

Optimasi Model Prediksi Kelulusan Mahasiswa Menggunakan Algoritma *Naive Bayes*

Hartatik

D3 Teknik Informatika Sekolah Vokasi Universitas Sebelas Maret

hartatik119@staff.uns.ac.id

Info Artikel

Kata Kunci :

Prediksi kelulusan, *naive bayes*, sistem informasi, indeks prestasi, prestasi akademik, distribusi gaussian

Keywords :

graduation prediction, naive bayes, grade point, academic achievement, gaussian distribution

Tanggal Artikel

Dikirim : 17 September 2020

Direvisi : 9 November 2020

Diterima : 30 November 2020

Abstrak

Prediksi tentang status kelulusan mahasiswa menjadi persoalan tersendiri di perguruan tinggi. Perguruan tinggi utamanya di era Big Data sangatlah penting untuk melakukan prediksi perilaku akademik mahasiswa aktif sehingga dapat di ketahui kemungkinan mahasiswa bisa studi secara tepat waktu serta dapat diketahui langkah preventive dalam membuat prgram perencanaan. Salah satu cara yang digunakan adalah teknik data mining yaitu menggunakan Algoritma *naive bayes*. Algoritma *Naive bayes* merupakan salah satu metode yang digunakan untuk memprediksi kelulusan mahasiswa. Peneliti dalam hal ini menerapkan metode *Naive bayes* menggunakan parameter Indeks prestasi kumulatif(IPK) dan membandingkan dengan menggunakan prediksi *naive bayes methods* berdasarkan parameter IPK dan sosial parameter yaitu jenis kelamin dan status tinggal. Dalam penelitian ini menggunakan parameter akademis dan dilakukan optimasi menggunakan parameter sosial yang melekat pada mahasiswa. Berdasarkan hasil evaluasi untuk mendapatkan akurasi, hasil dari penelitian ini mendapatkan nilai akurasi untuk metode *Naive bayes* sebesar 75% dan akurasi untuk model prediksi dengan parameter sosial sebesar 85% dengan selisih akurasi 10%.

Abstarct

Predictions about a student's graduation status are a problem in college. Major tertiary institutions in the era of Big Data are very important to predict the behavior of active students so that they can find out the possibility of students in a timely manner and can determine preventive steps in making program planning. One method used is data mining techniques using the Naive bayes Algorithm. The Naive bayes algorithm is one of the methods used to predict student graduation. Researchers in this case applied the Naive bayes method using the cumulative achievement index (GPA) parameter and compared using the prediction of the Naive bayes method based on the GPA parameters and social parameters, namely gender and status. This study uses academic parameters and is carried out optimally using social parameters inherent in students. Based on the results of the evaluation to get an accuracy value, the results of this study get an accurate value for the Naive bayes method of 75% and accurate for prediction models with social parameters of 85% with a difference of 10%.

1. PENDAHULUAN

1.1 Latar Belakang

Teknologi informasi dan komunikasi yang kita nikmati sekarang pada awalnya dimulai dengan hal-hal yang sederhana. Untuk berkomunikasi dan bertukar informasi pada awalnya orang menggunakan surat, telegram, atau dengan menggunakan tulisan tangan dan mesin ketik untuk menuliskannya. Perkembangan yang sangat pesat di dalam teknologi dapat dinikmati oleh semua orang. Banyak ditemukan penerapan teknologi baik di berbagai bidang dalam kehidupan. Aktivitas belajar mengajar menggambarkan aktivitas yang bernilai untuk dilakukan dalam suatu proses pembelajaran. Dimana dalam pelaksanaannya terjadi interaksi antara murid dengan guru. Pada interaksi tersebut, terjadilah pertukaran informasi berupa ilmu pengetahuan dan norma yang diterapkan dalam lingkup pendidikan.

