

CASE BASED REASONING PREDIKSI WAKTU STUDI MAHASISWA MENGUNAKAN METODE EUCLIDEAN DISTANCE DAN NORMALISASI MIN-MAX

Henni Endah Wahanani¹, Made Hanindia Prami Swari², Fawwaz Ali Akbar³

^{1,2,3}Universitas Pembangunan Nasional "Veteran" Jawa Timur

Email: ¹henniendah222@gmail.com, ²madehanindia.fik@upnjatim.ac.id, ³ewezali@gmail.com

*Penulis Korespondensi

(Naskah masuk: 10 Agustus 2020, diterima untuk diterbitkan: 26 November 2020)

Abstrak

Salah satu penyebab dari lamanya waktu tempuh mahasiswa di Jurusan Informatika UPN "Veteran" Jawa Timur adalah sulitnya memantau kemajuan studi mahasiswa secara seksama, mengingat jumlah mahasiswa yang cukup banyak serta pihak akademik belum memiliki metode yang akurat untuk memetakan mahasiswa yang diprediksi akan mengalami keterlambatan dalam penyelesaian studinya. Melalui perkembangan teknologi informasi yang berkembang pesat saat ini, maka sangat dimungkinkan untuk membuat sebuah sistem yang mampu memprediksi kemungkinan keterlambatan kelulusan mahasiswa melalui penggunaan berbagai metode komputasi yang ada. Salah satu pendekatan yang dapat digunakan untuk membuat sebuah sistem prediksi kelulusan adalah menggunakan pendekatan populer yang digunakan dalam pembuatan sistem cerdas (*intelligent system*) yaitu *case based reasoning* (CBR). Langkah-langkah yang dilakukan pada penelitian ini adalah melakukan pengumpulan dan memasukkan data kasus pada basis kasus, melakukan pra-prosesing yakni normalisasi atribut yang akan digunakan dalam perhitungan similaritas antar kasus menggunakan normalisasi min-max, implementasi CBR menggunakan metode Euclidean Distance, serta melakukan pengujian pada 141 data kasus. Dari sisi perhitungan akurasi, sistem mampu memberikan nilai akurasi paling tinggi sebesar 100% pada pengujian berdasarkan predikat kelulusan, sedangkan berdasarkan ketepatan waktu, sistem mampu memberikan akurasi tertinggi dengan nilai 85,71%, dan sistem mampu memberikan nilai akurasi tertinggi sebesar 71,43% pada pengujian berdasarkan massa studi. Untuk pengujian presisi, sistem mampu mengasalkan nilai terbesar berturut-turut sebesar 90,90%, 43,33%, dan 100%. Sedangkan pada pengujian sensitivitas, sistem berturut-turut mampu menghasilkan nilai sebesar 90,90%, 40,48%, dan 100%. Hasil pengujian ini tentunya sangat bergantung dari basis kasus yang dimiliki, oleh sebab itu perbaikan dan peningkatan jumlah kasus yang dimiliki diharapkan mampu meningkatkan performa sistem rekomendasi.

Kata kunci: *Case Based Reasoning, Sistem Prediksi, Euclidean Distance*

CASE BASED REASONING STUDENT TIME PREDICTION USING EUCLIDEAN DISTANCE METHOD AND MIN-MAX NORMALIZATION

Abstract

One of the reasons for the length of study time for students of the Informatics study program of UPN "Veteran" Jawa Timur is the difficulty of monitoring the progress, given the large number of students and academics do not have an accurate method to map students who are predicted to experience delays. It is possible to create a system that is able to predict the possibility of student graduation delay through the use of various existing computational methods. One approach that can be used to create a graduation prediction system is to use the popular approach namely case based reasoning (CB). The steps taken in this study are collecting and entering case data, normalizing the attributes using min-max normalization, implementing CBR using the Euclidean Distance, and system testing in 141 data case. System is able to provide the highest accuracy value of 100% in testing based on the predicate of graduation, while based on timeliness, the system is able to provide the highest accuracy value with a value of 85.71%, and the system is able to provide the highest accuracy value of 71.43%. on testing based on the study period. For precision testing, the system was able to produce the largest values of 90.90%, 43.33% and 100%, respectively. Whereas in the sensitivity test, the system was able to produce values of 90.90%, 40.48%, and 100% respectively. The results of this test are of course very dependent on the basis of cases that are owned, therefore improvements and an increase in the number of cases owned are expected to be able to improve the performance.

Keywords: *Case Based Reasoning, Sistem Prediksi, Euclidean Distance*

1. PENDAHULUAN

Pendidikan tinggi merupakan jenjang pendidikan setelah pendidikan menengah yang mencakup program pendidikan diploma, sarjana, magister, spesialis dan doktor yang diselenggarakan oleh perguruan tinggi. Perguruan tinggi memiliki peranan sangat besar dalam hal peningkatan kualitas Sumber Daya Manusia (SDM) yang berimbas pada percepatan kemajuan bangsa, khususnya di Indonesia. Perguruan tinggi diharapkan mampu menyediakan insan-insan muda terampil yang memiliki kemampuan dan kompetensi sesuai kebutuhan dunia usaha dan industri, baik di Indonesia maupun mancanegara. Perguruan tinggi yang berkualitas baik juga berkontribusi melahirkan pemimpin dan pemikir yang handal dalam mensukseskan pembangunan negara di masa depan (Nulhaqim *et al.*, 2016).

Untuk dapat menjadi seorang sarjana, mahasiswa diwajibkan untuk memenuhi standar kompetensi lulusan, dimana untuk seorang sarjana wajib menempuh minimal 144 Satuan Kredit Semester (SKS) dan waktu tempuh studi ideal 4 tahun (8 semester) dan maksimal selama 7 tahun atau 14 semester (Pendidikan, Kebudayaan and Indonesia, 2020).

Kesuksesan penyelenggaraan pendidikan di perguruan tinggi dapat diukur dari beberapa tolak ukur, salah satunya adalah tingginya jumlah mahasiswa yang lulus tepat waktu (Pambudi, Supianto and Setiawan, 2019). Berdasarkan data akademik yang diperoleh, adapun rata-rata masa studi mahasiswa di program studi Informatika, Fakultas Ilmu Komputer Universitas Pembangunan Nasional "Veteran" Jawa Timur program studi adalah selama 5,2 tahun (62,5 bulan) pada tahun akademik 2019/2020 dan selama 5,1 tahun (61 bulan) pada tahun akademik 2018/2019, dan rata-rata masa studi paling lama tercatat pada tahun akademik 2017/2018 yakni selama 6,29 tahun (75,5 bulan) dengan masa studi terlama adalah selama 6,75 tahun (81 bulan). Melihat data statistik tersebut, dapat dikatakan bahwa rata-rata waktu studi mahasiswa di program studi informatika masih berada cukup jauh dari rata-rata waktu studi normal mahasiswa S1, sesuai peraturan dari Kementerian Pendidikan dan Kebudayaan Republik Indonesia.

