

---

## ANALISIS PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGGUNAKAN METODE DATA MINING NAÏVE BAYES : *SYSTEMATIC REVIEW*

Lila Setiyani  
Mokhamad Wahidin  
Dudi Awaludin  
Sri Purwani

Program Studi Sistem Informasi STMIK ROSMA  
Jl. Kertabumi No.62 Karawang, Indonesia  
Email: lila.setiyani@dosen.rosma.ac.id, m.wahidin@dosen.rosma.ac.id,  
dudi@rosma.ac.id, sri.purwani@mhs.rosma.ac.id

### *Abstract*

This article attempts to systematically review the literatures of the data mining methods Naïve Bayes in predicting the timely graduation tertiary students. By comparing such literatures, it is found that this method can be used for the timely graduation prediction with 90% accuracy. This literature review is important to support the future research under this topic.

*Key words:* Data Mining, Educational Data Mining, Naïve Bayes, graduation prediction

### **Abstrak**

Artikel ini bertujuan untuk memberikan kajian literatur secara sistematis terhadap metode *data mining* Naïve Bayes dalam memprediksi jumlah kelulusan tepat waktu mahasiswa. Setelah membandingkan beberapa literatur, dapat disimpulkan bahwa metode ini dapat digunakan untuk prediksi tersebut dengan tingkat keakuratan 90%. Kajian literatur ini penting sebagai faktor pendukung bagi penelitian dengan topik itu.

**Kata Kunci:** Data Mining, Educational Data mining, Naïve Bayes, Prediksi Kelulusan

### **PENDAHULUAN**

Perguruan tinggi merupakan penyelenggara pendidikan akademik bagi mahasiswa. Lima perguruan tinggi diantaranya adalah universitas, institut, sekolah tinggi, akademik dan politeknik. Perguruan tinggi diharapkan menyelenggarakan pendidikan yang berkualitas bagi mahasiswa sehingga menghasilkan mahasiswa yang memiliki kompetensi dibidangnya. Di Indonesia sendiri kualitas suatu perguruan tinggi ditentukan oleh grade akreditasi yang di keluarkan oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT). Banyak aspek yang dapat dijadikan tolak ukur kualitas perguruan tinggi. Salah satunya dapat dilihat dari banyaknya mahasiswa yang dapat menyelesaikan studi sesuai dengan waktu yang ditentukan atau istilahnya mahasiswa lulus tepat waktu. Semakin banyak mahasiswa yang lulus tepat waktu maka semakin baik pula kinerja perguruan tinggi tersebut, sehingga tingkat kelulusan mahasiswa tepat waktu menjadi salah satu kriteria penilaian akreditasi bagi suatu perguruan tinggi atau program studi (Broto Legowo & Indiarso, 2017).

Banyak faktor yang dapat mendukung peningkatan jumlah kelulusan mahasiswa tepat waktu Antara lain:

1. Kualitas staf pengajar yang baik
2. Materi pembelajaran yang baik

3. Proses belajar-mengajar telah tertata dengan baik
4. Pengukuran keberhasilan proses pendidikan telah diterapkan
5. Administrasi pendidikan yang berjalan dengan baik
6. Sarana penunjang pendidikan tersedia dan mudah dimanfaatkan

Setiap perguruan tinggi memiliki database yang menyimpan data berupa data akademik dan biodata mahasiswa. Dengan melimpahnya database yang dimiliki oleh perguruan tinggi tersebut, informasi yang tersembunyi dapat diketahui dengan cara melakukan pengolahan terhadap data mahasiswa. Database tersebut dapat digunakan untuk mempelajari pola dan perilaku mahasiswa sehingga dapat meminimalisir keterlambatan kelulusan. Salah satu metode yang dapat digunakan untuk menganalisa database tersebut adalah metode data mining. Menurut Jing dan marceron(2017), data yang berlimpah membuka peluang diterapkannya data mining untuk pengelolaan pendidikan yang lebih baik dan data mining untuk pelaksanaan pembelajaran. Metode data mining dengan akurasi terbaik dapat digunakan sebagai pendukung keputusan bagi perguruan tinggi untuk mengambil kebijakan. Dukungan prediksi data mining dalam pengambilan keputusan ini diharapkan dapat meningkatkan jumlah kelulusan mahasiswa tepat waktu sesuai dengan peraturan Menteri Pendidikan(Ayub, 2018).