Keberhasilan mahasiswa dalam suatu proses belajar merupakan suatu nilai yang bisa diukur parameter keberhasilannya. Apabila nilai akhir mahasiswa baik maka dapat dikatakan adanya suatu keberhasilan dalam proses belajar mengajar artinya ada suatu kesuksesan transfer pengetahuan dan skill. Bisa dikatakan bahwa nilai akhir yang baik oleh mahasiswa tentunya mencerminkan nilai yang baik pada mata kuliah dan hal ini menggambarkan adanya keberhasilan dalam hal pengetahuan, keahlian dan kemampuan mahasiswa [1].

Permasalahan yang timbul dari hasil proses perkuliahan yaitu tidak semua proses yang terjadi menghasilkan hasil yang baik seperti yang diinginkan. Hasil yang tidak diinginkan dapat diperoleh akibat dari beberapa faktor, diantaranya adalah tidak adanya pantauan prestasi akademik yang baik dari institusi. Apabila hal tersebut terjadi, dapat menyulitkan mahasiswa untuk lulus tepat waktu. Selain itu juga dapat mempengaruhi kualitas program studi yang bersangkutan. Prestasi akademik menjadi hal yang utama yang menjadikan parameter keberhasilan suatu pendidikan. Salah satu indikator tercapainya tujuan tersebut adalah dari hasil prestasi akademik mahasiswa yang dinyatakan dengan Indek Prestasi Semester (IPMS) dan Indek Prestasi Kumulatif (IPK).

Indek Prestasi Semester ialah nilai prestasi akademik mahasiswa dengan seluruh mata kuliah yang sudah ditempuh untuk setiap semester tertentu. serta Indek Prestasi Kumulatif ialah prestasi akademik mahasiswa dengan mengkombinasikan seluruh mata kuliah yang sudah ditempuh hingga dalam semester tertentu [11]. Untuk mewujudkan pendidikan yang berkualitas berkaitan dengan peran dosen, motivasi mahasiswa, kedisiplinan mahasiswa, sosial ekonomi mahasiswa dan juga hasil belajar masa lalu. Data ini berpotensi untuk menghasilkan informasi baru yang berguna. Salah satu yang dapat dilakukan oleh data mining adalah memprediksi prestasi akademik mahasiswa [4] [9]. Jika prestasi akademik mahasiswa dapat di ketahui lebih dini maka pihak program studi dapat melakukan tindakan yang diperlukan agar mahasiswa dapat mencapai prestasi akademik yang baik. Harapan akhirnya adalah semua mahasiswa dari berbagai latar belakang faktor dapat maksimal dalam meningkatkan prestasi akademiknya. Berdasarkan penjelasan di atas maka fokus penelitian ini adalah memprediksi prestasi akademik mahasiswa dengan menggunakan metode data mining klasifikasi berdasarkan peran dosen, motivasi mahasiswa, kedisiplinan mahasiswa, sosial ekonomi mahasiswa dan juga hasil belajar masa lalu.

Studi tentang kelulusan dan prestasi mahasiswa sangatlah penting baik bagi mahasiswa, orang tua maupun program studi. Telah banyak dikembangkan penelitian tentang prediksi prestasi mahasiswa yaitu riset yang dikembangkan oleh Hendra, Mochammad Abdul Azis serta Suhardjono tentang Analisis Kelulusan Mahasiswa dengan menggunakan teknik pengambilan keputusan *Decision Tree* [8]. Atribut yang digunakan dalam penelitian ini sebagai indikator prestasi mahasiswa adalah IPMS semester 1,2,3,4. Dalam penelitian ini selanjutnya fokus pada algoritma *naive bayes*, dengan menggunakan model baru yaitu menambahkan variabel baru selain variabel IPMS 1,2,3,4 yaitu variabel nilai UN, jenis kelamin dan status tinggal mahasiswa. Penelitian ini selanjutnya akan membandingkan model prediksi waktu kelulusan mahasiswa dengan metode *naive bayes* berdasarkan parameter prestasi akademik yaitu IPK dan berdasarkan gabungan prestais akademik dan sosial parameter yaitu Jenis kelamin dan status tinggal. Tujuan dari penelitian ini adalah melakukan prediksi Perkembangan Prestasi Mahasiswa dengan Algoritma *Naive bayes* dengan parameter nilai akademik dan melakukan prediksi Perkembangan Prestasi Mahasiswa dengan Algoritma *Naive bayes* dengan parameter sosial parameter yang dibawa mahasiswa sebelum proses pembelajaran.

