



Prediksi Ketepatan Waktu Lulus Mahasiswa dengan *k-Nearest Neighbor* dan *Naïve Bayes Classifier* (Studi Kasus Prodi D3 Sistem Informasi Universitas Airlangga)

Wilda Imama Sabilla¹, Tesa Eranti Putri²

¹Fakultas Vokasi, Universitas Airlangga, email: wildaimama@vokasi.unair.ac.id

²Fakultas Vokasi, Universitas Airlangga, email: tesaep@gmail.com

Abstrak

Salah satu aspek pengukuran kualitas dalam evaluasi keberhasilan penyelenggaraan pendidikan tinggi adalah ketepatan lulus mahasiswa. Jumlah prosentase mahasiswa yang lulus tepat waktu menjadi indikator keberhasilan pelaksanaan proses belajar mengajar di suatu program studi. Penelitian ini menawarkan penggunaan metode penggalian data untuk memprediksi waktu lulus mahasiswa menggunakan dua metode yaitu *k-Nearest Neighbour* dan *Naïve Bayes Classifier*. Hasil dari penelitian ini berupa sistem yang dapat memprediksi ketepatan waktu lulus. Uji coba dilakukan dengan menggunakan data lulusan mahasiswa D3 Sistem Informasi Universitas Airlangga. Hasil uji coba menunjukkan bahwa metode *k-Nearest Neighbor* menghasilkan akurasi lebih tinggi dibandingkan dengan *Naïve Bayes Classifier*. Akurasi tertinggi diperoleh dengan menggunakan metode *k-Nearest Neighbor* yaitu sebesar 98.7%. Oleh karena itu dapat disimpulkan bahwa sistem yang dibangun pada penelitian ini mampu memprediksi ketepatan waktu lulus dengan akurasi cukup tinggi.

Kata kunci: prediksi ketepatan waktu lulus, *k-Nearest Neighbor*, *Naive Bayes Classifier*

Abstract

One aspect of quality measurement for successful implementation of higher education is the timeliness of students graduation. The percentage of students who graduate on time become an indicator of successful implementation of teaching and learning process in a study program. This study offers the use of data mining methods to predict the timeliness of students graduation using two methods which are *k-Nearest Neighbors* and *Naïve Bayes Classifier*. The results of this study is a system that can predict the timeliness of students graduation. The testing is done by using data of graduate student in D3 Information System of Airlangga University. The result shows that *k-Nearest Neighbor* method produces higher accuracy than *Naïve Bayes Classifier*. The highest accuracy which is produces by *k-Nearest Neighbor* is 98.7%. Therefore it can be concluded that the system built in this study is able to predict the timeliness of students graduation with high accuracy.

Keywords: timeliness of students graduation prediction, *k-Nearest Neighbor*, *Naive Bayes Classifier*

1. Pendahuluan

Perguruan tinggi merupakan penyelenggara pendidikan dengan jenjang yang lebih tinggi daripada pendidikan menengah di jalur pendidikan sekolah. Perguruan tinggi diselenggarakan untuk mempersiapkan peserta didik menjadi anggota masyarakat yang memiliki kemampuan akademis dan profesional yang dapat menerapkan, mengembangkan, dan menciptakan ilmu pengetahuan, teknologi, dan kesenian [1]. Pendidikan Tinggi memiliki peran besar dalam mencetak sumber daya manusia yang kompeten dan siap menghadapi era persaingan global. Oleh karena itu, peningkatan kualitas perguruan tinggi sangat dibutuhkan.

Salah satu aspek pengukuran kualitas dalam evaluasi keberhasilan penyelenggaraan pendidikan tinggi adalah mahasiswa. Mahasiswa menjadi unsur penting karena mahasiswa nantinya akan langsung berhubungan dengan stakeholder. Perguruan tinggi dengan kualitas lulusan yang baik akan mendapat kepercayaan stakeholder dalam merekrut maupun menggunakan lulusannya. Terdapat beberapa unsur penilaian perguruan tinggi yang berhubungan dengan mahasiswa. Unsur penilaian tersebut antara lain nilai ujian masuk, nilai akademik, prestasi yang dicapai mahasiswa, kompetensi yang dimiliki mahasiswa, dan prosentase mahasiswa lulus tepat waktu [2].