### **Data Mining**

Menurut Han dan Kamber (2011), data mining adalah proses menemukan pola yang menarik dan pengetahuan dari data yang berjumlah besar (Septiani, 2017).Sedangkan menurut Berry & Linoff (1997), data mining adalah suatu pencarian dan analisa dari jumlah data yang sangat besar dan bertujuan untuk mencari arti dari pola dan aturan(Chien, Wang, & Cheng, 2007). Lalu menurut Connolly (2005),data mining adalah suatu proses ekstraksi atau penggalian data yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari database yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting(Wulandari, Jatnika, & Purwanto, 2015).

Dari beberapa teori yang dijabarkan oleh ahli di atas maka dapat ditarik garis besar,data mining adalah suatu pencarian dan analisa pada suatu koleksi data (database) sehingga ditemukan suatu pola yang menarik dengan tujuan mengekstrak informasi dan pengetahuan yang akurat dan potensial,serta dapat dipahami dan berguna bagi pengambilan keputusan.

### **Fungsi Data Mining**

Menurut Maclennan ,Tang, & Crivat (2011), data mining memiliki beberapa fungsi yaitu (Saleh & Nasari, 2018) :

1. *Classification*  
Classification adalah proses untuk mencari model atau fungsi yang menggambarkan dan membedakan kelas – kelas atau konsep data.
2. *Clustering*  
Berfungsi untuk mencari pengelompokan atribut kedalam segmentasi – segmentasi berdasarkan similaritas.
3. *Associasion*  
Berfungsi untuk mencari keterkaitan antara atribut atau item set,berdasarkan item yang muncul dan rule asociasion yang ada.
4. *Regression*  
Bertujuan untuk mencari prediksi dari suatu pola yang ada.
5. *Forecasting*

Berfungsi untuk meramalkan waktu yang akan datang berdasarkan trend yang telah terjadi di waktu sebelumnya.

6. *Sequence Analysis*

Berfungsi untuk mencari pola urutan dari rangkaian kejadian.

7. *Deviation Analysis*

Berfungsi untuk mencari kejadian langka yang sangat berbeda dari keadaan normal (kejadian abnormal).

### Teknik Data Mining

Menurut Kantardzic (2009), teknik data mining yang paling umum, antara lain (Kantardzic, 2011):

1. Metode statistika klasik yaitu linear, quadratic dan logistic discriminate analyses.
2. Teknik statistika modern yaitu projection pursuit classification, density estimation, k-nearest neighbor, bayesian networks.
3. Artificial Neural Network (ANN), yaitu model sistematis yang meniru atau mensimulasikan struktur dan aspek fungsi dari jaringan saraf biologis.
4. Support Vektor Machine (SVM), yaitu rangkaian metode supervised learning yang digunakan untuk klasifikasi dan regresi.
5. Association Rules (AR), yaitu suatu metode riset untuk menemukan hubungan yang menarik antar variabel dalam suatu database.
6. Decision Tree (DT), yaitu tool pendukung suatu keputusan yang menggunakan graif seperti pohon atau model keputusan yang terdiri dari konsekuensi-konsekuensi.
7. Case Base Reasoning (CBS), yaitu proses untuk memecahkan suatu masalah baru berdasarkan solusi dari masalah – masalah lalu yang mirip.
8. Fuzzy Logic System (FLS), yaitu sebuah bentuk dari logika nilai ganda yang terkait dengan kesimpulan dari suatu alasan (reasoning) secara pendekatan. Logika fuzzy mempunyai nilai kebenaran diantara 0 dan 1.
9. Genetic Algorithms (GA), yaitu algoritma pencarian heuristic yang meniru proses evolusi alam (genetika), untuk mendapatkan solusi optimum.