2. METODE PENELITIAN

Dalam penelitian ini pendekatan yang dilakukan yaitu kuantitatif, yaitu penelitian dilaksanakan berdasarkan data pengamatan berdasarkan populasi dan sampel. Teknik pengambilan sampel dilakukan dengan mengambil beberapa mahasiswa pada angkatan tertentu dan dalam program studi tertentu yang mana selanjutnya diterapkan untuk menguji hipotesa [4][5]. Adapun langkah langkah penelitian ini adalah:

1) Mengumpulkan hasil pengamatan

Ditahap ini dicari hasil pengamatan yang ada, memperoleh hasil pengamatan tambahan yang dibutuhkan, mengintegrasikan semua hasil pengamatan kedalam dataset.

2) Seleksi data

Setelah hasil pengamatan diperoleh, data selanjutnya diseleksi untuk membuat sebuah target data, fokus dalam bagian dari variabel atau sampel data yang mana penelitian akan dilakukan.

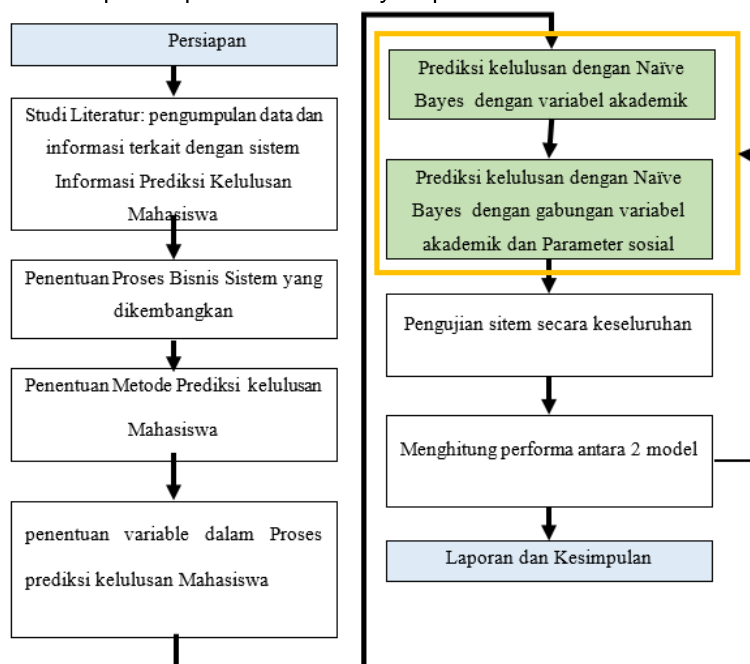
- 3) Teknik penelitian yang digunakan
 Pada tahap ini dianalisis hubungan antara satu data dengan data yang lainnya yang saling keterkaitan. Setelah analisis data diterapkan, selanjutnya dilakukan penerapan model yang sesuai dengan type data hasil amatan. Data hasil amatan terbagi dalam testing (data uji) dan data training (data latihan) yang digunakan untuk pembuatan model.
- 4) Penelitian dan pengujian algoritma
 Model yang digunakan dalam penelitian ini selanjutnya dilakukan uji sehingga diharapkan ada validasi dalam pengambilan keputusan.
- 5) Evaluasi
 Tahap evaluasi dilakukan terhadap model yang akan ditetapkan sehingga selanjutnya digunakan untuk mengetahui tingkat akurasi model.