Program studi D3 Sistem Informasi, yang berada di bawah Fakultas Vokasi Universitas Airlangga, memiliki permasalahan dengan ketepatan waktu lulusan dari tahun ke tahun. Berdasarkan data dari Laporan Evaluasi Diri D3 Sistem Informasi, dalam tiga tahun terakhir prosentase mahasiswa yang lulus tepat waktu kurang dari 40%. Hal ini menyebabkan menumpuknya jumlah mahasiswa dalam satu tahun ajaran serta kurangnya penyerapan lulusan oleh dunia kerja. Terlambatnya waktu lulus juga memberatkan dosen yang harus membimbing mahasiswa dalam jumlah yang lebih banyak dan menurunkan kepercayaan orang tua terhadap program studi.

Penelitian ini menawarkan penggunaan metode penggalian data (*data mining*) untuk memprediksi waktu lulus mahasiswa. Penggalian data (*data mining*), suatu bentuk dari eksplorasi dan analisis data, adalah proses untuk mengekstrak pola dan hubungan dari data berjumlah banyak secara otomatis [3]. Beberapa faktor yang mempengaruhi ketepatan waktu lulus akan dianalisa. Faktor-faktor tersebut yaitu nilai IPS (Indeks Prestasi Semester) mulai dari semester 1 hingga 4, jenis kelamin, dan alamat tinggal. Nilai IPS yang digunakan dalam analisa hanya mencapai semester 4 karena diharapkan hasilnya dapat digunakan oleh program studi sebagai pedoman dalam menentukan sikap pada semester 5 dan 6 sehingga mahasiswa dapat lulus tepat waktu.

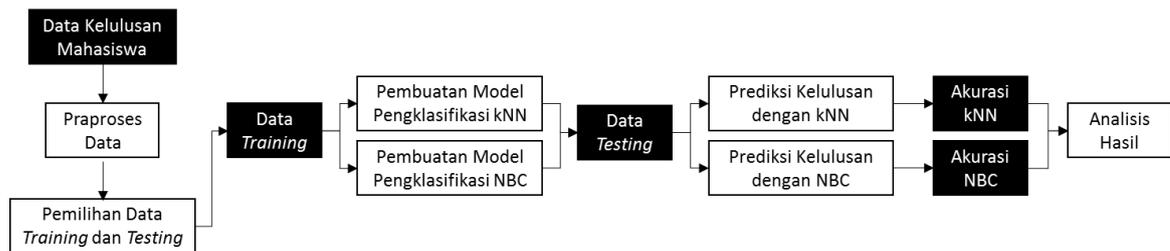
Beberapa metode *data mining* dapat digunakan untuk memprediksi ketepatan waktu lulus yaitu berupa metode klasifikasi. Metode klasifikasi menemukan utilitas pada aktivitas manusia dan secara khusus membantu pengambilan keputusan secara otomatis [4]. Metode klasifikasi antara lain *Naïve Bayes* [5][6], *information gain* [5], *decision tree* [7], *rule learner* dan *neural network* [8]. Pada penelitian ini akan diterapkan dua metode untuk memprediksi kelulusan mahasiswa yaitu kNN (*k-Nearest Neighbor*) dan NBC (*Naïve Bayes Classifier*). Dari kedua metode tersebut akan dianalisis metode yang lebih efektif untuk memprediksi ketepatan waktu lulus berdasarkan data lulusan D3 Sistem Informasi Universitas Airlangga.

Bagian selanjutnya dari artikel ini tersusun dari Metodologi Penelitian dalam bagian 2, pembahasan kinerja dari metode klasifikasi dalam Uji Coba dan Analisis Hasil pada bagian 3, serta Kesimpulan pada bagian 4.