Dalam systematic review ini focus pembahasan dibatasi pada literatur yang menggunakan fungsi data mining Classification dengan metode naïve bayes. Metode naïve bayes menurut Grainer (1998) memiliki beberapa kelebihan, yakni (Zatmiko, 2015):

1. Mudah untuk dipahami
2. Hanya memerlukan pengkodean yang sederhana
3. Lebih cepat dalam perhitungan

Merujuk dari kelebihan tersebut maka diharapkan prediksi kelulusan mahasiswa tepat waktu dengan menggunakan metode ini mendapatkan hasil dan akurasi yang baik.

### Tujuan dan Pertanyaan Penelitian

Tujuan dari keseluruhan review literatur ini adalah untuk mengidentifikasi tentang penggunaan metode data mining naïve bayes untuk memprediksi kelulusan mahasiswa tepat waktu. Oleh karena itu, dapat dirumuskan pertanyaan penelitian sebagai berikut :

1. Apakah Metode data mining naïve bayes dapat memprediksi kelulusan mahasiswa tepat waktu?
2. Seberapa besar akurasi yang didapat dari hasil metode data mining dalam memprediksi kelulusan mahasiswa tepat waktu?

3. Apakah atribut yang banyak digunakan dalam memprediksi kelulusan mahasiswa tepat waktu ?

Pada pertanyaan pertama, penulis mengidentifikasi penggunaan metode data mining naïve bayes untuk memprediksi kelulusan mahasiswa tepat waktu yang diperoleh dari systematic review. Hasil dari identifikasi tersebut membantu untuk menjawab pertanyaan kedua.

## **METODE**

Metode yang digunakan dalam melakukan systematic literature review ini berbasis pada protokol Preferred Reporting Item for Systematic Reviews and Meta-Analyses (PRISMA). Pada PRISMA terdapat tahapan yang detail dan lengkap untuk melakukan kajian suatu literatur. Literatur yang penulis bahas adalah literatur yang mengulas tentang prediksi kelulusan mahasiswa tepat waktu yang menggunakan metode data mining naïve bayes. Untuk literatur penulis menggunakan tipe literatur jurnal. Penulis membatasi topik literatur yang mengulas prediksi saja tidak untuk pembuatan aplikasi tertentu. Referensi literatur diambil dari mesin pencari literatur yaitu [garuda.ristekdikti.go.id](http://garuda.ristekdikti.go.id) dan [e-resources.perpusnas.go.id](http://e-resources.perpusnas.go.id). Penulis tidak membatasi penerbit tertentu. Penulis hanya menggunakan literatur yang terbit dari tahun 2015 demi mendapatkan literatur yang membahas metode terkini yang penulis akan ulas. Dalam mencari literatur penulis menggunakan kata kunci sebagai berikut:

1. Data mining & prediksi kelulusan
2. Naïve bayes & prediksi kelulusan
3. Naïve bayes & kelulusan
4. Naïve Bayes & prediksi akademik
5. Data mining & prediksi akademik
6. Naïve Bayes & kelulusan tepat waktu
7. Kelulusan & algoritma bayes