Ada beberapa tahapan proses dalam melakukan analisa data guna proses prediksi. Pemodelan serta algoritma yang digunakan mengambil referensi dari beberapa penelitian yang sudah ada sebelumnya dengan tambahan variabel untuk meningkatkan tingkat akurasi dalam proses prediksi kelulusan mahasiswa sehingga model prediksi bisa digunakan di institusi terkait. Variabel yang digunakan seperti pada Tabel 1 berikut:

Tabel 1. Variabel penelitian

No	Variabel	Keterangan
1	Nim	No Induk Mahasiswa
2	Nama	Nama Mahasiswa
3	IPK	Indeks Prestasi Mahasiswa
4	Status	Status Kelulusan Tepat Waktu dan Tidak Tepat Waktu
5	JK	Jenis Kelamin
6	UN	Nilai UN
7	Status Tinggal	Dalam Kota/Luar Kota PT

Berikut alur penelitian dalam model prediksi prestasi mahasiswa yaitu pada Gambar 1 berikut:



Gambar 1. Gambar metode prediksi

Penelitian yang dilakukan adalah dengan tahap modeling dengan menggunakan metode *naive bayesian* dengan membandingkan 2 parameter yaitu parameter akademik selama mahasiswa dan sosial parameter yaitu jenis kelamin, status tinggal dan nilai UN. Diharapkan penambahan parameter ini bisa memberikan optimasi terhadap model prediksi. Langkah dalam algoritma bayes adalah [2][3]:

1. Menyiapkan data yang akan dihitung

Untuk perhitungan data nominal menggunakan persamaan (1) dibawah ini.

$$P(A | I) = \frac{P(I | A)P(A)}{P(I)} \quad (1)$$

- I = amatan dengan kelompok yang belum diketahui
- A = Hipotesis amatan X merupakan suatu kelompok spesifik
- P(A|I) = Probabilitas A berdasarkan kondisi x (*posteriori prob.*)
- P(I) = Probabilitas A (*prior prob.*)
- P(I|A) = Probabilitas I berdasarkan kondisi tersebut
- P(I) = Probabilitas dari I

2. Untuk perhitungan data numerik menggunakan distribusi gaussian pada persamaan (2).

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \exp \frac{-(x-\mu)^2}{2\sigma^2}$$

Dengan rata rata = $\frac{\sum_i^n x_i}{n}$ (2)

$$\text{standart deviasi} = \sqrt{\frac{\sum_i^n (x_i - \mu)^2}{n - 1}}$$

3. Melakukan pencarian *mean* data terlebih dahulu dengan melakukan pembagian antara jumlah data dan banyaknya data
4. Perhitungan selanjutnya dengan mencari standar deviasi (σ) dengan melakukan pembagian antara penjumlahan data dikurangi mean data dengan banyak data dikurangi satu.
Perhitungan gaussian akan menghasilkan probabilitas kasus pada suatu kelas.
5. Lakukan perhitungan jumlah kelas pada data training dengan membagi jumlah data pada suatu kelas dengan jumlah data keseluruhan.
6. Kemudian lakukan perhitungan jumlah kasus yang sama di suatu kelas yang sama. Dengan membagi jumlah data kasus yang sama pada suatu kelas dengan jumlah data pada kelas tersebut.
Kemudian melakukan perkalian semua hasil perhitungan yang memiliki kelas yang sama.
7. Hasil dari perkalian tersebut kemudian dibandingkan antar kelas yang diuji. Kelas dengan nilai probabilitas paling tinggi ditetapkan sebagai hasil prediksi.