2. Metodologi Penelitian

Pada penelitian ini, metodologi yang digunakan untuk memprediksi ketepatan waktu lulus mahasiswa ditunjukkan pada Gambar 1. Data masukan berupa data kelulusan mahasiswa

berikut atribut penentu serta label kelas untuk setiap data. Data masukan dilakukan praproses terlebih dahulu agar data tersebut dapat digunakan ke dalam metode klasifikasi *k-Nearest Neighbor* maupun *Naïve Bayes Classifier*. Setelah itu, dilakukan pemilihan data *training* dan data *testing* secara acak. Data *training* merupakan data yang digunakan untuk membuat model pengklasifikasi baik pada metode *k-Nearest Neighbor* maupun *Naïve Bayes Classifier*. Model pengklasifikasi yang dihasilkan oleh data *training* selanjutnya digunakan untuk memprediksi kelas dari data *testing*, sehingga pada data *testing* label kelas tidak digunakan. Hasil prediksi berupa label kelas akan dibandingkan dengan label kelas sebenarnya untuk dihitung akurasi metode baik *k-Nearest Neighbor* maupun *Naïve Bayes Classifier*. Hasil prediksi kedua metode tersebut akan dianalisis dan dilihat metode mana yang mampu memprediksi lebih baik.



Gambar 1. Metodologi Penelitian untuk Prediksi Ketepatan Waktu Lulus

2.1 Praproses Data

Praproses data adalah sebuah tahap yang dilakukan untuk memperbaiki data masukan sehingga sesuai dengan format data yang digunakan pada metode klasifikasi. Pada penelitian ini, beberapa variabel yang memiliki dua nilai akan disederhanakan menjadi 1 atau 0. Sebagai contoh, nilai 'laki-laki' pada variabel jenis kelamin diubah menjadi 1 dan 'perempuan' menjadi 0. Begitu juga untuk label kelas yang berisi 'tepat waktu' atau 'tidak tepat waktu'. 'tepat waktu' akan diubah menjadi 1, sedangkan 'tidak tepat waktu' diubah menjadi 0.

2.2 Metode Klasifikasi *k-Nearest Neighbor*

Metode *k-Nearest Neighbour* (*k-NN*) adalah metode untuk melakukan klasifikasi objek berdasarkan data latih terdekat pada ruang fitur. *kNN* adalah sebuah jenis pembelajaran berdasarkan *instance* atau *lazy learning*, dimana fungsi hanya sesuai secara lokal dan komputasi ditangguhkan hingga klasifikasi.

Metode *kNN* termasuk salah satu pembelajaran mesin yang paling sederhana dimana sebuah objek diklasifikasikan berdasarkan kelas mayoritas sejumlah *k* objek tetangga. Pemilihan terbaik *k* tergantung pada data. Secara umum nilai *k* lebih besar akan mengurangi efek derau saat klasifikasi tetapi menyebabkan perbedaan antar kelas tidak mencolok. Akurasi dari metode *kNN* dapat berkurang karena adanya derau dan fitur atau atribut yang tidak relevan maupun fitur yang skalanya tidak konsisten dibandingkan fitur lainnya [7].

Terdapat beberapa metode yang digunakan untuk perhitungan jarak antar data dan tetangga antara lain *Euclidean*, *cityblock*, *cosine*, dan *correlation*. *Cityblock* dihitung menggunakan jumlah absolut dari perbedaan dua titik pada koordinat kartesian. *Cosine* dihitung menggunakan pengurangan antara satu oleh *cosine* dari sudut antar titik. *Correlation* dihitung berdasarkan dependensi statistik antar dua vektor [9].