Salah satu penyebab dari lamanya waktu tempuh mahasiswa di Jurusan Informatika tersebut adalah sulitnya memantau kemajuan studi mahasiswa secara seksama dari pihak akademik, mengingat jumlah mahasiswa yang cukup banyak serta pihak akademik belum memiliki metode yang akurat untuk memetakan mahasiswa yang diprediksi akan mengalami keterlambatan dalam penyelesaian studinya. Melalui perkembangan teknologi informasi yang berkembang pesat saat ini, maka sangat dimungkinkan untuk membuat sebuah sistem yang mampu memprediksi kemungkinan keterlambatan

kelulusan mahasiswa melalui penggunaan berbagai metode komputasi yang ada.

Salah satu pendekatan yang dapat digunakan untuk membuat sebuah sistem prediksi kelulusan adalah menggunakan pendekatan populer yang digunakan dalam pembuatan sistem cerdas (*intelligent system*) yaitu *case based reasoning* (CBR) (Swari and Winarko, 2014). Konsep dasar dari pendekatan ini adalah dengan menggunakan data-data (yang diinterpretasikan sebagai pengalaman) terdahulu untuk menyimpulkan sebuah hasil dari kasus yang dihadapi saat ini. Kasus yang baru akan diadopsi solusi dari kasus-kasus yang berada dalam kelompok kasus yang sama (Made Hanindia Prami Swari, Rahel Widya Arianti and Faisal Muttaqin, 2020). Ilmu CBR merupakan rumpun dari bidang ilmu sistem pakar, dimana pada sistem pakar terdapat dua pendekatan, yakni pendekatan berbasis aturan (*rule based reasoning*) dan pendekatan berbasis kasus seperti yang diadopsi pada sistem CBR. Keunggulan dari pendekatan CBR adalah akan sangat bermanfaat karena dapat menghilangkan kebutuhan untuk mengekstrak model seperti yang dibutuhkan oleh sistem berbasis aturan (Tedy and Hartati, 2012).

Teori similaritas merupakan teori utama yang dipakai dalam penelitian ini. Hasil prediksi akan didasarkan pada kasus lampau yang memiliki similaritas terbesar dengan kasus baru yang diuji. Untuk menghitung nilai similaritas maka digunakan pendekatan Euclidean Distance. Pendekatan ini dipilih karena cocok dengan karakteristik atribut yang digunakan pada basis kasus serta berdasarkan beberapa penelitian terdahulu mampu memberikan akurasi yang sangat baik, yakni 99,1% pada penelitian yang berjudul Aplikasi Case Based Reasoning untuk Identifikasi Serangan Hama pada Tanaman Jeruk (Silmina and Wardoyo, 2018) dan sebesar 94,83% pada penelitian berjudul Case Based Reasoning untuk Diagnosis Penyakit Jantung (Wahyudi and Hartati, 2017).

Data yang digunakan pada penelitian ini adalah 141 data mahasiswa program studi informatika dan sistem informasi yang terdapat pada sistem Pendaftaran Yudisium yang ada di lingkungan Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur. Atribut yang digunakan meliputi Indeks Prestasi Semester (IPS) 1-5, Indeks Prestasi Kumulatif (IPK) semester 5, Satuan Kredit Semester (SKS) yang ditempuh di semester 5, Keaktifan dalam Organisasi Jurusan (BEM, HIMA, Komunitas, dan Organisasi Non Kemahasiswaan lainnya), Jalur Masuk (SMNPTN, SBMPTN, Mandiri), serta asal daerah (Surabaya, Non Surabaya). Untuk menaggulangi bentuk data yang tidak beragam serta dengan jangkauan yang lebih dari rentang 0-1, maka akan dilakukan proses normalisasi menggunakan pendekatan *min-max*.

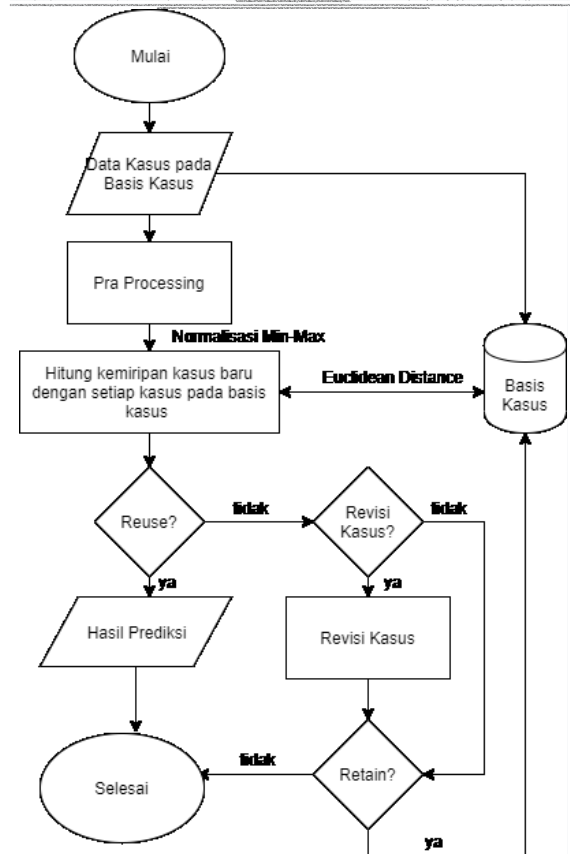
Beberapa penelitian terkait prediksi waktu studi mahasiswa telah banyak dilakukan pada penelitian-penelitian terdahulu namun penelitian sebelumnya

