

Implementasi *Particle Swarm Optimization* (PSO) untuk Optimisasi Algoritma *Naive Bayes* dalam Memprediksi Mahasiswa Lulus Tepat Waktu

Sudriyanto¹, Rudi Rizaldi², M. Ainun Rofiq Hariri³

^{1,2,3} Fakultas Teknik, Universitas Nurul Jadid Paiton Probolinggo, Indonesia

Info Artikel

Riwayat Artikel

Diterima: 03-05-2021

Disetujui: 14-06-2021

Kata Kunci

Naive Bayes;

Kelulusan;

PSO;

Email*

sudriyanto@unuja.ac.id

ABSTRAK

Istitusi perguruan tinggi merupakan wadah bagi mahasiswa dalam mendapatkan pengetahuan sebelum terjun langsung dan bersaing dalam dunia kerja. Istitusi perguruan tinggi berperan dalam menciptakan lulusan yang sesuai dengan kebutuhan dunia industri. Salah satu indikator dalam keberhasilan perguruan tinggi salah satunya ialah jumlah kelulusan mahasiswa yang mana berdampak kepada penilaian masyarakat dan akreditasi dari pemerintah. Peneliti sudah banyak melakukan penelitian dalam memprediksi kelulusan mahasiswa untuk mengetahui lulus terlambat atau tepat waktu. Menurut kondisi tersebut metode data mining yang cocok digunakan adalah *classification*. Salah satu metode data mining adalah *Naive Bayes*. Dalam penelitian ini algoritma yang digunakan *Naive Bayes* dengan PSO (*Particle Swarm Optimization*) sebagai penyeleksi atribut. Dari dataset yang digunakan adalah sebanyak 384 *record* data mahasiswa dari smestet 1 sampai dengan semester 8 diperoleh nilai akurasi sebesar 89.46%. Penelitian ini mempunyai tujuan agar universitas bisa memprediksi mahasiswa yang berpotensi tidak lulus tepat waktu, yang kemudian universitas akan memberikan penanganan khusus ataupun peringatan kepada mahasiswa agar supaya mahasiswa tersebut bisa lulus tepat pada waktunya.

1. PENDAHULUAN

Institusi pendidikan tinggi sebagai penyelenggara pendidikan di jenjang yang lebih tinggi dari pendidikan dasar dan menengah di bidang pendidikan. Institusi perguruan tinggi dilaksanakan untuk menyiapkan peserta didik yang memiliki kemampuan profesional dan akademis yang bisa mengimplementasikan, menciptakan ilmu pengetahuan, mengembangkan, kesenian dan teknologi untuk menjadi anggota masyarakat [1]. Pendidikan tinggi mempunyai peran yang begitu besar untuk mencetak generasi muda dan sumber daya manusia yang siap menghadapi persaingan di era global dan berkompeten. Oleh karenanya, Sangat dibutuhkan sekali untuk meningkatkan kualitas perguruan tinggi.

Salah satu aspek penting dalam pengukuran kualitas untuk mengevaluasi keberhasilan institusi pendidikan bidang perguruan tinggi adalah mahasiswa. Mahasiswa merupakan unsur yang sangat penting karenan nantinya mahasiswa langsung akan terkoneksi dengan *stakeholder*. Pendidikan tinggi akan mendapat kepercayaan darai *stakeholder* untuk menampung lulusannya bilamana kualitas mahasiswa dengan lulusan yang baik. Unsur penilaiannya antara lain, nilai akademik, nilai ujian masuk, prestasi yang didapat mahasiswa, prosentase lulus tepat waktu dan kompetensi mahasiswa [2].

Bagi mahasiswa kelulusan merupakan langkah awal dalam penentuan suatu karir yang mana untuk menunjang ke jejang yang lebih tinggi apalagi untuk penentuan masa depan. Bahkan bagi institusi perguruan tinggi, tingkat kelulusan tepat waktu menjadi salah satu dari

standar akreditasi institusi [3].