2.3 Metode Klasifikasi *Naïve Bayes Classifier*

Model statistik merupakan salah satu model yang cukup andal sebagai pendukung pengambilan keputusan. Konsep probabilitas merupakan salah satu bentuk model statistik, salah satu metode yang menggunakan konsep probabilitas adalah *Naïve Bayes Classifier*. Pada metode ini, semua atribut memiliki kontribusi dalam pengambilan keputusan. Setiap atribut

memiliki bobot sama dan independen (saling bebas). Jika terdapat sejumlah k atribut yang saling bebas, perhitungan nilai probabilitas ditunjukkan pada Persamaan 1 [10].

$$P(X_1, \dots, X_k|C) = P(X_1|C) * \dots * P(X_k|C) \quad (1)$$

dimana X adalah atribut dan $P(X|C)$ adalah probabilitas atribut X terhadap kelas C . Jika atribut ke- i bersifat diskrit, maka $P(X_i|C)$ diestimasi sebagai frekuensi relatif dari sampel yang memiliki nilai X_i sebagai atribut ke- i dalam kelas C . Namun, jika atribut ke- i bersifat kontinu, maka $P(X_i|C)$ diestimasi dengan fungsi densitas Gauss pada Persamaan (2).

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (2)$$

3. Uji Coba dan Analisis Hasil

3.1 Data Uji Coba

Data uji coba yang digunakan diambil dari data akademik mahasiswa program studi D3 Sistem Informasi Universitas Airlangga. Jumlah data uji dalam penelitian ini adalah 159 data berupa data kelulusan mahasiswa pada tiga tahun terakhir. Data terdiri dari beberapa atribut yaitu nomor induk mahasiswa, nama mahasiswa, IPS semester 1 hingga semester 4, jenis kelamin, dan alamat asal mahasiswa. Alamat asal mahasiswa dibagi ke dalam dua kategori yaitu dalam kota dan luar kota. Dalam kota adalah ketika mahasiswa beralamat di kota Surabaya, sedangkan bagi mahasiswa yang alamatnya di luar Surabaya akan masuk ke kategori luar kota. Pengkategorian alamat ini dilakukan untuk memudahkan pengidentifikasian atribut alamat. Data kemudian diklasifikasikan ke dalam dua kelas yaitu lulus tepat waktu dan tidak tepat waktu. Terdapat 82 buah data yang berlabelkan lulus tepat waktu, sedangkan lulus tidak tepat waktu berjumlah 77 data.

Tabel 1 menunjukkan karakteristik data uji berdasarkan nilai atribut yang masuk ke dalam kelas lulus tepat waktu dan tidak tepat waktu. Untuk atribut jenis kelamin, terdapat 42 laki-laki dan 40 perempuan yang lulus tepat waktu. Pada kelas tidak tepat waktu terdapat 52 laki-laki dan 25 perempuan. Pada kelas tepat waktu terdapat 24 mahasiswa beralamat dalam kota dan 58 dari luar kota. Untuk kelas tidak tepat waktu terdapat 47 mahasiswa dari dalam Surabaya dan 30 mahasiswa dari luar Surabaya. Data IPS semester 1 hingga 4 berupa data desimal antara 0,00 sampai 4,00 dengan ketepatan dua angka dibelakang koma.

Tabel 1. Karakteristik Data Uji Coba

| Kelas | Jenis Kelamin | | Alamat | |
|-------------------|---------------|-----------|------------|-----------|
| | Laki-laki | Perempuan | Dalam Kota | Luar Kota |
| Tepat Waktu | 42 | 40 | 24 | 58 |
| Tidak Tepat Waktu | 52 | 25 | 47 | 30 |
| Total | 94 | 65 | 71 | 88 |

Untuk melakukan prediksi kelas menggunakan klasifikasi, data kelulusan mahasiswa dibagi ke dalam dua bagian yaitu data *training* (data latihan) dan data *testing* (data uji). Pemilihan data *training* dan *testing* dilakukan secara acak dengan prosentase tertentu. Pada penelitian ini prosentase data *training* terhadap data *testing* yang dipakai pada uji coba yaitu 80%:20%, 60%:40%, dan 50%:50%.

