

## Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes, Support Vector Machine (SVM), dan Random Forest untuk Prediksi Ketidakhadiran di Tempat Kerja

Hiya Nalatissifa<sup>1</sup>, Windu Gata<sup>2</sup>, Sri Diantika<sup>3</sup>, Khoirun Nisa<sup>4</sup>

<sup>1,2,3,4</sup>Ilmu Komputer, STMIK Nusa Mandiri, Jl. Raya Margonda No.545, Pondok Cina, Beji, Kota Depok, Jawa Barat, Indonesia, 16424  
e-mail: <sup>1</sup>14002363@nusamandiri.ac.id, <sup>2</sup>windu@nusamandiri.ac.id, <sup>3</sup>14002378@nusamandiri.ac.id, <sup>4</sup>khoirun.khn@nusamandiri.ac.id

Submitted Date: November 03<sup>rd</sup>, 2020  
Revised Date: January 02<sup>nd</sup>, 2021

Reviewed Date: December 30<sup>th</sup>, 2020  
Accepted Date: January 05<sup>th</sup>, 2021

### Abstract

Absence is a problem for the company. Absenteeism is defined as a task that is assigned to an individual, but the individual cannot complete the task when he is not present. Absence from work is influenced by many factors, including mismatched working hours, job demand and other factors such as serious accidents / illness, low morale, poor working conditions, boredom, lack of supervision, personal problems, insufficient nutrition, transportation problems, stress, workload, and dissatisfaction. The purpose of this study is to predict absenteeism at work based on the Absenteeism at work dataset obtained from the UCI Machine Learning repository site using the Weka 3.8 application and the Naïve Bayes algorithm, Support Vector Machine (SVM), and Random Forest. In the results of the study, the Random Forest algorithm obtained the highest accuracy, precision, and recall values compared to the Naïve Bayes and SVM algorithms, which resulted in an accuracy value of 99.38%, 99.42% precision and a recall of 99.39%.

Keywords: Data Mining; Clasification; Naïve Bayes; Super Vector Machine; Random Forest

### Abstrak

Ketidakhadiran merupakan permasalahan bagi perusahaan. Ketidakhadiran diartikan sebagai tugas yang dibebankan kepada individu, namun individu tersebut tidak dapat menyelesaikan tugasnya ketika tidak hadir. ketidakhadiran di tempat kerja dipengaruhi oleh banyak faktor, di antaranya ketidaksesuaian jam kerja, job demand dan faktor lain seperti kecelakaan/sakit yang serius, rendahnya moral, buruknya kondisi kerja, kebosanan, kurangnya supervisi, masalah personal, nutrisi yang tidak cukup, masalah transportasi, stres, beban kerja, dan ketidakpuasan. Tujuan dari penelitian ini adalah untuk memprediksi ketidakhadiran di tempat kerja berdasarkan dataset Absenteeism at work yang di peroleh dari situs repository UCI Machine Learning dengan menggunakan aplikasi Weka 3.8 dan algoritma Naïve Bayes, Support Vector Machine (SVM), dan Random Forest. Pada hasil penelitian, algoritma Random Forest memperoleh nilai akurasi, presisi, dan recall yang paling tinggi dibandingkan dengan algoritma Naïve Bayes dan SVM, yaitu menghasilkan nilai akurasi sebesar 99.38%, presisi 99.42% dan recall 99.39%.

Kata Kunci: Data Mining; Klasifikasi; Naïve Bayes; Super Vector Machine; Random Forest

### 1. Pendahuluan

Mengendalikan sumber daya manusia merupakan suatu kewajiban bagi semua organisasi atau perusahaan. Sebagai upaya mencapai sebuah tujuan organisasi, sumber daya manusia di tuntut untuk melakukan pekerjaan secara efisien dan efektif. Tercapainya tujuan organisasi apabila

karyawan dapat menunjukkan kinerja terbaiknya, salah satunya yaitu diukur melalui tingkat kehadiran (Setiani & Martono, 2018).