3. HASIL DAN PEMBAHASAN

3.1 Perhitungan Algoritma *Naive bayes*

Perhitungan algoritma *naive bayes* menggunakan data observasi pada masa lalu [6][7]. Adalah simulasi perhitungan sistem menggunakan Microsoft Excel terhadap data latih yang tersedia menggunakan perhitungan algoritma *naive bayes*. Tabel data yang digunakan sebagai data latih untuk prediksi perkembangan prestasi mahasiswa dengan algoritma *naive bayes* dapat dilihat pada Tabel 2. Berdasarkan data Tabel 2, dalam pengujian perhitungan diperlukan data uji untuk melihat kesesuaian hasil dari pengujian. Data uji yang digunakan dapat dilihat pada Tabel 3.

Perhitungan dengan menggunakan microsoft excel dilakukan dengan menggunakan rumus matematika. Untuk data bertipe numerik seperti indeks prestasi semester dan nilai rata-rata ujian nasional maka digunakan perhitungan distribusi Gaussian. Sedangkan untuk data bertipe nominal yaitu asal daerah menggunakan perhitungan teorema bayes.

Perhitungan yang dilakukan dibagi menjadi dua, yaitu perhitungan tipe numerik dan nominal. Perhitungan tipe numerik yaitu sebagai berikut:

1. Mencari *mean* data (μ) dengan membagi jumlah data keseluruhan dengan banyaknya data.

Perintah Excel yang digunakan untuk mencari *mean* dari indeks prestasi semester 1 dengan status kelulusan tepat yaitu :
 =SUM(C2;C3;C4;C6;C9;C11;C12;C14;C15;C16;C17;C18;C19;C22;C23;C25;C24;C26;C27;C28;C29;C30)/COUNTIF(K2;K3
 1;"tepat")

2. Mencari standart deviasi (σ).

Perintah Excel yang digunakan untuk mencari standart deviasi dari indeks prestasi semester 1 dengan status kelulusan tepat yaitu :
 =SQRT(((C2-D41)^2+(C3-D41)^2+(C4-D41)^2+(C6-D41)^2+(C9-D41)^2+(C12-D41)^2+(C14-D41)^2+(C11-D41)^2+(C15-D41)^2+(C16-D41)^2+(C17-D41)^2+(C18-D41)^2+(C19-D41)^2+(C22-D41)^2+(C23-D41)^2+(C24-D41)^2+(C25-D41)^2+(C26-D41)^2+(C27-D41)^2+(C28-D41)^2+(C29-D41)^2+(C30-D41)^2)/(D40-1))

3. Melakukan perhitungan distribusi gaussian.

Perintah Excel yang digunakan untuk mencari probabilitas dengan melakukan perhitungan distribusi gaussian dari indeks prestasi semester 1 dengan status kelulusan tepat yaitu :

$$=(1/(E51*D42))*E52^(-1*((D59-D41)^2)/((2*D42)^2))$$

Berdasarkan perhitungan ini didapatkan nilai tepat lulus yaitu 0.7272 dan peluang lulus terlambat 0.2728 artinya prediksi untuk mahasiswa X3112053 adalah lulus tepat waktu..

Tabel 2. Data Latih Perkembangan Prestasi Mahasiswa Dengan Algoritma Naive bayes

No	NIM	JK	Nilai UN	TINGGAL	IPMS1	IPMS2	IPMS3	IPMS4	IPMS5	IPMS6	kelulusan
1	X3112001	L	51.55	Luar Kota	3.45	3.05	3.2	3.33	3.5	3.5	Tepat waktu
2	X3112002	P	50.6	Dalam Kota	3.3	3.76	3.46	3.33	3.73	3.8	Tepat waktu
3	X3112003	L	53.15	Luar Kota	3.35	3.33	3	3.19	3.5	2.87	Tepat waktu
4	X3112004	L	48.1	Luar Kota	3.4	3.24	3	3.38	2.27	4	Tepat waktu
5	X3112005	P	43.6	Luar Kota	3.1	3.48	3.14	3.19	3.36	3.8	Tepat waktu
6	X3112006	P	52.5	Luar Kota	3.6	3.38	3.36	3.71	3.73	3.8	Tepat waktu
....											
20	X3112020	L	51.75	Dalam Kota	3.3	3.76	3.46	3.33	3.73	3.8	Tepat waktu