Sejalan dengan melonjaknya minat terhadap pendidikan khususnya program studi teknik informatika di fakultas Teknik Universitas Nurul Jadid, maka dari itu universitas seharusnya sudah mempunyai tindakan khusus untuk menyeimbangkan jumlah yang masuk dengan mahasiswa yang sudah keluar. Salah satunya dengan cara melakukan pengawasan terhadap kelulusan mahasiswa tepat waktu.

Beberapa peneliti telah banyak melakukan penelitian dengan penggunaan data mining untuk menggali serta memperoleh model dari data yang sudah di ukur tingkat keakuratannya. Beberapa metode dalam data mining digunakan dalam memprediksi kelulusan tepat waktu berupa metode klasifikasi. Metode klasifikasi secara khusus membantu untuk secara otomatis dalam pengambilan keputusan untuk menemukan *utilitas* dalam aktivitas manusia. Metode klasifikasi diantaranya, *Naive Bayes* [4] [5], *decision tree* [6], *information gain* [5], *neural network* dan *rule learner* [7]. Pada penelitian ini akan menerapkan dua metode yaitu *Naive bayes* dan *optimalisasi Particle Swarm Optimization (PSO)* sebagai penentuan atribut mana saja yang akan dihilangkan [8], dalam memprediksi tingkat kelulusan mahasiswa tepat waktu. Algoritma *naive bayes* dipilih karena sangat baik dalam pengklasifikasian [9], dari metode ini dihasilkan pula model yang mudah bila diimplementasikan kedalam aplikasi.

Dalam penelitian ini peneliti menggunakan data mahasiswa angkatan 2019 dengan data kelulusan mahasiswa sebanyak 384 *record*. Dari penelitian yang dihasilkan diharapkan untuk mendapatkan model yang dapat diimplementasikan, untuk mendapatkan klasifikasi mahasiswa yang diprediksi akan tidak lulus tepat waktu. Sehingga pihak kampus mendapatkan tingkat kelulusan tepat waktu yang tinggi. Setelah didapat model serta proses yang sebelumnya, peneliti akan mengimplemintasikan dalam bentuk *software* yang berbasis web. Diharapkan aplikasi ini dapat membantu universitas dalam pengambilan sikap dan keputusan terhadap mahasiswa yang diprediksi lulus tidak tepat waktu.

2. METODE

Teknik penelitian ini mengacu kepada instrumen dan perilaku yang kita gunakan untuk melakukan penelitian seperti melaksanakan pengamatan, mencatat data, metode pengolahan data dan semacamnya [10]. Dalam penelitian ini peneliti memanfaatkan algoritma *naive bayes* dengan metode optimasi *particle swarm optimization (PSO)* dimana dataset yang digunakan dalam penelitian ini adalah data kelulusan mahasiswa angkatan tahun 2019 dari Program Studi Teknik Informatika, Fakultas Teknik, Universitas Nurul Jadid dengan dataset sebanyak 384 *record* data mahasiswa. Pada penelitian ini metode yang digunakan ialah model CRISP-DM [11], dengan langkah-langkahnya sebagai berikut :

a. Pemahaman bisnis (*Bussiness understanding*)

Perguruan tinggi dengan misi masing-masing bersaing dalam meningkatkan kelulusan mahasiswanya, yaitu dengan menghasilkan kelulusan yang menjadi anggota produktif dalam masyarakat dan memberi kontribusi kepada kesejahteraan dan kemajuan ekonomi bangsa. Selain dari itu, tiap-tiap perguruan tinggi tahu bahwa jumlah mahasiswa yang putus diartikan oleh lembaga pendidikan sebagai hilangnya pemasukan bagi lembaga [12].

b. Pemahaman data (*Data understanding*)

Wisuda tepat pada waktunya merupakan isu penting karena tingkat kelulusan merupakan dasar efektifnya dalam lembaga perguruan tinggi. Jika tingkat kelulusan terjadi penurunan yang signifikan dan terus menerus berkembang akan jadi permasalahan serius. Bahkan berpengaruh terhadap akreditasi perguruan tinggi tersebut. Maka dari itu perlu dilakukan evaluasi dan pemantauan secara berkala terhadap tingkat kecenderungan kelulusan mahasiswa. Jadi, prediksi kelulusan mahasiswa

sangatlah diperlukan agar dapat memberi pertimbangan dalam mengambil kebijakan oleh perguruan tinggi setelah mengetahui aset mereka yaitu mahasiswa kemungkinan akan terlambat untuk kelulusannya [13].