Ketidakhadiran merupakan permasalahan bagi perusahaan. Ketidakhadiran diartikan sebagai tugas yang dibebankan kepada individu, namun individu tersebut tidak dapat menyelesaikan

tugasnya ketika tidak hadir (Fitri, 2017). Beberapa peneliti menemukan bahwa ketidakhadiran karyawan dipengaruhi oleh banyak faktor, diantaranya: ketidaksesuaian jam kerja, job demand (Vignoli, Guglielmi, Bonfiglioli, & Saverio, 2015) dan faktor lain seperti kecelakaan/sakit yang serius, rendahnya moral, buruknya kondisi kerja, kebosanan, kurangnya supervisi, masalah personal, nutrisi yang tidak cukup, masalah transportasi, stres, beban kerja, dan ketidakpuasan (Setiani & Martono, 2018). Biaya ketidakhadiran karena sakit di banyak organisasi lebih besar daripada gaji yang dibayarkan kepada karyawan yang sakit yang tidak hadir di tempat kerja dan tanpa ada pengganti, sehingga implikasinya terhadap produktivitas kerja jauh lebih besar (Martiniano, Ferreira, & Produção, 2012).

Untuk memprediksi ketidakhadiran di tempat kerja berdasarkan faktor-faktor tersebut, maka diperlukan data mining yang merupakan pendekatan kuat dan efektif dengan menyediakan proses penemuan pola pada dataset besar (Prajarini, 2016). Metode yang melakukan pengelompokan data dengan sistem partisi dan proses pemodelan tanpa supervisi dilakukan oleh data mining (Hidayati, Rizmayanti, Bunga, Dewi, & Fatmasari, 2020). Penelitian ini dilakukan dengan menggunakan metode klasifikasi.

Metode klasifikasi/taksonomi merupakan penempatan objek tertentu dalam suatu set kategori berdasarkan objek/konsep property (Septiani, 2017). Klasifikasi membangun model dengan data latih yang ada, selanjutnya menggunakan model tersebut untuk mengklasifikasikan pada data yang baru. Klasifikasi diartikan sebagai pekerjaan dengan melakukan pembelajaran fungsi target yang melakukan pemetaan setiap set atribut/fitur ke satu jumlah label kelas (Utomo, 2020). Proses klasifikasi berdasarkan empat komponen mendasar yaitu prediktor, kelas, pengujian dataset, dan training set.

Penelitian terkait mengenai data mining dengan menggunakan metode klasifikasi diantaranya, (Prajarini, 2016) meneliti penyakit kulit dengan menggunakan dataset yang diunduh dari situs repository UCI menggunakan perbandingan algoritma klasifikasi Decision Tree, Naive Bayes, KNN dan SVM. (Pratama, Wihandika, & Ratnawati, 2018) memprediksi ketepatan waktu kelulusan mahasiswa dengan melakukan klasifikasi data berdasarkan parameter yang telah ditentukan dengan menggunakan

algoritma SVM (Rahmadi et al., 2020) melakukan pengujian akurasi terhadap dataset pasien pasca operasi dengan weka tools menggunakan algoritma Naive Bayes. (Septiani, 2017) meneliti dataset penyakit hepatitis dengan melakukan komparasi terhadap algoritma C4.5 dan Naive Bayes. (Mardhiyah et al., 2020) meneliti dataset tumpukan data nasabah pemegang kartu kredit yang belum diklasifikasikan dan bertujuan memprediksi macet atau tidaknya pembayaran kredit oleh nasabah. (Setiyorini & Asmono, 2018) menggunakan perbandingan algoritma Decision Tree, Naive Bayes, dan KNN untuk menganalisa faktor yang mempengaruhi kinerja siswa. (Nugroho & Emiliyawati, 2017) melakukan penelitian menggunakan algoritma Random Forest dengan melakukan klasifikasi faktor yang mempengaruhi tingkat penerimaan konsumen pada mobil. (Cendana, Dian, & Permana, 2019) melakukan prediksi bisnis pada data yang ada dengan cara menganalisa dan menguji data untuk proyeksi pelanggan mana yang dapat melakukan transaksi kembali pada bisnis UMKM menggunakan algoritma Naive Bayes, Random Forest Tree, dan J48. (Derisman, 2020) membandingkan hasil algoritma terbaik dari penelitian sebelumnya yaitu Algoritma Naive Bayes, algoritma random forest, Neural Network, untuk mendapatkan algoritma paling baik dalam memprediksi penyakit jantung. (Yusuf, Qalbi, Dwitawati, & Ellyadi, 2020) melakukan penelitian dengan mengimplementasi algoritma Naive Bayes dan Random Forest untuk prediksi prestasi akademik mahasiswa.