3.2 Penghitungan Akurasi

Kebenaran metode klasifikasi dapat diperoleh dengan menghitung akurasinya. Akurasi dihitung melalui nilai TP, FN, FP, dan TN menggunakan Persamaan 3 [11]. Berdasarkan Tabel 2 TP terjadi ketika kelas prediksi dan kelas sebenarnya memiliki label ‘tepat waktu’. TN terjadi ketika kelas prediksi dan kelas sebenarnya memiliki label yang sama yaitu ‘tidak tepat waktu’. FN terjadi ketika kelas prediksi menunjukkan ‘tidak tepat waktu’ dan kelas sebenarnya memiliki label ‘tepat waktu’. FP berkebalikan dengan FN yaitu ketika kelas prediksi memiliki label ‘tepat waktu’ sedangkan kelas sebenarnya memiliki label ‘tidak tepat waktu’.

$$akurasi = \frac{TP + TN}{TP + FP + FN + TN} \quad (3)$$

Tabel 2. Confusion Matrix untuk Akurasi

| Kelas Sebenarnya | Kelas Prediksi | |
|-------------------|----------------|-------------------|
| | Tepat Waktu | Tidak Tepat Waktu |
| Tepat Waktu | TP | FN |
| Tidak Tepat Waktu | FP | TN |

3.3 Uji Coba Kebenaran Klasifikasi dengan *k-Nearest Neighbor*

Uji coba kebenaran klasifikasi dengan *k-Nearest Neighbor* dihitung menggunakan nilai akurasi. Untuk pengujian pertama, akurasi dihitung berdasarkan perbandingan jumlah data *training* dan data *testing*. Pengujian berdasarkan perbandingan jumlah data menggunakan perhitungan jarak ‘*cityblock*’ dan jumlah *neighbor* 2. Berdasarkan Tabel 3, nilai akurasi terbaik diperoleh ketika perbandingan jumlah data *training* dan *testing* adalah 50%:50% yaitu sebesar 98.7%. Tabel tersebut menunjukkan semakin besar jumlah data *testing* yang digunakan maka akurasinya akan naik.

Tabel 3. Nilai Akurasi k-NN Berdasarkan Perbandingan Jumlah Data *Training* dan *Testing*

| Jumlah Data <i>Training</i> (%) | Jumlah Data <i>Testing</i> (%) | Akurasi (%) | TP | FN | FP | TN |
|---------------------------------|--------------------------------|-------------|----|----|----|----|
| 80 | 20 | 96.9 | 16 | 1 | 0 | 15 |
| 60 | 40 | 98.4 | 31 | 1 | 0 | 31 |
| 50 | 50 | 98.7 | 40 | 1 | 0 | 39 |

Pengujian kedua dilakukan dengan membandingkan pengaruh metode perhitungan jarak (*distance*) terhadap akurasi k-NN. Pengujian ini menggunakan perbandingan jumlah data *training* dan *testing* 50%:50% dan jumlah *neighbor* 2. Dengan merujuk pada Tabel 4, metode terbaik untuk menghitung jarak yaitu ‘*cityblock*’ yang menghasilkan akurasi 98.7%. Sedangkan penggunaan ‘*cosine*’ dan ‘*correlation*’ menghasilkan akurasi kurang dari 80%.

Pengujian ketiga adalah membandingkan pengaruh jumlah *neighbor* (tetangga) terhadap akurasi k-NN. Pengujian ini menggunakan perbandingan jumlah data *training* dan *testing* 50%:50% dan metode penghitungan jarak ‘*cityblock*’. Berdasarkan data pada Tabel 5, jumlah *neighbor* yang menghasilkan akurasi terbaik sebesar 98.7% adalah 2 buah *neighbor*. Pada penggunaan 3 *neighbor* nilai akurasi turun menjadi 80% dan naik lagi setelah penggunaan 4 *neighbor* menjadi 93.7%. Tetapi penggunaan *neighbor* sejumlah 5 menghasilkan akurasi terendah yaitu 76.2%.