Dataset yang digunakan dalam pemodelan ini yang berupa data kelulusan mahasiswa angkatan tahun 2019 yang berjumlah 384 *record* dan terdiri dari 15 atribut, dengan 5 atribut tipe kategorikal dan 10 dengan tipe numerik.

1	NAMA	JENIS KELAMIN	STATUS MAHASISWA	UMUR	STATUS NIKAH	IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	IPS 6	IPS 7	IPS 8	IPK	STATUS KELU
2	AGUNG DUMADI	LAKI - LAKI	BEKERJA	23	BELUM MENIKAH	3,5	3,1	3,8	3,1	3,9	3,6	2,3	0	3,6	TEPAT
3	MARWAH AISAH	PEREMPUAN	MAHASISWA	27	BELUM MENIKAH	3,3	2,9	3,5	2,7	2,7	2,1	2,1	0	3	TEPAT
4	MUHAMMAD ALI	LAKI - LAKI	MAHASISWA	24	BELUM MENIKAH	2,9	2,8	3,5	2,8	2,9	2,7	2,6	0	3,1	TERLAMBAT
5	DINDA RETNO ARIYAN	PEREMPUAN	MAHASISWA	25	BELUM MENIKAH	2,7	2,6	3	3,1	3,4	3,1	2,7	0	3,2	TERLAMBAT
6	IDRIS MAULANA FIKRI	LAKI - LAKI	MAHASISWA	27	BELUM MENIKAH	3	2,5	3,1	2,5	2,6	2,9	2,2	1	2,9	TERLAMBAT
7	ERVINA DWI WIJAYAN	PEREMPUAN	MAHASISWA	25	BELUM MENIKAH	2,8	2,5	2,6	2,6	2,3	2,8	2,7	0,8	2,9	TERLAMBAT
8	JOKO SUSANTO	LAKI - LAKI	MAHASISWA	26	BELUM MENIKAH	3,5	3	3,3	3,2	3,1	3,3	3,5	0	3,3	TEPAT
9	GALANG MAHENDRA	LAKI - LAKI	MAHASISWA	27	BELUM MENIKAH	3,1	3	3,1	2,4	3	3,4	2,9	0	3,1	TEPAT
10	LAILATUL MUNAWAR	PEREMPUAN	BEKERJA	25	BELUM MENIKAH	2,8	1,9	2,1	1,1	1,4	1,7	1,8	2,4	2,6	TERLAMBAT
11	IMAM HANAFI	LAKI - LAKI	BEKERJA	24	BELUM MENIKAH	3,2	2,7	3,2	2,4	3	2,5	1,8	0	2,8	TEPAT
12	NOOR FARIDA	PEREMPUAN	BEKERJA	26	BELUM MENIKAH	3,6	3,5	3,4	2,9	3,3	3	2,2	0	3,4	TEPAT
13	ARIE RETNO WATI	PEREMPUAN	MAHASISWA	29	BELUM MENIKAH	2,7	2,7	2,9	3,1	2,9	2,6	2,9	0,5	2,8	TEPAT
14	SUDARSONO	LAKI - LAKI	BEKERJA	27	BELUM MENIKAH	2,5	2,9	2,1	2,6	2,6	2,4	2,6	2,2	2,8	TEPAT
15	IKA RIYANI	PEREMPUAN	MAHASISWA	25	BELUM MENIKAH	3,2	3,1	3,3	2,8	3,4	2,7	3,1	0	3,1	TEPAT
16	KHOTIMATUS SA'ADA	PEREMPUAN	BEKERJA	24	BELUM MENIKAH	3,1	3	3,2	3,3	3,4	3,1	3,4	3	3,2	TEPAT
17	LANI PUSPAWATI	PEREMPUAN	MAHASISWA	24	BELUM MENIKAH	3	2,7	2,2	2,9	2,3	2,6	1,5	2,1	2,5	TEPAT