kebanyakan menggunakan pendekatan pada konsep *data mining* dan klasifikasi, misalnya seperti yang telah dilakukan pada penelitian berjudul Komparasi Data Mining Naive Bayes dan Neural Network memprediksi Masa Studi Mahasiswa S1. Penelitian ini bertujuan untuk prediksi kelulusan mahasiswa di STMIK Widya Cipta Dharma (WiCiDa) Samarinda. Atribut yang digunakan yaitu, umur saat masuk kuliah, klasifikasi kota asal Sekolah Menengah Atas, pekerjaan ayah, program studi, kelas, jumlah saudara, dan Indeks Prestasi Kumulatif (IPK) (Azahari *et al.*, 2020). Penelitian tentang prediksi kelulusan mahasiswa S1 juga pernah dilakukan oleh Ryan Dwi Pambudi dkk dengan judul penelitian Prediksi Kelulusan Mahasiswa Berdasarkan Kinerja Akademik Menggunakan Pendekatan Data Mining Pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya. Penelitian ini menghasilkan sebuah sistem prediksi dengan menggunakan algoritma klasifikasi Naive-Bayes menggunakan tools WeKa (Pambudi, Supianto and Setiawan, 2019). Penelitian lain yang pernah dilakukan terkait prediksi kelulusan S1 adalah penelitian yang dilakukan oleh Mongan Winny Amelia yang berjudul Prediksi Masa Studi Mahasiswa dengan Menggunakan Algoritma Naive Bayes. Penelitian ini menggunakan atribut Indeks Prestasi Semester (IPS), Indeks Prestasi Kumulatif (IPK), serta jumlah SKS tempuh. Akurasi yang dihasilkan sistem dari penelitian ini adalah sebesar 85,17% (Amelia, Lumenta and Jacobus, 2017). Penelitian terakhir terkait prediksi kelulusan mahasiswa dilakukan pada penelitian berjudul Aplikasi Prediksi Kelulusan Mahasiswa Dengan Metode Linear Discriminant Analysis yang dilakukan oleh Sulistio. Cara penentuan skor prediksi didapat dengan menggunakan metode analisa diskriminan. Sebelum menentukan metode diskriminan, akan dilakukan terlebih dahulu perhitungan korelasi antara masing masing mata kuliah dengan status kelulusan, mata kuliah yang akan dipilih adalah mata kuliah dengan nilai korelasi ≥ 0.5 atau korelasi positif sedang. Penelitian ini menghasilkan nilai akurasi sebesar 97% dalam melakukan prediksi kelulusan mahasiswa (Sulistio, 2017).

Perbedaan mendasar dari penelitian ini dengan penelitian-penelitian yang diacu adalah pada pendekatan yang digunakan, dimana penelitian-penelitian terdahulu menggunakan pendekatan data mining dan penggunaan teori komputasi, sedangkan penelitian ini mendasarkan hasil prediksi pada kasus-kasus yang pernah terjadi pada masa lampau. Perbedaan lain adalah penggunaan atribut yang lebih beragam pada penelitian ini. Melalui pembuatan sistem prediksi berbasis CBR ini, maka diharapkan pihak akademik dapat mendeteksi mahasiswa-mahasiswa yang kemungkinan mengalami keterlambatan dalam penyelesaian masa studinya, sehingga dapat dipantau dan diambil tindakan yang dianggap perlu.

2. METODE PENELITIAN

Secara umum, langkah-langkah yang dilakukan pada penelitian ini adalah melakukan pengumpulan dan memasukkan data kasus pada basis kasus, melakukan pra-prosesing yakni normalisasi atribut yang akan digunakan dalam perhitungan similaritas antar kasus menggunakan normalisasi min-max, implementasi CBR menggunakan metode Euclidean Distance, serta melakukan pengujian sistem. Alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur Sistem Prediksi Kelulusan Mahasiswa

2.1. Pengumpulan Data Basis Kasus

Data mahasiswa yang dijadikan pada basis kasus pada penelitian ini didapatkan dari Tata Usaha di Fakultas Ilmu Komputer Universitas Pembangunan Nasional "Veteran" Jawa Timur serta data yang telah dimiliki pada sistem pendaftaran yudisium fakultas. Data yang dikumpulkan merupakan data jalur masuk mahasiswa, IPS semester 1 sampai 5, IPK semester 5, keaktifan organisasi dalam jurusan, daerah asal, serta predikat kelulusan sehingga terdapat 20 buah atribut yang digunakan dalam proses CBR yang akan diproses pada proses prediksi kelulusan mahasiswa. Data yang terkumpul sejumlah 141 data kasus yang terdiri dari mahasiswa angkatan 2015 dan 2016 yang telah dinyatakan lulus.

2.2. Pra Processing Data

Nilai similaritas mempunyai rentang antara 0-1. Nilai 0 berarti tidak ada kemiripan antara kasus baru dengan kasus lama yang diuji, sebaliknya nilai 1 mengindikasikan bahwa kasus baru sangat mirip dengan kasus lama yang diuji, sehingga kesimpulan dari kasus lama tersebut akan dipercaya untuk dijadikan solusi pada kasus baru. Untuk dapat menghasilkan nilai dengan rentang 0-1 tersebut, maka harus dilakukan pra processing pada atribut-atribut yang tidak berada pada rentang tersebut. Adapun atribut yang dinormalisasi adalah sebagai berikut :

2.2.1 Atribut Daerah Asal

Daerah asal merupakan data yang bertipe kategori (bukan berupa angka), sehingga perlu dilakukan transformasi agar berbentuk angka. Atribut daerah asal diklasifikasikan kedalam 2 kelas, yakni daerah asal Surabaya (diberi kode 1) dan daerah asal Luar Surabaya (diberi kode 2). Kode 1 bukan berarti bahwa mahasiswa yang berasal dari Surabaya memiliki nilai lebih kecil dari mahasiswa yang berasal dari daerah luar Surabaya. Nilai atribut akan dikonversi menjadi nilai 0 apabila mahasiswa pada kasus lama yang diuji berasal dari Surabaya namun mahasiswa pada kasus baru berasal dari Luar Surabaya.

2.2.2 Atribut Organisasi Kemahasiswaan

Atribut organisasi kemahasiswaan dibedakan menjadi BEM (diberi kode 1), Himpunan Mahasiswa Program Studi (diberi kode 2), Komunitas (diberi kode 3), dan tidak mengikuti organisasi (diberi kode 4). Proses normalisasi pada atribut ini menggunakan pendekatan yang sama dengan normalisasi pada atribut daerah asal, dimana nilai 1 akan diberikan jika mahasiswa pada basis kasus yang diuji mengikuti organisasi kemahasiswaan yang sama dengan data pada kasus baru yang diuji, begitu pula sebaliknya nilai 0 akan diberikan jika mahasiswa pada basis kasus yang diuji tidak bernilai sama dengan data mahasiswa pada kasus baru.