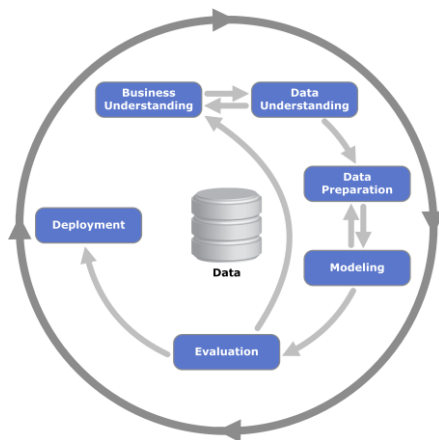
Berdasarkan literature review di atas, dalam penelitian ini data yang akan diolah adalah dataset ketidakhadiran di tempat kerja (Absenteeism at work dataset) yang diperoleh dari situs repository UCI Machine Learning. Algoritma yang digunakan untuk klasifikasi prediksi ketidakhadiran di tempat kerja pada penelitian ini adalah algoritma Naive Bayes, SVM, dan Random Forest dengan menggunakan aplikasi Weka versi 3.8.

Tujuan penelitian ini yaitu untuk prediksi ketidakhadiran di tempat kerja dengan melakukan pengklasifikasian algoritma klasifikasi yang dapat memperoleh hasil yang akurat dan presisi.

## 2. Metode Penelitian

Pada penelitian ini, metode penelitian yang digunakan adalah model Cross Standard Industry for Data Mining (CRISP-DM). CRISP-DM memiliki enam fase siklus hidup proyek data

mining (Feblian et al., 2017). Adapun gambaran fase pada model CRISP-DM adalah sebagai berikut:



Gambar 1. Diagram CRISP-DM (Mardhiyah et al., 2020)

Penjelasan mengenai fase pada model CRISP-DM adalah:

- Tahap Pemahaman bisnis (Business Understanding)**  
 Tahap ini merupakan tahap pemahaman terhadap substansi kegiatan data mining yang ingin dilakukan, kebutuhan perspektif penelitian. Kegiatannya yaitu menentukan sasaran/tujuan penelitian, memahami situasi penelitian, menerjemahkan tujuan penelitian dalam tujuan data mining.
- Tahap Pemahaman Data (Data Understanding)**  
 mengumpulkan data, mempelajari data agar data dapat dipahami kemudian digunakan pada penelitian, melakukan identifikasi masalah terkait data.
- Tahap persiapan Data (Data Preparation)**  
 Mempersiapkan struktur basis datayang dapat memudahkan proses mining.
- Tahap Pemodelan (Modeling Phase)**  
 merupakan fase penentuan penggunaan teknik data mining, algoritma data mining, tools data mining, serta penentuan parameter dengan nilai optimal.
- Tahap Evaluasi (Evaluation Phase)**  
 Merupakan interpretasi hasil data mining pada proses pemodelan. Evaluasi bertujuan untuk menyesuaikan model yang diperoleh agar sesuai dengan sasaran yang akan dicapai.
- Tahap Penerapan (Deployment Phase)**  
 Merupakan tahap menyusun laporan pengetahuan yang diperoleh dari evaluasi proses data mining.