Gambar 1. Dataset kelulusan mahasiswa dalam bentuk excel

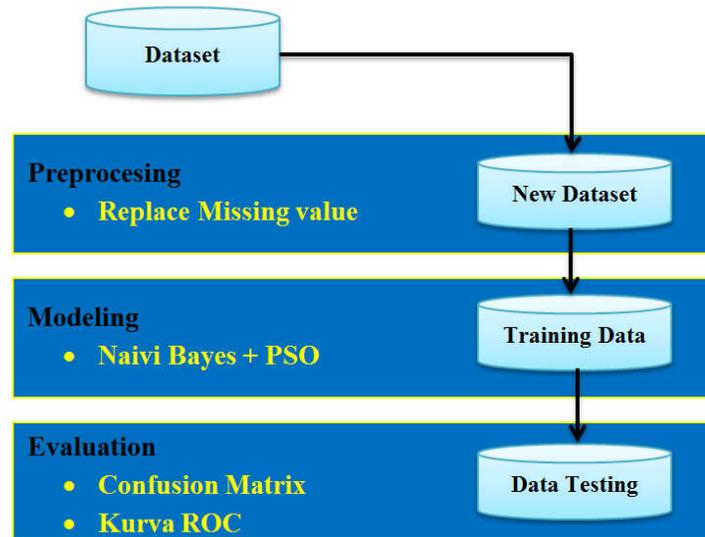
c. Pengolahan Data (*Data preparation*)

Pada tahapan ini, atribut dataset yang akan digunakan ialah sebagai berikut pada tabel 1 dibawah ini:

Tabel 1. Atribut yang digunakan dan kedudukannya

Atribut	Kegunaan	
NAMA	X	No
JENIS KELAMIN	√	regular (nilai model)
STATUS MAHASISWA	√	regular (nilai model)
UMUR	√	regular (nilai model)
STATUS NIKAH	√	regular (nilai model)
IPS 1	√	regular (nilai model)
IPS 2	√	regular (nilai model)
IPS 3	√	regular (nilai model)
IPS 4	√	regular (nilai model)
IPS 5	√	regular (nilai model)
IPS 6	√	regular (nilai model)
IPS 7	√	regular (nilai model)
IPS 8	√	regular (nilai model)
IPK	√	regular (nilai model)
STATUS KELULUSAN	√	Label(Hasil)

d. Pemodelan



Gambar 2. Model yang di usulkan

e. Evaluasi

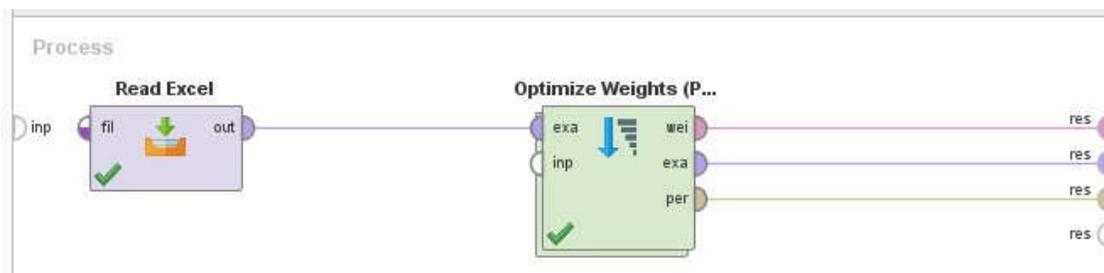
Pada tahapan ini akan dilakukan validasi dan proses mengukur keakuratan pada hasil yang telah tercapai oleh model dengan menggunakan teknik *confusion matrix* serta *kurva* ROC dalam mengukur akurasi model, dan untuk validasi menggunakan pengukuran *cross-validation*.

f. Penyebaran (*Deployment*)

