. 56–63, 2017.
- [14] M. H. Meinanda, M. Annisa, N. Muhandri, and K. Suryadi, "Prediksi Masa Studi Sarjana dengan Artificial Neural Network," *Internetworking Indonesia Journal*, vol. 1, no. 2, pp. 31–35, 2009.
- [15] I. Tahyudin, E. Utami, and A. Amborowati, "Comparing Classification Algorithm of Data Mining to Predict the Graduation Students on Time," in *Proc. of the Information Systems International Conference (ISICO)*, 2-4 Dec 2013, Surabaya, Indonesia, pp. 379–384.
- [16] R. Wirth and J. Hipp, "CRISP-DM: Towards a Standard Process Model for Data Mining," in *Proc. of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 2012, pp. 23-39.