## 2.1. Naive Bayes

Naive Bayes adalah klasifikasi probabilitas berdasarkan Teorema Bayes. Naive Bayes beranggapan efek pada nilai atribut terhadap kelas tertentu independen dari nilai atribut lain, hal ini bertujuan untuk menyederhanakan perhitungan (Prajarini, 2016). Rumus probabilitas bayes:

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$$

Keterangan:

$P(C_i|X)$  : Probabilitas  $C_i$  dengan bukti  $X$

$P(C_i)$  : Probabilitas  $C_i$

$P(X|C_i)$  : Probabilitas  $X$  berpengaruh terhadap  $C_i$

$P(X)$  : Probabilitas  $X$

## 2.2. Support Vector Machine (SVM)

SVM adalah algoritma klasifikasi yang berfungsi untuk melakukan klasifikasi data non linier dan data liner (Prajarini, 2016). Prinsip SVM yaitu menemukan hyperlane yang bisa melakukan klasifikasi data menjadi 2 kelas. Rumus perhitungannya adalah sebagai berikut:

Titik data :  $x_i = \{x_1, x_2, \dots, x_n\} \in R^n$

Kelas data :  $y_i \in \{-1, +1\}$

Pasangan data dan kelas :

$$\{(x_i, y_i)\}_{i=1}^N$$

Maksimalkan fungsi:

$$Ld = \sum_{i=1}^N \alpha_i - \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) \text{ syarat : } 0 \leq \alpha_i \leq C \text{ dan } \sum_{i=1}^N \alpha_i y_i = 0$$

Menghitung nilai  $w$  dan  $b$  :

$$w = \sum_{i=1}^N \alpha_i y_i x_i \quad b = -\frac{1}{2}(w \cdot x^+ + w \cdot x^-)$$

Fungsi keputusan klasifikasi  $\text{sign}(f(x))$  :

$$f(x) = w \cdot x + b \quad \text{atau} \quad f(x) = \sum_{i=1}^m \alpha_i y_i K(x, x_i) + b$$

Keterangan :

$N$  : Jumlah data

$n$  : banyak fitur/dimensi data

$C$  : nilai konstanta

$m$  : jumlah support vector/titik data yang memiliki  $\alpha_i > 0$

$K(x, x_i)$  : fungsi kernel

Ld : Dualitas Lagrange Multiplier  
 $\alpha_i$  : nilai bobot setiap titik data

### 2.3. Random Forest

Metode Random Forest (RF) adalah metode untuk meningkatkan nilai akurasi. Metode ini bertujuan untuk membangun pohon keputusan yang terdiri dari root node, internal node, dan leaf node dengan mengambil atribut dan data secara acak sesuai ketentuan. Root node adalah simpul yang terletak di atas (akar dari pohon keputusan). Sedangkan internal node merupakan simpul percabangan, node ini mempunyai satu input dan minimal dua output. Leaf node/terminal node yaitu simpul terakhir yang tidak memiliki output dan hanya memiliki satu input. Perhitungan nilai entropy menggunakan rumus pada persamaan 1, dan nilai information gain pada rumus persamaan 2 (Nugroho & Emiliyawati, 2017).

$$Entropy(Y) = -\sum_i p(c|Y) \log_2 p(c|Y) \quad (1)$$

Y merupakan himpunan kasus dan  $p(c|Y)$  adalah proporsi nilai Y pada kelas c.

$$Information\ Gain(Y, a) = Entropy(Y) - \sum_{v \in Values(a)} \frac{|Y_v|}{|Y_a|} Entropy(Y_v) \quad (2)$$

Values(a) adalah semua nilai yang mungkin pada himpunan kasus a.  $Y_v$  merupakan subkelas dari Y dengan kelas v yang berkaitan dengan kelas a.  $Y_a$  merupakan semua nilai yang sama dengan a.

## 3. Hasil

### 3.1. Tahap Pemahaman Bisnis (Business Understanding)

Dengan memanfaatkan sumber data yang ada, maka dapat dianalisis dan diprediksi menggunakan teknik data mining yang bertujuan untuk pengklasifikasian dataset ketidakhadiran di tempat kerja (Absenteeism at work dataset) menggunakan algoritma naïve bayes, SVM, dan Random Forest.

### 3.2. Tahap Pemahaman Data (Data Understanding)

Merupakan dataset ketidakhadiran di tempat kerja yang di peroleh dari UCI Machine Learning Repository dengan alamat URL (<https://archive.ics.uci.edu/ml/datasets/Absenteeism+at+work>). Tahap pemahaman data yaitu dengan mendeskripsikan atribut-atribut yang

terdapat pada dataset dan mendeskripsikan tipe nilai atribut pada dataset yang digunakan.