Tabel 3. Tabel Data Uji Sistem Informasi Perkembangan Prestasi Mahasiswa

No	NIM	JK	Nilai UN	TINGGAL	IPMS1	IPMS2	IPMS3	IPMS4	IPMS5	IPMS6	kelulusan
1	X3112001	L	51.55	Luar Kota	3.45	3.05	3.2	3.33	3.5	3.5	Tepat waktu

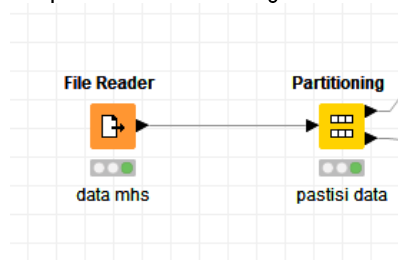
Berdasarkan perhitungan dengan algoritma *naive bayes* didapatkan hasil seperti ditunjukkan Tabel 4.:

Tabel 4. Hasil Perhitungan Prediksi prestasi Terhadap Data Uji

	Model 1	
	Probabilitas Tepat	Probabilitas Terlambat
Model 1	0.338057271	0.661942729
Model 2	0.192607694	0.807392306

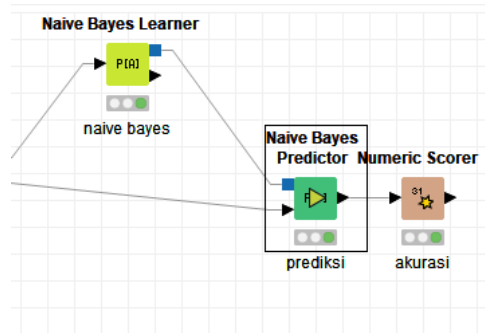
Dari perhitungan probabilitas kelulusan Tabel 4 di atas, dapat disimpulkan bahwa berdasarkan indeks prestasi, rata-rata ujian nasional, Jenis kelamin dan parameter asal daerah pada data uji, memiliki hasil akhir yaitu "Tepat Lulus" baik pada model 1 maupun model 2.

Model prediksi diimplementasikan untuk data besar (*big data*) pada data kurikulum. Dengan menggunakan teknik klasifikasi pada data mining untuk prediksi prestasi mahasiswa yaitu kelulusan mahasiswa dengan menggunakan program KNIME sebagai berikut Gambar 1 yaitu untuk implementasi data training dan data testing :



Gambar 1. Gambar partisi data latih

Pada Gambar 2 untuk implementasi algoritma *naive bayes* pada data besar.



Gambar 2. model prediksi prestasi mahasiswa

Hasil dari pengujian model yang dilakukan adalah optimasi parameter dalam algoritma *naive bayes* untuk memprediksi kelulusan mahasiswa. Dalam menentukan nilai tingkat akurasi dalam model prediksi dengan algoritma *naive bayes*. Model prediksi pada penelitian ini adalah model 1 yaitu model prediksi berdasarkan parameter akademis IPMS 1, IPMS2 dan Model 2 yaitu model prediksi berdasarkan parameter akademis IPMS 1, IPMS2, IPMS3, IPMS4, UN, status tinggal dan Jenis kelamin. Berikut dalam Tabel 5 merupakan komparasi hasil prediksi untuk model prediksi dengan menggunakan data nilai akademik IPK dan juga prediksi berdasarkan gabungan nilai akademik dan nilai sosial parameter sebagai dasar prediksi model.