Tabel 4. Nilai Akurasi k-NN Berdasarkan Metode Perhitungan Jarak

| <i>Distance</i> | Akurasi (%) | TP | FN | FP | TN |
|-----------------|-------------|----|----|----|----|
| euclidean | 88.7 | 37 | 4 | 5 | 34 |
| cityblock | 98.7 | 40 | 1 | 0 | 39 |
| cosine | 77.5 | 30 | 11 | 7 | 32 |
| correlation | 65.0 | 24 | 17 | 11 | 28 |

Tabel 5. Nilai Akurasi Berdasarkan Jumlah *Neighbor* (Tetangga)

| Jumlah <i>Neighbor</i> | Akurasi (%) | TP | FN | FP | TN |
|------------------------|-------------|----|----|----|----|
| 2 | 98.7 | 40 | 1 | 0 | 39 |
| 3 | 80 | 35 | 6 | 10 | 29 |
| 4 | 93.7 | 38 | 3 | 2 | 37 |
| 5 | 76.2 | 33 | 8 | 11 | 28 |

3.4 Uji Coba Kebenaran Klasifikasi dengan *Naïve Bayes Classifier*

Kebenaran klasifikasi dengan *Naïve Bayes Classifier* dilakukan dengan menghitung nilai akurasi yang dihasilkan oleh data uji coba. Pengujian dilakukan dengan membandingkan nilai akurasi yang dihasilkan berdasarkan perbandingan jumlah data *training* dan data *testing*. Pada Tabel 5, nilai akurasi terbaik diperoleh ketika perbandingan jumlah data *training* dan *testing* adalah 60%:40% yaitu sebesar 80.9%. Pada penggunaan data *training* 80% diperoleh akurasi terendah yaitu 75%. Sedangkan untuk data *training* sejumlah 50%, akurasi yang dihasilkan adalah 80%.

Tabel 6. Nilai Akurasi NBC Berdasarkan Perbandingan Data *Training* dan *Testing*

| Jumlah Data <i>Training</i> (%) | Jumlah Data <i>Testing</i> (%) | Akurasi (%) | TP | FN | FP | TN |
|---------------------------------|--------------------------------|-------------|----|----|----|----|
| 80 | 20 | 75 | 15 | 2 | 6 | 9 |
| 60 | 40 | 80.9 | 29 | 3 | 9 | 22 |
| 50 | 50 | 80 | 38 | 3 | 13 | 26 |

3.5 Analisis Hasil Klasifikasi

Kebenaran metode klasifikasi baik dengan kNN maupun NBC diperoleh menggunakan perhitungan nilai akurasi. Berdasarkan uji coba yang telah dilakukan pada sub bab 3.3 dan 3.4 nilai akurasi terbaik diperoleh menggunakan metode kNN yaitu sebesar 98.7%. Metode kNN menghasilkan akurasi lebih baik karena tipe data masukan lebih banyak yang berupa numerik dibanding kategorikal. Empat dari enam variabel masukan bertipe numerik yaitu IPS semester 1, IPS semester 2, IPS semester 3, dan IPS semester 4. Metode kNN menghitung jarak antar data yang memungkinkan ketika tipe data berupa angka. Sedangkan NBC menghitung probabilitas suatu nilai variabel yang akan lebih mudah dilakukan ketika data bertipe kategorikal. Sehingga penggunaan NBC kurang efektif untuk data uji berupa data lulusan mahasiswa.

Pada klasifikasi menggunakan kNN, terdapat dua parameter yang dibandingkan yaitu metode perhitungan jarak dan jumlah *neighbor*. Jumlah *neighbor* yang menghasilkan akurasi terbaik adalah 2 buah. Hal ini dimungkinkan karena semakin banyak data *neighbor* yang digunakan akan menambah kebiasaan data yang berakibat pada kesalahan pelabelan kelas.