### 3.3. Tahap persiapan Data (Data Preparation)

Data yang didapatkan pada penelitian ini sejumlah 740 data dengan 21 atribut yang akan diproses untuk menghasilkan prediksi ketidakhadiran di tempat kerja. Beberapa atribut yang terdapat pada dataset tersebut adalah:

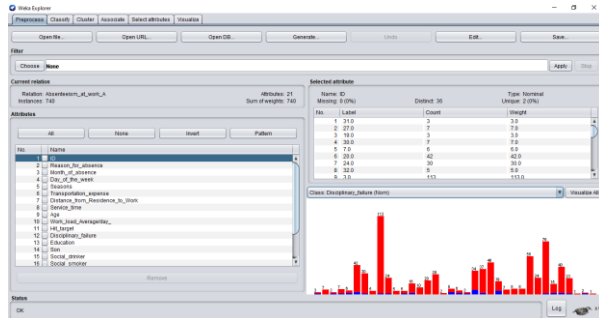
Tabel 1. Atribut dataset Absenteeism at work

Nama Atribut	Tipe Atribut	Keterangan
<i>ID</i>	Kategori	Identifikasi individu
<i>Reason for absence</i>	Kategori	Alasan ketidakhadiran
<i>Month of absence</i>	Real	Bulan absen
<i>Day of the week</i>	Kategori	Hari dalam seminggu
<i>Seasons</i>	Kategori	Musim
<i>Transportation expense</i>	Real	Biaya transportasi
<i>Distance from Residence to Work</i>	Real	Jarak tempat tinggal ke tempat kerja
<i>Service time</i>	Integer	Waktu pelayanan
<i>Age</i>	Integer	Usia
<i>Work load Average/day</i>	Real	Beban kerja rata-rata/hari
<i>Hit target</i>	Real	Target
<i>Disciplinary failure</i>	Kategori	Kegagalan disiplin
<i>Education</i>	Real	Pendidikan
<i>Son</i>	Real	Jumlah anak
<i>Social drinker</i>	Kategori	Peminum
<i>Social smoker</i>	Kategori	Perokok
<i>Pet</i>	Real	Jumlah hewan
<i>Weight</i>	Real	Berat
<i>Height</i>	Real	Tinggi
<i>Body mass index</i>	Real	Indeks masa tubuh
<i>Absenteeism time in hours (target)</i>	Real	Waktu ketidakhadiran dalam jam

Class yang digunakan pada penelitian ini adalah field *Disciplinary failure* (kegagalan disiplin) yang digunakan untuk membagi data menjadi dua set (Label) dengan nilai 1 dan 0. Dimana, 1 adalah tidak disiplin dan 0 berarti disiplin.

### 3.4. Tahap Pemodelan (Modeling Phase)

Pada tahap ini, langkah yang dilakukan yaitu mengaplikasikan dataset yang digunakan, yaitu dataset ketidakhadiran di tempat kerja (Absenteeism at work dataset) menggunakan tools Weka 3.8 dengan format \*.aff seperti gambar berikut:



Gambar 2. Pengaplikasian dataset

Setelah mengaplikasikan dataset, tahap pemodelan pada penelitian ini adalah melakukan proses klasifikasi menggunakan algoritma Naïve Bayes, SVM, dan Random Forest. Untuk mengevaluasi permorma dari masing-masing algoritma yang digunakan, penelitian ini melakukan pembagian persentasi data testing dan data training. Berikut penjelasan dari masing-masing algoritma tersebut:

#### a. Naïve Bayes

Berikut perolehan hasil akurasi, presisi, dan recall yang diperoleh algoritma Naïve Bayes berdasarkan pengujian menggunakan tools Weka dengan membagi persentasi data testing dan data training:

Tabel 2. Hasil Pengujian Algoritma Naïve Bayes

Data Training (%)	Data Testing (%)	Hasil Akurasi (%)	Presisi (%)	Recall (%)
50	50	96.5	96.9	96,5
60	40	96.9	97.5	97
70	30	96.8	98	96.8
80	20	97.3	98	97.3
90	10	95.9	97.5	95.9

Berdasarkan tabel di atas, hasil perhitungan nilai rata-rata algoritma Naive Bayes diperoleh akurasi 96,68%, presisi 97,58% dan recall sebesar 96,7%.