2.2.3 Atribut Jalur Masuk

Terdapat 3 kategori yang dimiliki oleh atribut jalur masuk, yakni melalui jalur SMNPTN (diberi kode 1), jalur SBMPTN (diberi kode 2), dan jalur Mandiri (diberi kode 3). Proses normalisasi pada atribut ini juga menggunakan pendekatan yang sama dengan atribut daerah asal serta atribut organisasi kemahasiswaan yang diikuti.

2.2.4 Atribut IPS, IPK, dan Jumlah SKS

Proses normalisasi dengan menggunakan pendekatan Min-Max baru akan digunakan untuk melakukan normalisasi pada atribut IPS dan IPK. Hal ini dikarenakan atribut-atribut ini merupakan atribut numerik yang memiliki nilai komputasi. Selain itu, rentang nilai dari atribut-atribut ini juga bernilai lebih dari 1 (batas maksimal nilai similaritas), sehingga atribut-atribut ini perlu dikonversi agar perhitungan similaritas nantinya dapat berada di rentang 0 hingga 1. Adapun persamaan normalisasi min-max yang digunakan terdapat pada persamaan 1.

$$N = \frac{MinRange+(X-MinValue)(MaxRange-MinRange)}{MaxValue-MinValue} \quad (1)$$

Dimana

N = Normalisasi Min-Max

MinRange = Nilai Konversi terkecil yang ditentukan

MaxRange = Nilai Konversi terbesar yang ditentukan

MaxValue = Nilai terbesar pada atribut yang dibandingkan

MinValue = Nilai terkecil pada atribut yang dibandingkan

Proses normalisasi untuk seluruh atribut yang digunakan pada penelitian ini yang meliputi atribut daerah asal, keikutsertaan organisasi, jalur masuk, IPS, IPK, dan jumlah SKS diilustrasikan pada Tabel 1.

Tabel 1. Atribut yang belum ternormalisasi

No	Kriteria	Ks01	K02	Ks03	Kb01
1	IPS Smt 1	3,2	3,3	3,1	3,2
2	IPS Smt 2	3,1	3,4	4	3,1
3	IPS Smt 3	2,99	3,25	3,3	3,1
4	IPS Smt 4	3,2	2,99	3,4	3,2
5	IPS Smt 5	3	3,1	3,4	3,2
6	SKS Smt 1	20	21	23	20
7	SKS Smt 2	22	20	21	22
8	SKS Smt 3	21	22	24	21
9	SKS Smt 4	23	22	21	23
10	SKS Smt 5	24	24	20	24
11	IPK Smt 5	3,09	3,20	3,44	3,12
12	SNMPTN	0	1	0	0
13	SBMPTN	1	0	0	1
14	Mandiri	0	0	1	0
15	BEM	0	0	1	1
16	Himpunan	0	0	1	0
17	Komunitas	1	0	0	1
18	Tidak Ikut	0	1	0	0
19	Surabaya	0	1	0	0
20	Luar Sby	1	0	1	1

Keterangan :

Ks01 : Data kasus nomor 1 pada basis kasus

Ks02 : Data kasus nomor 2 pada basis kasus

Ks03 : Data kasus nomor 3 pada basis kasus

Kb01 : Data kasus baru yang dicari similaritasnya

Berikut merupakan contoh perhitungan normalisasi Min-Max pada atribut IPS Smt 1 pada Kasus Baru (Kb01).

Diketahui :

MinRange = 0

MaxRange = 1

X = 3,2

MinValue = 3,1

MaxValue = 3,3

$$Normalisasi = \frac{0 + (3,2 - 3,1) * (1 - 0)}{3,3 - 3,1} = 0,5$$

Proses ini juga dihitung untuk atribut IPS Semester 2, IPS Semester 3, IPS Semester 4, IPS Semester 5, SKS Semester 1, SKS Semester 2, SKS Semester 3, SKS Semester 4, SKS Semester 5, dan IPK Semester 5 pada seluruh atribut kasus yang terlibat. Setelah dilakukan proses normalisasi, maka

terbentuk tabel hasil normalisasi seperti yang terlihat pada Tabel 2.

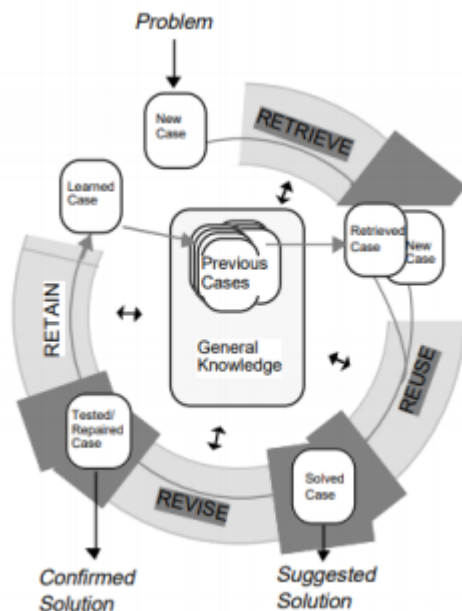
Tabel 2. Hasil Normalisasi Atribut

No	Kriteria	Ks01	K02	Ks03	Kb01
1	IPS Smt 1	0,5	1	0	0,5
2	IPS Smt 2	0	0,33	1	0
3	IPS Smt 3	0	0,83	1	0,35
4	IPS Smt 4	0,51	0	1	0,51
5	IPS Smt 5	0	0,25	1	0
6	SKS Smt 1	0	0,33	1	0
7	SKS Smt 2	1	0	0,5	1
8	SKS Smt 3	0	0,33	1	0
9	SKS Smt 4	1	0,5	0	1
10	SKS Smt 5	1	1	0	1
11	IPK Smt 5	0	0,32	1	0,06
12	SNMPTN	0	1	0	0
13	SBMPTN	1	0	0	1
14	Mandiri	0	0	1	0
15	BEM	0	0	1	1
16	Himpunan	0	0	1	0
17	Komunitas	1	0	0	1
18	Tidak Ikut	0	1	0	0
19	Surabaya	0	1	0	0
20	Luar Sby	1	0	1	1

Seperti yang terlihat pada Tabel 2, maka seluruh atribut yang akan digunakan pada proses perhitungan similaritas dalam sistem CBR telah berada pada rentang 0 sampai 1, sesuai yang dibutuhkan pada perhitungan similaritas menggunakan pendekatan Euclidean Distance.

2.3. Implementasi Sistem CBR

Case Based Reasoning memiliki 4 tahapan didalamnya seperti yang terlihat pada Gambar 2.