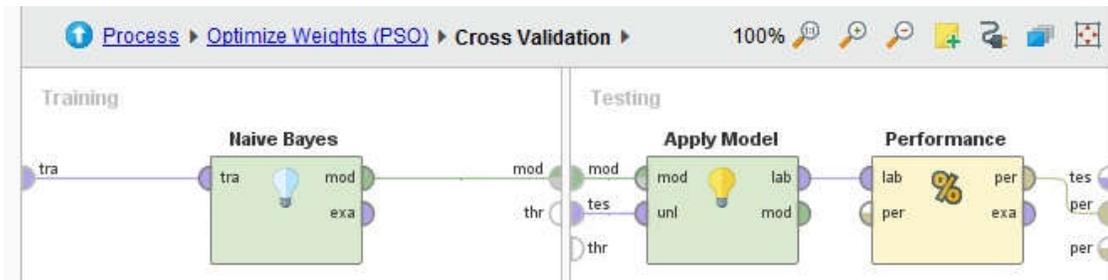
Hasil dari pemodelan ini, analisisnya mengarah kepada *Decision Support System* (DSS) bisa di manfaatkan oleh institusi perguruan tinggi dalam memberi pertimbangan dalam menentukan langkah selanjutnya terkait masalah kelulusan tidak tepat waktu dari mahasiswa yang berpengaruh terhadap hasil akreditasi institusi perguruan tinggi, nilai sosial kemasyarakatan, dan juga bisa dimanfaatkan untuk acuan pemodelan yang berikutnya serta juga bisa digunakan sebagai landasan dasar dalam perancangan suatu sistem.

3. HASIL DAN PEMBAHASAN

Proses desain model di implementasikan pada gambar 3 menggunakan *tool* rapidminer dengan jumlah dataset sebanyak 384 *record* data kelulusan mahasiswa angkatan tahun 2019 yang terdiri dari beberapa atribut diantaranya jenis_kelamin, nama mahasiswa, umur, status mahasiswa, status nikah, IPS smester 1 sampai 8, IPK dan yang terakhir status kelulusan.

Gambar 3. Implementasi model menggunakan *tool* rapidminer

Proses pada gambar 3 menunjukkan bahwa dalam penerapannya menggunakan model dari metode yang di gunakan dengan menggunakan *particle swarm optimization (PSO)* dalam penentuan bobot dari atribut. Pada *Optimize Weight (PSO)* digunakan 10 *k-fold validation* dengan memanfaatkan operator *Cross Validation* ditunjukkan pada gambar 4.



Gambar 4. Proses yang ada pada operator *Cross Validation*

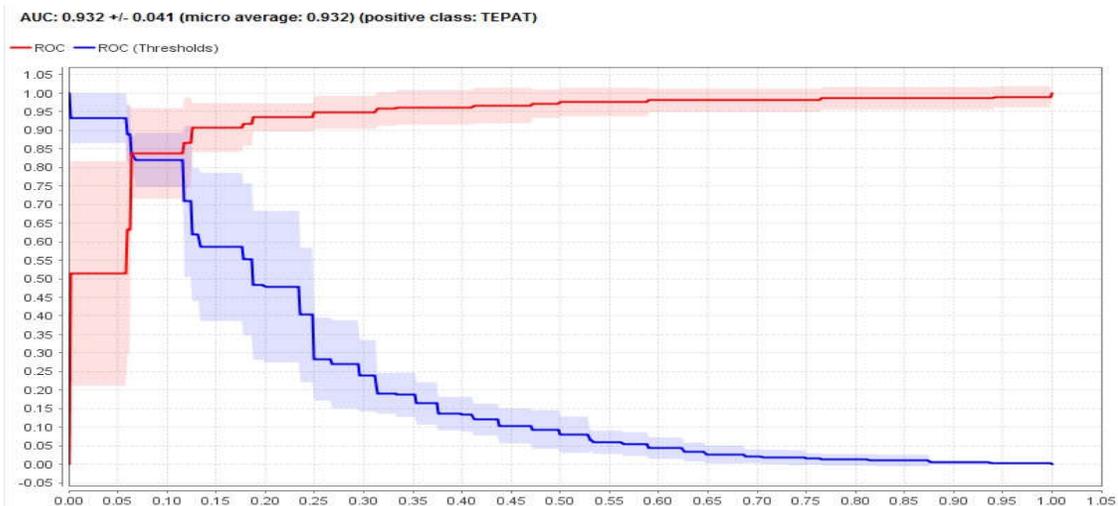
Dari proses implementasi pada gambar 4 diperoleh hasil pengujian akurasi menggunakan metode *Naive Bayes* berbasis *Particle Swarm Optimization (PSO)* dengan proses *10-fold validation* untuk kedua algoritma yang digunakan dinyatakan bahwa tingkat akurasi pada metode ini sebesar 89.46%, dimana *class precision* untuk pred. terlambat adalah sebesar 89.68%, pred. tepat adalah sebesar 89.29%, ditunjukkan pada gambar 5.