#### b. Support Vector Machine (SVM)

Berikut perolehan hasil akurasi, presisi, dan recall yang diperoleh algoritma Support Vector Machine (SVM) berdasarkan pengujian menggunakan tools Weka dengan membagi persentasi data testing dan data training:

Tabel 3. Hasil Pengujian Algoritma Support Vector Machine (SVM)

Data Training (%)	Data Testing (%)	Hasil Akurasi (%)	Presisi (%)	Recall (%)
50	50	99.2	99.3	99.2
60	40	99.3	99.3	99.3
70	30	99.5	99.6	99.5
80	20	99.3	99.4	99.3
90	10	98.6	98.7	98.6

Berdasarkan tabel di atas, hasil perhitungan nilai rata-rata algoritma Support Vector Machine (SVM) diperoleh akurasi 99.18% , presisi 99.26% dan recall sebesar 99.18%.

#### c. Random Forest

Berikut perolehan hasil akurasi, presisi, dan recall yang diperoleh algoritma Random Forest berdasarkan pengujian menggunakan tools Weka dengan membagi persentasi data testing dan data training:

Tabel 4. Hasil Pengujian Algoritma Random Forest

Data Training (%)	Data Testing (%)	Hasil Akurasi (%)	Presisi (%)	Recall (%)
50	50	99.2	99.3	99.2
60	40	99.3	99.3	99.3
70	30	99.5	99.6	99.5
80	20	99.3	99.4	99.3
90	10	98.6	98.7	98.6

Berdasarkan tabel di atas, hasil perhitungan nilai rata-rata algoritma Random Forest diperoleh akurasi 99.38%, presisi 99.42% dan recall sebesar 99.39%.

### 3.5. Evaluation (Tahap Evaluasi)

Penelitian pada tahap evaluasi, melakukan klasifikasi menggunakan algoritma Naïve Bayes, SVM, dan Random Forest untuk pengukuran keakuratan hasil yang dicapai. Berdasarkan hasil pengujian dari masing-masing algoritma yang di gunakan dalam penelitian, berikut evaluasi perbandingan hasil pengujian algoritma Naïve

Bayes, Support Vector Machine (SVM), dan Random Forest:

Tabel 5. Perbandingan Hasil Pengujian

Data Training (%)	Data Testing (%)	Hasil Akurasi (%)			Presisi (%)			Recall (%)		
		Naïve Bayes	SVM	Random Forest	Naïve Bayes	SVM	Random Forest	Naïve Bayes	SVM	Random Forest
50	50	96.5	99.2	99.2	96.9	99.3	99.3	96,5	99.2	99.2
60	40	96.9	99.3	99.3	97.5	99.3	99.3	97	99.3	99.3
70	30	96.8	99.5	99.5	98	99.6	99.6	96.8	99.5	99.5
80	20	97.3	99.3	99.3	98	99.4	99.4	97.3	99.3	99.3
90	10	95.9	98.6	98.6	97.5	98.7	98.7	95.9	98.6	98.6
<b>Nilai Rata-rata</b>		96.68	99.18	99.38	97.58	99.26	99.42	96,7	99.18	99.39

#### 4. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan pada dataset ketidakhadiran di tempat kerja (Absenteeism at work dataset) yang di peroleh dari situs repository UCI Machine Learning menggunakan algoritma klasifikasi Naïve Bayes, SVM, dan Random Forest yang digunakan untuk prediksi ketidakhadiran di tempat kerja dapat di ambil beberapa kesimpulan, yaitu:

- Algoritma Naïve Bayes, SVM, dan Random Forest dapat diterapkan untuk prediksi ketidakhadiran di tempat kerja dengan nilai akurasi, presisi, dan recall di atas 96%.
- Hasil penelitian menggunakan beberapa algoritma klasifikasi yang telah digunakan dalam prediksi ketidakhadiran di tempat kerja, menunjukkan algoritma Random Forest memperoleh nilai akurasi, presisi, dan recall yang paling tinggi dibandingkan dengan algoritma Naïve Bayes dan SVM, yaitu dengan hasil akurasi sebesar 99.38%, presisi 99.42% dan recall 99.39%.