Gambar 2. Siklus CBR (Aamodi and Plaza, 1994)

Berdasarkan siklus CBR yang tergambar pada Gambar maka secara umum langkah-langkah yang dilakukan dalam melakukan proses prediksi menggunakan CBR adalah sebagai berikut :

2.3.1 Proses Retrieval

Proses retrieval merupakan proses utama dalam melakukan prediksi kelulusan pada sistem yang dibangun pada penelitian ini. Pada proses retrieval dilakukan proses pencarian nilai similaritas antar kasus yang terdapat pada basis kasus dengan kasus baru yang akan diprediksi. Proses perhitungan similaritas antar kasus dilakukan menggunakan rumus euclidean distance seperti yang tercantum pada persamaan 2.

$$SimED(S, T) = \frac{\sum_{i=1}^n f(S_i, T_i)^2 * W_i^2}{\sum_{i=1}^m W_i^2} \quad (2)$$

Keterangan :

- $SimED(S, T)$ = Similaritas Euclidean Distance
- $f(S_i, T_i)$ = Fitur ke-i dari kasus pada basis kasus (*source*) dan kasus baru yang diuji (*Target*)
- n = jumlah atribut pada *Target*
- m = jumlah atribut pada *Source*
- i = atribut individu, antara 1 s/d n
- W_i = bobot fitur ke-i

Proses retrieval yang menggunakan rumus jarak (*distance*) memiliki sedikit perbedaan dalam pembacaan hasil rekomendasi yang dihasilkan sistem, dimana jarak terbesar berarti bahwa antar kasus memiliki nilai similaritas terkecil, begitu pula sebaliknya, nilai jarak terkecil memiliki arti bahwa kedua kasus memiliki similaritas terbesar. Sehingga dapat disimpulkan bahwa kasus dengan hasil perhitungan euclidean distance terkecil merupakan kasus yang akan direkomendasikan sebagai hasil prediksi sistem. Untuk mempermudah pembacaan hasil, dimana similaritas identik dengan penggunaan nilai terbesar sebagai hasil rekomendasi, maka sistem yang dibangun akan menampilkan hasil similaritas menggunakan metode Euclidean Distance dengan memodifikasi pendekatan menggunakan pendekatan pada persamaan 3.

$$Similarity = 1 - SimED(S, T) \quad (3)$$

2.3.2 Proses Reuse

Prose Reuse pada sistem Case Based Reasoning merupakan proses untuk menggunakan kembali pengetahuan dan informasi yang telah didapatkan pada kasus lama lalu menggunakan solusi dari kasus terpilih tersebut sebagai solusi pada kasus baru yang diuji . Kasus yang terpilih merupakan kasus yang memiliki nilai similaritas terbesar berdasarkan proses retrieval menggunakan metode Euclidean Distance.

2.3.3 Proses Revise

Pada sebuah sistem Case Based Reasoning, sangat dimungkinkan dihasilkan atau direkomendasikannya solusi yang tidak relevan atau memenuhi kebutuhan pengguna saat ini. hal ini dapat disebabkan oleh beberapa hal, misalnya terdapat anomali kasus yang disimpan pada basis kasus ataupun kualitas basis kasus yang belum baik. Untuk

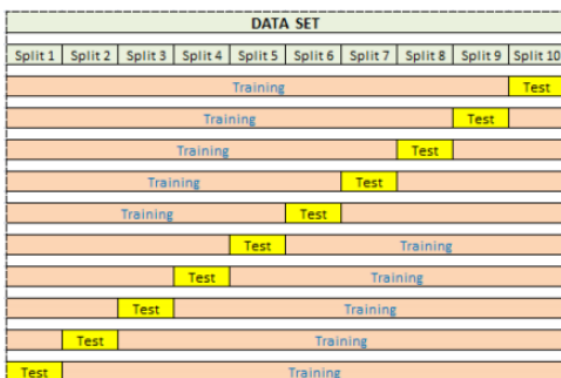
menanggulangi kasus-kasus tersebut, maka sebuah sistem CBR dapat dilengkapi dengan fitur revisi hasil (*revise*). Proses revisi pada penelitian ini dirancang masih dilakukan secara manual oleh pakar yang berkepentingan, dalam hal ini pakar yang dimaksud adalah Wakil Dekan I bidang kemahasiswaan di Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur. Sistem yang dibangun menyediakan halaman yang berisi hasil prediksi untuk setiap kasus baru yang diuji untuk dapat ditinjau oleh Wadep I dan diputuskan apakah kasus tersebut dapat dinilai valid atau membutuhkan proses revisi lebih lanjut.

2.3.4 Proses Retain

Proses retain dilakukan setelah proses revise dilakukan oleh Wadep I. kasus-kasus baik yang belum atau telah diperbaiki dapat dimasukkan sebagai basis kasus baru setelah divalidasi oleh Wadep I. Proses revise ini merupakan salah satu proses penting dalam CBR yang digunakan dalam penambahan pengetahuan bagi sistem yang dibangun, mengingat letak kecerdasan atau pengetahuan dari sebuah sistem CBR bergantung pada kekayaan basis kasus yang dimiliki.

2.4. Pengujian Sistem

Untuk mengetahui akurasi dari sistem prediksi yang dibuat, maka dilakukan pengujian menggunakan metode *k-fold cross validation*. Metode ini bekerja dengan mengelompokkan data secara acak kedalam beberapa (k) bagian sesuai kebutuhan dan pada setiap bagian yang terbentuk akan dilakukan proses klasifikasi (Han and Kamber, 2006). Nilai k yang digunakan pada penelitian ini adalah 10, dimana dalam proses pengujian nantinya data training akan dibagi menjadi 10 kelas. Masing-masing kelas memiliki data uji yang berbeda, sehingga seluruh data yang terdapat pada basis kasus akan diuji secara bergiliran pada akhir proses pengujian. Gambar 3 merupakan ilustrasi proses pengujian menggunakan skema *10-fold cross validation*.



Gambar 3. *10-fold cross validation* (Han and Kamber, 2006)

Setelah dilakukan pengujian menggunakan *10-fold cross validation*, maka kemudia dihitung nilai akurasi, presisi, serta sensitivitas sistem

menggunakan persamaan 4, 5, dan 6 (Zhu, Zeng and Wang, 2010).

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \tag{4}$$

$$Presisi = \frac{TP}{TP+FP} * 100\% \tag{5}$$

$$Sensitivitas = \frac{TP}{TP+FN} * 100\% \tag{6}$$

Keterangan :

True Positive (TP) : Data positif yang diprediksi benar

True Negative (TN) : Data negatif yang diprediksi benar

False Positive (FP) : Data negatif yang diprediksi sebagai data positif

False Negative (FN) : Data positif namun diprediksi sebagai data negatif.