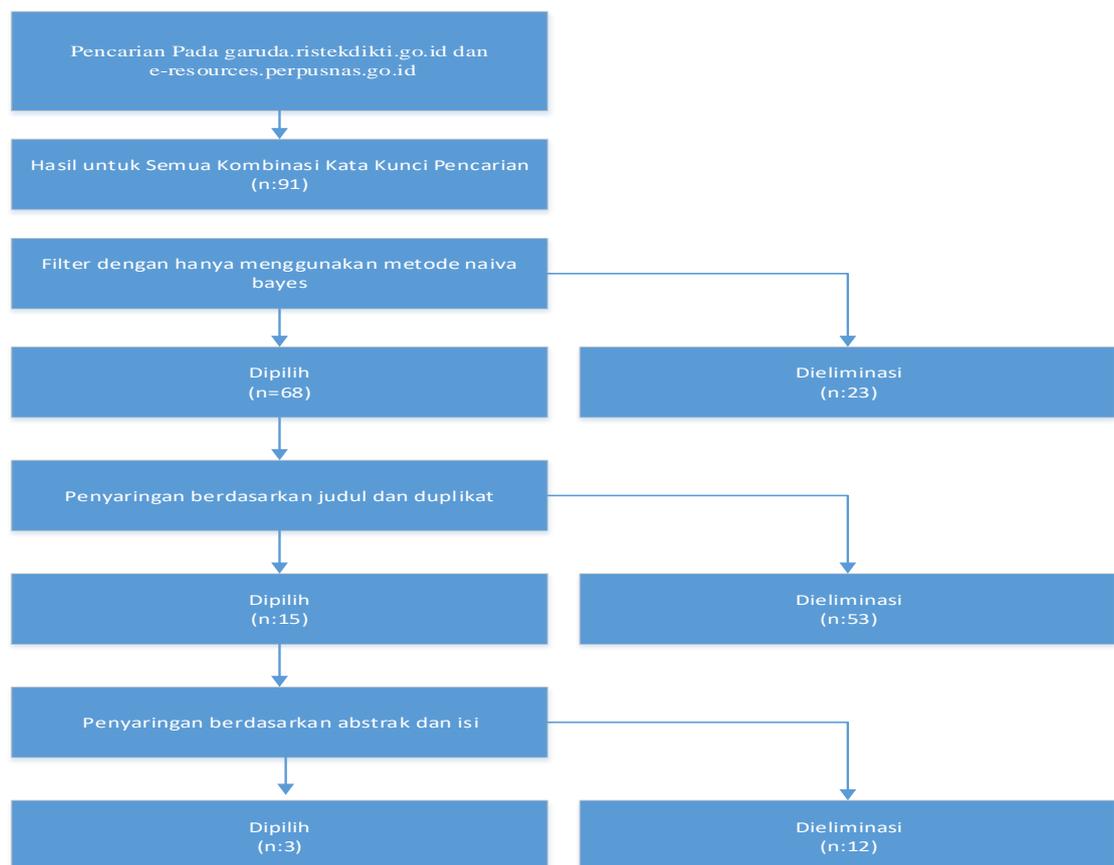
Referensi dari proses pencarian akan mengalami beberapa tahapan penyaringan. Pada penyaringan pertama penulis mengeliminasi jurnal yang menggunakan metode selain metode naïve bayes dan hanya menggunakan literatur yang menggunakan metode naïve bayes. Kemudian dilakukan screening judul untuk mengeliminasi beberapa literatur yang sama. Pada penyaringan selanjutnya melakukan kesesuaian antara judul dengan topik yang diulas. Selanjutnya membaca abstrak untuk mengetahui apakah literatur tersebut sesuai dengan topik secara garis besar. Setelah itu dilakukan proses pembacaan isi literatur secara skimming. Kemudian memilih beberapa literatur yang benar-benar sesuai dengan topik untuk direview. Setelah mendapat literatur yang sesuai maka literatur tersebut di baca dengan lebih detail untuk meninjau isi dari literatur-literatur yang ada. Hasil tinjauan kemudian dibuat kedalam sebuah ringkasan untuk memudahkan dalam proses banding. Ringkasan tersebut berisi tentang: 1) Banyaknya dataset yang digunakan, 2) Metode yang digunakan, 3) Aplikasi yang digunakan, 4) Hasil prediksi yang diperoleh dari masing-masing literatur. Selanjutnya hasil dari ringkasan tersebut dituangkan kedalam sebuah table yang akan menjadi acuan pembading dari literatur yang digunakan. Sehingga menghasilkan data yang specific dari masing-masing literatur dalam efisiensi penggunaan metode naïve bayes terhadap prediksi kelulusan mahasiswa tepat waktu. Kemungkinan terjadinya kesalahan penafsiran penulis sadari dikarenakan perbedaan penggunaan dataset dari masing-masing literatur. Penggunaan perangkat hardware yang berbeda juga berpengaruh terhadap kecepatan proses penghitungan algoritma. Serta beberapa factor lain yang dapat menjadi pemicu terjadinya kesalahan penafsiran. Maka dari itu kunci utama yang

difokuskan dalam systematic review ini adalah hasil yang di dapat dari metode naïve bayes yang diperoleh dari setiap literatur. Penulis menganalisa perbedaan dan persamaan yang ada pada masing-masing literatur.