#### Referensi

- Cendana, M., Dian, S., & Permana, H. (2019). Analisis Perbandingan Algoritma Naive Bayes , J48 , Dan Random Forest Tree Dalam Peningkatan Loyalitas Pelanggan Umkm Dengan Voucher Belanja, *11*(2), 140–145.
- Derisman. (2020). Perbandingan Kinerja Algoritma Untuk Prediksi Penyakit Jantung Dengan Teknik Data Mining, *4*(1), 84–88.
- Feblian, D., Daihani, D. U., Magister, M., Industri, T., Industri, F. T., Trisakti, U., Trisakti, U. (2017). Implementasi Model Crisp-Dm Untuk Menentukan Sales, 1–12.
- Fitri. (2017). Pengaruh Intensi Turnover Dan Ketidakhadiran Terhadap Kinerja Pegawai Dinas Pendidikan Dki Jakarta Mutiara Annisa Fitri. *Manajemen Pendidikan*, (52), 125–138.
- Hidayati, N., Rizmayanti, A. I., Bunga, C., Dewi, S., & Fatmasari, R. (2020). Penerapan Algoritma Klasterisasi Dan Klasifikasi Pada Tingkat Kepentingan Sistem Pembelajaran Di Universitas Terbuka, *8*(2), 134–142.
- Mardhiyah Et Al. (2020). Jurnal Teknologia Klasifikasi Untuk Memprediksi Pembayaran Kartu Kredit Macet Jurnal Teknologia, *3*(1), 91–101.
- Martiniano, A., Ferreira, R. P., & Produção, P. E. De. (2012). Application Of A Neuro Fuzzy Network In Prediction Of Absenteeism At Work.
- Nugroho, S., & Emiliyawati, N. (2017). Sistem Klasifikasi Variabel Tingkat Penerimaan Konsumen Terhadap Mobil Menggunakan Metode Random Forest, *9*(1).
- Prajarini, D. (2016). Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Kulit, *1*(3), 1–5.
- Pratama, A., Wihandika, R. C., & Ratnawati, D. E. (2018). Implementasi Algoritme Support Vector Machine ( Svm ) Untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa, *2*(4), 1704–1708.
- Rahmadi, M., Kaurie, F., & Susanti, T. (2020). Uji Akurasi Dataset Pasien Pasca Operasi Menggunakan Algoritma Naïve Bayes Menggunakan Weka Tools, *7*(1), 134–139. <https://doi.org/10.30865/Jurikom.V7i1.1761>
- Septiani, W. D. (2017). Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis, *13*(1), 76–84.
- Setiani & Martono. (2018). Pengaruh Konflik Pekerjaan-Keluarga, Persepsi Hukuman Organisasi Dan Kepuasan Kerja Pada Ketidakhadiran Pekerja Wanita Bagian Produksi Pt Hyup Sung Indonesia. *Management Analysis Journal*, *7*(1), 90–97.
- Setiyorini, T. &, & Asmono, R. T. (2018). Komparasi Metode Decision Tree , Naive Bayes Dan K-Nearest Neighbor Pada Klasifikasi Kinerja Siswa, *15*(2), 85–92.
- Utomo, D. P. (2020). Analisis Komparasi Metode Klasifikasi Data Mining Dan Reduksi Atribut Pada Data Set Penyakit Jantung, *4*(April), 437–444. <https://doi.org/10.30865/Mib.V4i2.2080>

- Vignoli, M., Guglielmi, D., Bonfiglioli, R., & Saverio, F. (2015). How Job Demands Affect Absenteeism? The Mediating Role Of Work – Family Conflict And Exhaustion. <https://doi.org/10.1007/S00420-015-1048-8>
- Yusuf, B., Qalbi, M., Dwitawati, I., & Ellyadi, M. (2020). Implementasi Algoritma Naive Bayes Dan Random Forest Dalam Memprediksi Prestasi Akademik Mahasiswa Universitas Islam Negeri Ar-Raniry, 4, 50–58.