Tabel 5. Komparasi ketepatan model 1 dan model 2

No	Modell	Data latih/ Data Training	Tepat Waktu	Terlambat	Akurasi
1	Model 1	70%	76%	24%	77%
		80%	82%	18%	78%
2	Model 2	70%	85%	15%	85%
		80%	88%	12%	89%

Berdasarkan hasil penelitan yang dilakukan untuk memecahkan hasil prediksi prestasi mahasiswa yaitu kelulusan mahasiswa didapatkan nilai komparasi model 1 dan model 2 adalah sebesar 77% dan 85% . Hal ini berarti bahwa parameter sosial yaitu UN, Jenis kelamin dan status tinggal mahasiswa mempunyai pengaruh terhadap prestasi mahasiswa khususnya waktu tempuh studi mahasiswa.

4. KESIMPULAN

Dalam penelitian ini ada 2 model yang digunakan untuk prediksi prestasi mahasiswa yaitu model 1 yang menekankan pada nilai akademik selama mahasiswa menempuh studi dan model 2 yaitu menggabungkan antara model 1 dan parameter sosial mahasiswa yaitu nilai UN, JK dan status tinggal mahasiswa. Hasil perhitungan dan analisis akurasi dari model 2 didapatkan hasilnya lebih tinggi dibandingkan dengan model 1 artinya bahwa parameter sosial yaitu UN, Jenis kelamin dan status tinggal mahasiswa mempunyai pengaruh terhadap prestasi mahasiswa khususnya waktu tempuh studi mahasiswa dan bisa meningkatkan optimasi performance dari prediksi prestasi mahasiswa yaitu dari 75% menjadi 85%.

DAFTAR PUSTAKA

- [1] Siagan, S. P., "Manajemen Sumber Daya Manusia". Jakarta: Bumi Aksara, 2003.
- [2] Budi, S., "Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis". 2007.
- [3] Bramer, M., "Principles of Data Mining", vol. 98, no. 461. London: Springer London, 2016.
- [4] Aryasanti, A., "Sistem Komparasi *Naive bayes* dan Decision Trees untuk Menentukan Klasifikasi Kegagalan Studi Mahasiswa", Jurnal TICOM, Vol. 6, No. 3, 2018, pp. 103-111, 2018.
- [5] Jananto, A., "Algoritma *Naive bayes* untuk Mencari Perkiraan Waktu Studi Mahasiswa". Jurnal Teknologi Informasi DINAMIK, 18(1), 09-16, 2013.

- [6] Farida, I. N., & Maulidina, M., "Pengembangan Sistem Evaluasi Akademik Berdasarkan Predikat Nilai Indeks Prestasi Mahasiswa". *Jurnal Ilmiah Teknik Informatika*, 12(2), 2018.
- [7] Hartatik Aziz, Abdul Hartono, Rudi et al, "Decision Support System for Detection of Skin Diseases in Smart Health development planning", *IOP Conference Series: Materials Science and Engineering Vol 858*, 2020.
- [8] Hartatik, Purnomo, A., Hartono, R., Munawaroh, H., "*Naive bayes* Approach for Expert System Design of Children Skin Identification Based on Android", *IOP Conference Series: Materials Science and Engineering Vol 333*, 2020.
- [9] Hendra, Mochammad Abdul Azis dan Suhardjono, "Analisis Prediksi Kelulusan Mahasiswa Menggunakan Decission Tree Berbasis Particle Swarm Optimization", *Jurnal SISFOKOM (Sistem Informasi dan Komputer)*, Volume 09, Nomor 01, PP 102 – 107, 2020 .
- [10] Nurrohmat, M. A., & Nugroho, Y. S., "Aplikasi Pemrediksi Masa Studi dan Predikat Kelulusan Mahasiswa Informatika Universitas Muhammadiyah Surakarta Menggunakan Metode *Naive bayes*". *Jurnal Ilmu Komputer dan Informatika*, 1(1), 29-34, 2015.
- [11] Magadalena. H, "Sistem Pendukung Keputusan Untuk Menentukan Mahasiswa Lulusan Terbaik Di Perguruan Tinggi". *Seminar Nasional Teknologi Informasi dan komunikasi*. ISSN : 2089-9815, 2012.