accuracy: 89.46% +/- 3.09% (micro average: 89.45%)

	true TERLAMBAT	true TEPAT	class precision
pred. TERLAMBAT	139	16	89.68%
pred. TEPAT	24	200	89.29%
class recall	85.28%	92.59%	

Gambar 5. Hasil hitung akurasi dari *Naive bayes-Particle Swarm Optimization (PSO)*

Sedangkan pada gambar 6 menunjukkan hasil pada diagram AUC pengolahan kurva dengan ROC menghasilkan nilai AUC 0.932 yang dikategorikan masuk dalam kategori *Fair Classification*.



Gambar 6. Grafik hasil dari performa AUC dari *Naive bayes - PSO*

Selain tabel *Confusion Matrix*, terdapat juga *performance vector* untuk metode *Naive Bayes* berbasis *Particle Swarm Optimization* (PSO) seperti pada gambar 7.

PerformanceVector

```

PerformanceVector:
accuracy: 89.46% +/- 3.09% (micro average: 89.45%)
ConfusionMatrix:
True:   TERLAMBAT   TEPAT
TERLAMBAT:   139    16
TEPAT:    24    200
precision: 89.36% +/- 3.15% (micro average: 89.29%) (positive class: TEPAT)
ConfusionMatrix:
True:   TERLAMBAT   TEPAT
TERLAMBAT:   139    16
TEPAT:    24    200
recall: 92.53% +/- 4.42% (micro average: 92.59%) (positive class: TEPAT)
ConfusionMatrix:
True:   TERLAMBAT   TEPAT
TERLAMBAT:   139    16
TEPAT:    24    200
AUC (optimistic): 0.932 +/- 0.041 (micro average: 0.932) (positive class: TEPAT)
AUC: 0.932 +/- 0.041 (micro average: 0.932) (positive class: TEPAT)
AUC (pessimistic): 0.932 +/- 0.041 (micro average: 0.932) (positive class: TEPAT)

```

Gambar 7. *Performance Vector Naive Bayes* berbasis *Particle Swarm Optimization* (PSO)

4. KESIMPULAN

Hasil dari pengolahan data kelulusan mahasiswa dan dilakukan pemodelan dengan metode *Naive Bayes* berbasis *Particle Swarm Optimization* (PSO) dengan menggunakan *framework* Rapidminer, dapat disimpulkan hasil sebagai berikut :

- a. Model algoritma dalam memprediksi kelulusan mahasiswa diperoleh menggunakan model algoritma *Naive Bayes* berbasis *Particle Swarm Optimization* (PSO)
- b. Hasil akurasi memprediksi kelulusan mahasiswa 89.46% Dimana class precision untuk pred. terlambat adalah 89.68%, pred. tepat adalah 89.29%. dengan prediksi tepat dan true tepat sebanyak 200 data.

5. SARAN

Saran penulis bagi penelitian yang selanjutnya disarankan beberapa hal sebagai berikut :

- a. Pada penelitian yang selanjutnya disarankan Jumlah dataset yang digunakan diharapkan agar lebih di perbanyak dari dataset yang sebelumnya untuk meningkatkan pemberian informasi terhadap layanan Akademik dan agar lebih meningkatkan keakuratan akurasi dari analisis yang dilakukan.
- b. Penelitian ini sangat perlu untuk penelitian lanjutan dan disarankan untuk menggunakan metode-metode yang lainnya sehingga bisa mengetahui berbagai macam-macam metode yang lainnya.