## HASIL DAN PEMBAHASAN

Proses pencarian referensi menghasilkan total sitasi sebanyak 91 literatur. Setelah proses eliminasi pertama dengan menghilangkan literatur yang menggunakan metode selain naïve bayes maka tereliminasi 23 literatur dan menyisakan 68 literatur saja. Kemudian dilakukan proses penyaringan kedua dengan screening judul untuk mengeliminasi literatur duplikat.

Proses screening judul menyisakan 21 yang kemudian dilakukan pemilihan literasi yang memiliki judul yang sesuai dengan topik yang diulas. Dari proses kesesuaian judul menghasilkan 15 literatur yang oleh penulis dilakukan pembacaan abstrak untuk mendapatkan garis besar literatur yang sesuai dengan kriteria topik systematic review ini. Kemudian literatur yang lolos dari tahap pembacaan abstrak penulis baca secara skimming isi dari literatur tersebut sehingga menghasilkan 3 literatur yang sesuai dan memenuhi syarat untuk dijadikan acuan dalam penulisan systematic review ini.



Gambar 1. Flow proses pencarian literatur

Literatur yang digunakan adalah literatur yang memiliki kriteria yang sesuai dengan topik tentang prediksi kelulusan mahasiswa tepat waktu dengan menggunakan metode naïve bayes. Literatur yang kami gunakan adalah literatur yang meneliti database akademik perguruan tinggi baik berupa NIM, nama mahasiswa, IPK, umur, jenis

kelamin, status mahasiswa dan objek lain yang dapat menjadi bahan untuk dapat membuat suatu prediksi kelulusan mahasiswa tepat waktu.

Untuk mencegah terjadinya bias, penulis menetapkan 4 variabel kontrol, yaitu jumlah data set yang digunakan, jumlah atribut yang digunakan, ada tidaknya metode validasi dan penggunaan aplikasi data mining pada proses penelitiannya.

Tabel 1. Validasi dengan empat variabel kontrol yang dijadikan hasil akurasi

<i>Literatur</i>	<i>Data set</i>	<i>Atribut lebih dari 5</i>	<i>Menggunakan</i>	<i>Menggunakan</i>
	di atas 200		Aplikasi data mining	Metode Validasi
(Suardika, 2019)	ya	ya	ya	ya
(Effendi, 2014)	ya	ya	ya	tidak mencantumkan
(Nugroho & Wibowo, 2017)	ya	ya	ya	tidak mencantumkan

Untuk mempermudah dalam proses analisa maka hasil dari tiga literatur kemudian dituangkan dalam bentuk tabel 2.

Tabel 2. Hasil Literatur

<i>Literatur</i>	<i>Data set</i>	<i>Atribut</i>	<i>Aplikasi data mining</i>	<i>Akurasi</i>	<i>Keterangan</i>
(Suardika, 2019)	226	6	Weka	Relative absolute error 0.0359% Root relative squared error 0,0654%	Screenshot hasil perhitungan
(Effendi, 2014)	1043	6	Rapid Miner 7.0	99,33%	Screenshot hasil perhitungan
(Nugroho & Wibowo, 2017)	210	13	Rapid Miner	90,95%	Screenshot hasil perhitungan