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan 141 data kasus mahasiswa yangtelah lulus dari Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur. Pada masing-masing kasus terdapat 20 atribut yang akan diperbandingkan antara kasus lama yang telah disimpan pada basis kasus dengan kasus baru yang akan diuji.

3.1. Proses Perhitungan Similaritas antar Kasus

Proses retrieval dilakukan menggunakan pendekatan Euclidean Distance.

Tabel 3. Contoh Data yang akan Dihitung Nilai Similaritasnya

No	Kriteria	W	Ks01	K02	Ks03	Kb01
1	IPS Smt 1	1	0,5	1	0	0,5
2	IPS Smt 2	1	0	0,33	1	0
3	IPS Smt 3	1	0	0,83	1	0,35
4	IPS Smt 4	1	0,51	0	1	0,51
5	IPS Smt 5	1	0	0,25	1	0
6	SKS Smt 1	1	0	0,33	1	0
7	SKS Smt 2	1	1	0	0,5	1
8	SKS Smt 3	1	0	0,33	1	0
9	SKS Smt 4	1	1	0,5	0	1
10	SKS Smt 5	1	1	1	0	1
11	IPK Smt 5	1	0	0,32	1	0,06
12	SNMPTN	1	0	1	0	0
13	SBMPTN	1	1	0	0	1
14	Mandiri	1	0	0	1	0
15	BEM	0,5	0	0	1	1
16	Himpunan	0,5	0	0	1	0
17	Komunitas	0,5	1	0	0	1
18	Tidak Ikut	0,5	0	1	0	0
19	Surabaya	1	0	1	0	0
20	Luar Sby	1	1	0	1	1
	Masa Studi		9 Smt	7 Smt	8 Smt	?
	Ketepatan		Tidak	Tepat	Tepat	?
	Predikat		SM	SM	DP	?

Nilai W merupakan nilai bobot yang diberikan oleh pakar (bidang akademik) yang menunjukkan pengaruh suatu kriteria terhadap hasil prediksi, dimana semakin tinggi nilai W pada suatu kriteria

maka kriteria tersebut dianggap memiliki pengaruh lebih tinggi terhadap hasil prediksi. Sebelum menghitung nilai similaritas, maka dilakukan praprosesing data menggunakan pendekatan Min-Max untuk membuat seluruh atribut yang terlibat berada pada rentan 0-1. Tabel 3 merupakan tabel yang memuat atribut dari 3 basis kasus yang akan dihitung nilai similaritasnya dengan kasus baru yang akan diprediksi.

Berdasarkan data kasus pada tabel 3, maka dihitung nilai similaritas masing-masing kasus yang terdapat pada basis kasus (Ks01, Ks02, dan Ks03) dengan kasus yang diuji (Kb01) dengan hasil sebagai berikut :

a. Similaritas Ks01 dengan Kb01

SimED (Ks01, Kb01)

$$= \sqrt{\frac{0 + 0 + 0,12 + 0 + 0 + 0 + 0 + 0 + 0 + 0}{0,0036 + 0 + 0 + 0 + 0 + 0,5 + 0 + 0 + 0 + 0 + 0}}{\frac{1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1}{1 + 1 + 0,25 + 0,25 + 0,25 + 0,25 + 1 + 1}}$$

= 0,19

Similarity = 1 – 0,19 = 0,81

b. Similaritas Ks02 dengan Kb01

SimED (Ks02, Kb01)

$$= \sqrt{\frac{0,25 + 0,11 + 0,23 + 0,26 + 0,06 + 0,11 + 1 + 0,11 + 0,25 + 0 + 0,07 + 1 + 1 + 0 + 0,5 + 0 + 0,5 + 0,5 + 1 + 1}{1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1}}{\frac{1 + 1 + 0,25 + 0,25 + 0,25 + 0,25 + 1 + 1}{1 + 1 + 0,25 + 0,25 + 0,25 + 0,25 + 1 + 1}}$$

= 0,68

Similarity = 1 – 0,68 = 0,32

c. Similaritas Ks03 dengan Kb01

SimED (Ks01, Kb01)

$$= \sqrt{\frac{0,25 + 1 + 0,42 + 0,24 + 1 + 1 + 0,25 + 1 + 1 + 1 + 0,88 + 0 + 1 + 0,5 + 0 + 0,5 + 1 + 0 + 0 + 0}{1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1}}{\frac{1 + 1 + 0,25 + 0,25 + 0,25 + 0,25 + 1 + 1}{1 + 1 + 0,25 + 0,25 + 0,25 + 0,25 + 1 + 1}}$$

= 0,81

Similarity = 1 – 0,81 = 0,19

Berdasarkan hasil perhitungan similaritas yang dilakukan, maka didapatkan bahwa nilai jarak terkecil (similaritas terbesar) adalah pada kasus 1, dengan demikian sistem CBR akan merekomendasikan prediksi bahwa mahasiswa pada kasus baru akan menempuh perkuliahan selama 9 semester yang berarti tidak tepat waktu dengan predikat kelulusan sangat memuaskan.

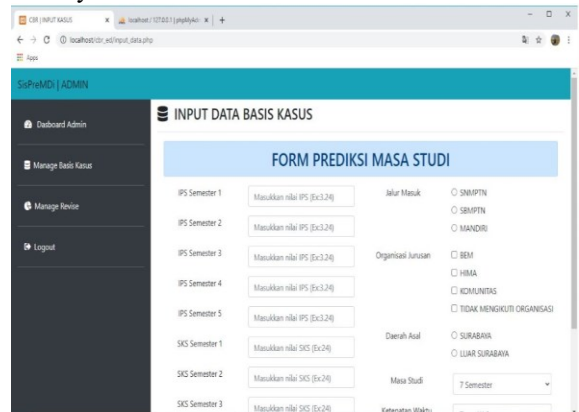
3.2. Implementasi Sistem Prediksi

Sistem prediksi menggunakan *pendekatan case based reasoning* ini dibuat berbasis web menggunakan bahasa pemrograman PHP dan basis data MSQl. Terdapat 2 fungsi utama pada sistem ini, yakni proses memasukkan basis kasus dan proses pengisian atribut untuk melakukan rekomendasi

3.2.1 Proses Pembentukan Basis Kasus

Berkebalikan dengan sistem pakar berbasis pengetahuan, dimana letak kemampuan dalam melakukan tugasnya bergantung pada basis pengetahuan yang dimiliki, maka pada sistem CBR basis kasus merupakan penentu keberhasilan performa sistem. Semakin banyak dan bervariasi basis kasus yang dimiliki, maka kemungkinan sistem mampu memberikan rekomendasi yang akurat akan semakin meningkat.