6. DAFTAR PUSTAKA

- [1] Imama Sabilla Wilda and Eranti Putri Tesa , "Prediksi Ketepatan Waktu Lulus Mahasiswa dengan kNearest Neighbor dan Naïve Bayes Classifier (Studi Kasus Prodi D3 Sistem Informasi Universitas Airlangga)," *Jurnal Komputer Terapan*, vol. III, no. 2, pp. 233-240, November 2017.

- [2] Mustafa M. Syukri , Rizky Ramadhan Muh , and P. Thenata Angelina , "Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Citec Journal*, vol. IV, no. 2, pp. 151-162, April 2017.
- [3] Maulida Rusyda and Bani, "Prediksi Kelulusan Mahasiswa Tepat Waktu dengan Algoritma C4.5 dengan Particle Swarm Optimization pada Univeristas XYZ," *Journal of Artificial Intelligence and Innovative Applications*, vol. I, no. 3, pp. 138-144, Agustus 2020.
- [4] Sutoyo Edi and Almaarif Ahmad , "Educational Data Mining untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritme Naïve Bayes Classifier," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. IV, no. 1, pp. 95 – 101, Februari 2020.
- [5] Muqorobin, Kusriani, and Taufiq Luthf Emha , "Optimasi Metode Naïve Bayes Dengan Feature Selection Information Gain Untuk Prediksi Keterlambatan Pembayaran Sumbangan Pembinaan Pendidikan Sekolah," *Jurnal Ilmiah Sinus (JIS)*, vol. 17, no. 01, pp. 2548-4028, Januari 2019.
- [6] Budiyantra Agus , Irwansyah, Prengki Egi , Ahmad Pratama Pandi , and Wiliani Ninuk , "Komparasi Algoritma Decision Tree, Naive Bayes Dan K-Nearest Neighbor Untuk Emprediksi Mahasiswa Lulus Tepat Waktu," *JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER*, vol. V, no. 2, pp. 2527-4864, Februari 2020.
- [7] Hamdani Nahrowi , Setyanto Arief , and Sudarmawan, "Perbandingan Algoritma Regresi Logistic Dan Neural Network Pada Prediksi Nilai Hasil Pembinaan Dan Kelulusan Tepat Waktu," *Jurnal Teknologi Informasi*, vol. XV, no. 1, pp. 30-36, Maret 2020.
- [8] Moh. Zainuddin, "Perbandingan 4 Algoritma Berbasis Particle Swarm Optimization (pso) Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa," *Jurnal Ilmiah Teknologi Informasi Asia*, vol. 13, no. 1, pp. 1-12, 2019.
- [9] Sri Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4.5, Naïve Bayes, KNN, Dan SVM," *Jurnal Tekno Insentif*, vol. 13, no. 1, pp. 16-25, April 2019.
- [10] Basias Nikolaos and Pollalis Yannis , "Quantitative and Qualitative Research in Business & Technology: Justifying a Suitable Research Methodology," *Review of Integrative Business and Economics Research*, vol. VII, no. 1, pp. 91-105, 2018.
- [11] Yahya Norzam and Jananto Arief , "Komparasi Kinerja Algoritma C.45 Dan Naive Bayes Untuk Prediksi Kegiatan Penerimaanmahasiswa Baru (Studi Kasus : Universitas Stikubank Semarang)," in *Prosiding SENDI_U*, Semarang, 2019, pp. 221-228.
- [12] Kurniawati Ika, Eko Indrajit Richardus , and Fauzi Muh. , "Peran Bussines Intelligence Dalam Menentukan Strategi Promosi Penerimaan Mahasiswa Baru," *IKRAITH-INFORMATIKA*, vol. I, no. 2, pp. 70-79, November 2017.
- [13] Agung Prakoso Satrio and Tias Tutik Eli , "Komparasi Algoritma C4.5 Dengan Naïve Bayes Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu Di Pts “Kzx”," *KOMPUTAKI*, vol. III, no. 1, pp. 103-132, Februari 2017.