Literatur pertama menggunakan data set sebanyak 226 data dengan 6 atribut dan menggunakan aplikasi Weka menghasilkan prediksi presentasi lulus tepat waktu adalah 98 % dan tidak lulus tepat waktu adalah 2 % dengan Relative absolute error 0.0359% dan Root

relative squared error 0,0654%.Literatur Kedua menggunakan data set sebanyak 1.043 dengan 6 atribut dan menggunakan aplikasi Rapid Miner 7.0 menghasilkan tingkat akurasi sebesar 99,30%.Sedangkan literatur ketiga menggunakan dataset sebanyak 210 dengan atribut 13 dan menggunakan aplikasi Rapid Miner menghasilkan tingkat akurasi sebesar 90,95%. Dari ketiga literatur dihasilkan prediksi kelulusan lebih dari 90% , terutama pada literatur kedua yang menggunakan dataset paling banyak.Literatur pertama menggunakan aplikasi data mining yang berbeda dari dua litaratur lainnya.Hal ini tidak mempengaruhi hasil karena algoritma yang digunakan sama yaitu algoritma naïve bayes.

Untuk literatur pertama dan ketiga menggunakan dataset yang hampir sama tetapi literatur ketiga menggunakan atribut lebih banyak.Faktor banyaknya atribut yang digunakan berpengaruh terhadap prediksi hasil dari literatur.Semakin banyak atribut yang digunakan semakin banyak pula kemungkinan yang didapat sehingga akan mempengaruhi keakurasian hasil akhir.

Dari tabel 2 dapat disimpulkan bahwa kelulusan mahasiswa tepat waktu dapat diprediksi menggunakan metode data mining naïve bayes dengan memperhatikan atribut yang digunakan dalam proses data mining tersebut hal ini menjawab pertanyaan penelitian pertama.Selain itu dari tabel 2 juga dapat dilihat hasil akurasi yang diperoleh dengan menggunakan apliasi data mining yaitu weka dan Rapid Miner dengan metode naïve bayes mencapai akurasi lebih dari 90%.Data akurasi tersebut dapat menjadi acuan dalam menjawab pertanyaan penelitian kedua.

Tabel 3. Perbandingan Atribut

<i>Literatur</i>	<i>Data set</i>	<i>Atribut</i>	<i>Aplikasi data mining</i>	<i>Akurasi</i>
(Suardika, 2019)	226	Nama_prodi, nama_fakultas, angkatan, date_wisuda, ipk, masa studi	Weka	Relative absolute error 0.0359% Root relative squared error 0,0654%
(Effendi, 2014)	1043	Nama_mahasiswa, umur, jenis kelamin, status mahasiswa, IPS&IPK, status_lulus	Rapid Miner 7.0	99,33%
(Nugroho & Wibowo, 2017)	210	Nim,nama, jurusan, jenis_kelamin, umur, daerah_asal, status pernikahan, status_pekerjaan, kelompok, IPS1,IPS2,IPS3,IPS4,IPS5,IPS6,IPS7,IPS8, Konsentarsi	Rapid Miner	90,95%

Dari tabel 3. diatas dapat dibandingkan bahwa atribut yang selalui digunakan untuk prediksi kelulusan pada setiap literatur adalah nama mahasiswa dan IPK atau IPS.

## **PENUTUP**

### **Simpulan**

Tujuan dari dibuatnya systematic review ini adalah untuk mengetahui apakah kelulusan mahasiswa tepat waktu dapat diprediksi dengan metode data mining naïve bayes.

---

Setelah dianalisa dan hasilnya dituangkan ke dalam bentuk table pembandingan, maka didapatkan jawaban untuk pertanyaan penelitian yaitu :

1. Apakah Metode data mining naive bayes dapat memprediksi kelulusan mahasiswa tepat waktu?
2. Seberapa besar akurasi yang didapat dari hasil metode data mining dalam memprediksi kelulusan mahasiswa tepat waktu?
3. Apakah atribut yang banyak digunakan dalam memprediksi kelulusan mahasiswa tepat waktu ?