Pengetahuan sistem dalam bentuk basis kasus akan disimpan sedemikian rupa pada basis data pada sistem yang dibangun. Gambar 4 merupakan tampilan proses memasukkan kasus-kasus yang menjadi basis kasus pada sistem yang dibangun. Kasus awal yang dimasukkan pada sistem adalah sejumlah 141 data mahasiswa di Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur yang telah menyelesaikan studinya.

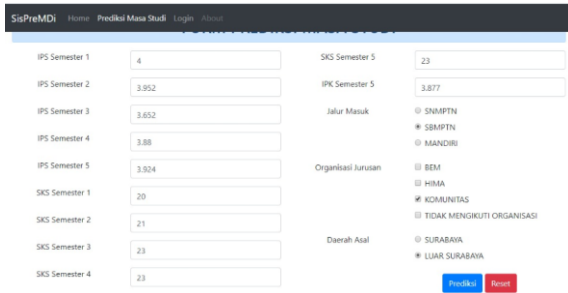


Gambar 4. Proses Pengisian Basis Kasus

Kasus-kasus dapat dimasukkan ke dalam basis kasus menggunakan *form* sesuai Gambar 4. Setelah seluruh kasus dimasukkan maka sistem prediksi sudah dapat digunakann untuk memprediksi kasus baru.

3.2.2 Proses Rekomendasi

Untuk melakukan proses rekomendasi, maka pengguna sistem harus memasukkan seluruh atribut yang digunakan sebagai parameter dalam mencari jarak terdekat (nilai similaritas terbesar) antara kasus baru dengan kasus lama. Gambar 5 merupakan halamn yang digunakan untuk memasukkan data atribut yang akan diprediksi.



Gambar 5. Proses Memasukkan Atribut Kasus Baru

Setelah pengguna selesai memasukkan seluruh atribut yang dibutuhkan, maka sistem akan menghitung nilai jarak dan similaritas antar kasus baru yang diprediksi dengan seluruh kasus yang telah disimpan pada basis kasus. Gambar 6 merupakan hasil rekomendasi sistem.



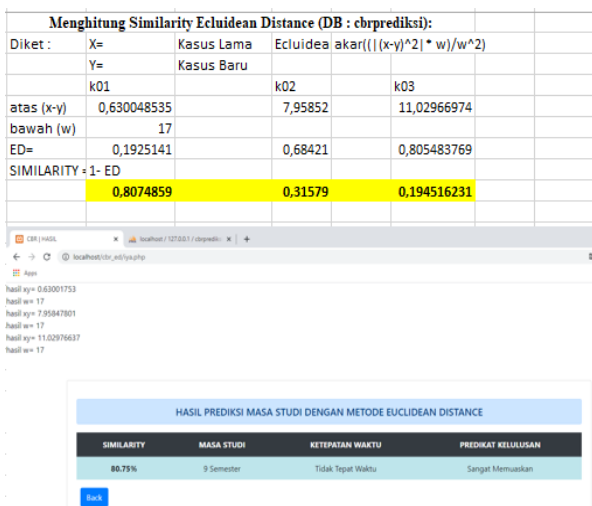
Gambar 6. Hasil Prediksi Sistem

3.3. Pengujian Sistem

Terdapat 2 jenis pengujian yang dilakukan pada penelitian ini, yakni pengujian untuk memastikan bahwa hasil perhitungan sistem sama persis dengan perhitungan manual, serta melakukan pengujian akurasi hasil rekomendasi sistem menggunakan 10 *k-fold cross validation*.

3.4 Pengujian Manual dan Sistem

Pengujian ini dilakukan untuk menjamin bahwa sistem rekomendasi yang dibangun sudah menghasilkan nilai similaritas sesuai teori yang digunakan. Gambar 6 merupakan hasil perbandingan perhitungan manual dengan hasil rekomendasi sistem pada perhitungan similaritas Ks01 dengan Kb01.



Gambar 6. Perbandingan Perhitungan Manual dan Hasil Prediksi Sistem

Berdasarkan hasil pengujian yang dilakukan, maka dapat disimpulkan bahwa sistem telah mampu menghasilkan hasil rekomendasi sistem yang sama persis dengan perhitungan manual.

3.3. Pengujian Akurasi

Terdapat 3 parameter yang digunakan pada proses pengujian akurasi sistem. Tabel 4 merupakan hasil pengujian sistem dari sisi akurasi.

Tabel 4. Pengujian Akurasi Sistem

Pengujian	Berdasarkan Ketepatan Waktu	Berdasarkan Masa Studi	Berdasarkan Predikat
1	71,43%	57,14%	100%
2	71,43%	50%	100%
3	71,43%	71,43%	100%
4	57,14%	50,00%	78,57%
5	50%	35,71%	92,86%
6	71,43%	35,71%	93%
7	78,57%	35,71%	92,86%
8	50,00%	36%	78,57%
9	53%	26,67%	80%
10	85,71%	50%	92,85%
Rata-rata	66,05%	44,81%	91%

Berdasarkan hasil pengujian yang terdapat pada Tabel 4, maka dapat disimpulkan bahwa rata-rata akurasi sistem adalah sebesar 66,05% pada prediksi ketepatan waktu, sebesar 44,81% pada prediksi berdasarkan masa studi, dan sebesar 91% pada prediksi berdasarkan predikat kelulusan. Sedangkan hasil pengujian Presisi sistem terdapat pada Tabel 5.

Tabel 5. Pengujian Presisi Sistem

Pengujian	Berdasarkan Ketepatan Waktu	Berdasarkan Masa Studi	Berdasarkan Predikat
1	90,00%	34,26%	100%
2	90,90%	25,93%	100%
3	91%	45%	100%
4	78%	43,33%	72,72%
5	62,50%	35,56%	100,00%
6	76,92%	25,40%	100%
7	90,90%	35,55%	90%
8	66,66%	37,62%	77,77%
9	100%	26,67%	84,62%
10	83,33%	50%	100%
Rata-rata	82,98%	35,98%	92,51%

Hasil pengujian sistem dari sisi presisi menghasilkan nilai yang cukup baik yakni dengan rata-rata sebesar 82,98% dan 92,51% pada prediksi berdasarkan ketepatan waktu dan predikat kelulusan, namun menghasilkan nilai yang cukup kecil pada prediksi berdasarkan masa studi yang hanya menghasilkan nilai 35,98%. Pengujian terakhir yang dilakukan adalah pengujian sensitivitas pada Tabel 6.