4. Kesimpulan

Pada penelitian ini diterapkan metode penggalian data berupa klasifikasi untuk memprediksi ketepatan waktu lulus mahasiswa. Terdapat dua metode yang digunakan yaitu *k-Nearest Neighbor* dan *Naïve Bayes Classifier*. Sedangkan untuk data yang digunakan pada pengujian metode klasifikasi berasal dari data kelulusan mahasiswa pada prodi D3 Sistem Informasi, Universitas Airlangga. Berdasarkan uji coba dapat disimpulkan bahwa sistem mampu memprediksi ketepatan waktu lulus dengan akurasi cukup tinggi. Uji coba yang dilakukan menunjukkan bahwa metode *k-Nearest Neighbor* menghasilkan akurasi lebih tinggi dibandingkan dengan *Naïve Bayes Classifier*. Akurasi tertinggi diperoleh dengan menggunakan metode *k-Nearest Neighbor* yaitu sebesar 98.7%. Akurasi ini diperoleh dengan menerapkan metode '*cityblock*' untuk parameter metode perhitungan jarak dan 2 buah *neighbor* untuk parameter jumlah *neighbor*.

Penelitian ini masih dapat dikembangkan dan diperbaiki. Salah satu pengembangan yang bisa dilakukan yaitu dengan mencoba menggunakan metode klasifikasi lain untuk dibandingkan akurasinya. Selain itu dapat pula dilakukan analisis variabel paling berpengaruh dalam memprediksi ketepatan waktu lulus mahasiswa.

Daftar Pustaka

- [1] Menteri Riset, Teknologi, dan Pendidikan Tinggi, 2015, *Peraturan Menteri Riset, Teknologi, dan Pendidikan Tinggi Nomor 44 Tahun 2015 tentang Standar Nasional Pendidikan Tinggi*, Jakarta: Kementerian Riset, Teknologi, dan Pendidikan Tinggi.
- [2] Ridwan, Mujib, Hadi Suyono and M. Sarosa, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier", *Jurnal EECCIS* 7(1), 59-64, 2013.
- [3] Yu, Chong Ho, et al, "A Data Mining Approach for Identifying Predictors of Student Retention from Sophomore to Junior Year", *Journal of Data Science* 8, 307-325, 2010.
- [4] Hssina, Badr, et al, "A comparative study of decision tree ID3 and C4.5", *International Journal of Advanced Computer Science and Applications Special Issue*, 13-19, 2013.
- [5] Alfajri, Rd Muhammad, et al, "Pengklasifikasian Kemampuan Akademik Mahasiswa Menggunakan Metode Information Gain dan Naive Bayes Classifier dalam Prediksi Penyelesaian Studi Tepat Waktu", dalam *Prosiding SNST ke-7*, pp. 144-149, 2016.
- [6] Jananto, Arief, "Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa", *Jurnal Teknologi Informasi DINAMIK* 18(1), 09-16, 2013.
- [7] Kabakchieva, Dorina, "Predicting Student Performance by Using Data Mining Methods for Classification", *Cybernetics and Information Technologies* 13(1), 61-72, 2013.
- [8] Kabakchieva, Dorina, "Student Performance Prediction by Using Data Mining Classification Algorithms", *International Journal of Computer Science and Management* 1(4), 686-690, 2012.

- [9] Chomboon, Kittipong, “An Empirical Study of Distance Metrics for k-Nearest Neighbor Algorithm”, dalam Proceedings of the 3rd International Conference on Industrial Application Engineering 2015, pp. 280-285, 2015.
- [10] Kusumadewi, Sri, “Klasifikasi Status Gizi Menggunakan Naive Bayesian Classification”, CommIT 3(1), 6-11, 2009.
- [11] Sabilla, Wilda Imama, et al, “Automatic Detection of Proliferative Diabetic Retinopathy With Hybrid Feature Extraction Based on Scale Space Analysis and Tracking”, dalam The 1 st International Seminar on Science and Technology, pp. E313-95 – E313-96, 2015.