Berdasarkan hasil identifikasi literatur yang digunakan, didapat bahwa metode data mining naïve bayes dapat membuat suatu prediksi mengenai kelulusan mahasiswa tepat waktu dengan memperhitungkan atribut – atribut dari database perguruan tinggi yang digunakan. Sedangkan untuk tingkat akurasi ketiga literatur menghasilkan akurasi di atas 90% walaupun dengan menggunakan jumlah atribut dan aplikasi data mining yang berbeda. Atribut yang terdapat pada semua literatur yang dapat menentukan prediksi adalah atribut IPK (Indeks Prestasi Kumulatif)

### **Saran**

Dengan adanya prediksi sedari awal mengenai kelulusan mahasiswa tepat waktu diharapkan dapat menjadi acuan bagi perguruan tinggi untuk meningkatkan faktor yang mendukung peningkatan jumlah kelulusan mahasiswa tepat waktu. Selain itu prediksi yang di hasilkan dapat pula menjadi catatan untuk dapat menghindari dan mengurangi penyebab dari mahasiswa yang telat lulus agar perguruan tinggi tersebut menjadi perguruan tinggi yang memiliki kredibilitas dalam penyelenggaraan pendidikan. Peneliti selanjutnya dapat menggunakan atribut – atribut yang menjadi acuan dalam prediksi kelulusan tersebut.

### **DAFTAR PUSTAKA**

- Ayub, M. (2018). *Proses Data Mining dalam Sistem Pembelajaran Berbantuan Komputer*. (January 2012), 21–30.
- Broto Legowo, M., & Indiarso, B. (2017). Model Sistem Penjaminan Mutu Berbasis Integrasi Standar Akreditasi BAN-PT dan ISO 9001:2008. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 1(2), 90. <https://doi.org/10.29207/resti.v1i2.51>
- Chien, C. F., Wang, W. C., & Cheng, J. C. (2007). Data mining for yield enhancement in semiconductor manufacturing and an empirical study. *Expert Systems with Applications*, 33(1), 192–198. <https://doi.org/10.1016/j.eswa.2006.04.014>
- Effendi, M. R. (2014). Akurasi Data Mining Untuk Menghasilkan Pola Kelulusan Mahasiswa dengan Metode NAÏVE BAYES. *Jurnal Sistem Informasi Universitas Suryadarma*, 3(2), 101–106. <https://doi.org/10.35968/jsi.v3i2.66>
- Kantardzic, M. (2011). Data Mining: Concepts, Models, Methods, and Algorithms: Second Edition. In *Data Mining: Concepts, Models, Methods, and Algorithms: Second Edition*. <https://doi.org/10.1002/9781118029145>
- Nugroho, M. F., & Wibowo, S. (2017). Fitur Seleksi Forward Selection Untuk Menentukan Atribut Yang Berpengaruh Pada Klasifikasi Kelulusan Mahasiswa Fakultas Ilmu Komputer UNAKI Semarang Menggunakan Algoritma Naive Bayes. *Jurnal Informatika Upgris*, 3(1), 63–70. <https://doi.org/10.26877/jiu.v3i1.1669>
- Saleh, A., & Nasari, F. (2018). Penerapan Equal-Width Interval Discretization Dalam Metode Naive Bayes Untuk Meningkatkan Akurasi Prediksi Pemilihan

- Jurusan Siswa. *Masyarakat Telematika Dan Informasi : Jurnal Penelitian Teknologi Informasi Dan Komunikasi*, 9(1), 1. <https://doi.org/10.17933/mti.v9i1.113>
- Septiani, W. D. (2017). Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis. *None*, 13(1), 76–84.
- Suardika, I. G. I. (2019). Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naive Bayes: Studi Kasus Fakultas Ekonomi Dan Bisnis Universitas Pendidikan Nasional. *Jurnal Ilmu Komputer Indonesia*, 4(2), 37–44. <https://doi.org/10.23887/jik.v4i2.2775>
- Wulandari, D. ., Jatnika, H., & Purwanto, Y. . (2015). Rancang Bangun Aplikasi Data Mining Untuk. *Jurnal Kajian Ilmu Dan Teknologi*, 1(1), 31–38.
- Zatmiko, R. . (2015). ANALISIS PENERAPAN METODE. In *Universitas Nusantara PGRI Kediri* (Vol. 01).