Pada pengujian sensitivitas, sistem juga menghasilkan nilai yang baik pada prediksi berdasarkan ketepatan waktu dan predikat kelulusan, yakni menghasilkan nilai 73,56% dan 95,52%. Sedangkan pengujian berdasarkan masa studi menghasilkan hasil yang cukup kecil, yakni 32,66%.

Tabel 6. Pengujian Sensitivitas Sistem

Pengujian	Berdasarkan Ketepatan Waktu	Berdasarkan Masa Studi	Berdasarkan Predikat
1	75,00%	37,88%	100%
2	76,92%	19,44%	100%
3	76,92%	38,46%	100%
4	63,63%	31,11%	100%
5	55,55%	39,52%	90,90%
6	90,90%	31,11%	93%
7	83,33%	39,52%	100,00%
8	60%	33,73%	87,50%
9	53%	15,38%	91,66%
10	100%	40,48%	92,31%
Rata-rata	73,56%	32,66%	95,52%

4. KESIMPULAN

Setelah penelitian ini dilakukan, maka didapat beberapa kesimpulan diantaranya bahwa berdasarkan pengujian yang telah dilakukan pada sistem baik pada pengujian kesesuaian hasil perhitungan manual dengan sistem maupun pengujian akurasi memperoleh hasil yang baik. Dari sisi perhitungan akurasi, sistem mampu memberikan nilai akurasi paling tinggi sebesar 100% pada pada pengujian berdasarkan predikat kelulusan pada skenario uji ke 1 dan 2, sedangkan berdasarkan ketepatan waktu, sistem mampu memberikan akurasi tertinggi pada skenario uji ke 10 dengan nilai 85,71%, dan sistem mampu memberikan nilai akurasi tertinggi sebesar 71,43% pada pengujian berdasarkan massa studi. Untuk pengujian presisi, sistem mampu mengasilkan nilai terbesar berturut-turut sebesar 90,90%, 43,33%, dan 100%. Sedangkan pada pengujian sensitivitas, sistem berturut-turut mampu menghasilkan nilai sebesar 90,90%, 40,48%, dan 100%. Hasil pengujian ini tentunya sangat bergantung dari basis kasus yang dimiliki, oleh sebab itu perbaikan dan peningkatan jumlah kasus yang dimiliki diharapkan mampu meningkatkan performa sistem rekomendasi yang dibuat.

UCAPAN TERIMA KASIH

Peneliti mengucapkan terima kasih kepada LPPM dan UPN "Veteran" Jawa Timur yang telah mengusahakan dan membiayai penelitian dalam Program Penelitian Mandiri Skim Riset Dasar Lanjutan UPN "Veteran" Jawa Timur Tahun Anggaran 2020

DAFTAR PUSTAKA

- AMELIA, M. W., LUMENTA, A. S., JACOBUS, A., 2017 . Prediksi Masa Studi Mahasiswa dengan Menggunakan Algoritma Naïve Bayes. *Jurnal Teknik Informatika*, 11(1). doi: 10.35793/jti.11.1.2017.17652.
- AZAHARI, A., *ET AL.*, 2020. Komparasi Data Mining Naive Bayes dan Neural Network memprediksi Masa Studi Mahasiswa S1. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 7(3), p. 443. doi: 10.25126/jtiik.2020732093.
- SWARI, M.H.P.S., ARIANTI, R.W., MUTTAQIN,

F., 2020. Case-Based Reasoning Pemberian Rekomendasi Profesi Berdasarkan Minat Dan Bakat Siswa Menggunakan Simple Matching Coefficient Similarity. *SINTECH (Science and Information Technology) Journal*. doi: 10.31598/sintechjournal.v3i1.505.

NULHAQIM, S. A., *ET AL.*, 2016. Peranan Perguruan Tinggi Dalam Meningkatkan Kualitas Pendidikan Di Indonesia Untuk Menghadapi Asean Community 2015 Studi Kasus: Universitas Indonesia, Universitas Padjadjaran, Institut Teknologi Bandung. *Share : Social Work Journal*, 6(2), p. 197. doi: 10.24198/share.v6i2.13209.

PAMBUDI, R. D., SUPianto, A. A., SETIAWAN, N. Y., 2019. Prediksi Kelulusan Mahasiswa Berdasarkan Kinerja Akademik Menggunakan Pendekatan Data Mining Pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya'. 3(3), pp. 2194–2200.

Peraturan Menteri Pendidikan Dan Kebudayaan Nomor 03 Tahun 2020 Tentang Standar Nasional Perguruan Tinggi.

SILMINA, E. P., WARDOYO, R., 2018. Aplikasi Case Based Reasoning Untuk Identifikasi Serangan Hama Pada Tanaman Jeruk. *Transmisi*, 20(3), p. 96. doi: 10.14710/transmisi.20.3.96-104.

SULISTIO, S., 2017. Aplikasi Prediksi Kelulusan Mahasiswa Dengan Metode Linear Discriminant Analysis, *Computatio : Journal of Computer Science and Information Systems*, 1(1), p. 58. doi: 10.24912/computatio.v1i1.241.

SWARI, M. H. P., WINARKO, E., 2014. Sistem Diagnosis Penyakit Gigi Dan Mulut Menggunakan Kombinasi Case Based Reasoning Dan Rule Based Reasoning. Universitas Gadjah Mada. Available at: http://etd.repository.ugm.ac.id/home/detail_pencarian/74315.

TEDY, R., HARTATI, S., 2012. Case-Based Reasoning untuk Diagnosa Penyakit THT (Telinga Hidung dan Tenggorokan). *Indonesiann Journal of Computing and Cybernetics Systems*, 6(2), pp. 67–78. Available at: <https://jurnal.ugm.ac.id/ijccs/article/view/2154/1934>.

WAHYUDI, E., HARTATI, S., 2017. Case-Based Reasoning untuk Diagnosis Penyakit Jantung. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 11(1), p. 1. doi: 10.22146/ijccs.15523.

Zhu, W., Zeng, N., Wang, N., 2010. Sensitivity, specificity, accuracy, associated

confidence interval and ROC analysis with practical SAS® implementations. Northeast SAS Users Group 2010: Health Care and Life Sciences, pp. 1–9. Available at:
<https://www.lexjansen.com/nesug/nesug10/hl/hl07.pdf>.