

**PERBANDINGAN METODE RANDOM FOREST DAN NAIVE BAYES PADA
KLASIFIKASI PERILAKU MAHASISWA DI LMS SPADA INDONESIA**

**COMPARISON OF RANDOM FOREST AND NAIVE BAYES METHODS IN
STUDENT BEHAVIOR CLASSIFICATION AT LMS SPADA INDONESIA**



**MITA ASTUTI
H062201003**



**PROGRAM STUDI MAGISTER STATISTIKA
SEKOLAH PASCASARJANA
UNIVERSITAS HASANUDDIN
MAKASSAR
2024**

Optimized using
trial version
www.balesio.com

**PERBANDINGAN METODE RANDOM FOREST DAN NAIVE BAYES PADA
KLASIFIKASI PERILAKU MAHASISWA DI LMS SPADA INDONESIA**

*COMPARISON OF RANDOM FOREST AND NAIVE BAYES METHODS
IN STUDENT BEHAVIOR CLASSIFICATION AT LMS SPADA
INDONESIA*

**MITA ASTUTI
H062201003**



**PROGRAM STUDI MAGISTER STATISTIKA
SEKOLAH PASCASARJANA
UNIVERSITAS HASANUDDIN
MAKASSAR
2024**

Optimized using
trial version
www.balesio.com

PERBANDINGAN METODE RANDOM FOREST DAN NAIVE BAYES PADA KLASIFIKASI PERILAKU MAHASISWA DI LMS SPADA INDONESIA

Tesis
sebagai salah satu syarat untuk mencapai gelar Magister

Program Studi Statistika

Disusun dan diajukan oleh :

MITA ASTUTI
H062201003

kepada

**PROGRAM STUDI STATISTIKA
SEKOLAH PASCASARJANA
UNIVERSITAS HASANUDDIN
MAKASSAR
2024**



Optimized using
trial version
www.balesio.com

TESIS

PENERAPAN PETA KENDALI *EXPONENTIALLY WEIGHTED MOVING AVERAGE*
PADA DATA DEMERIT

ALFI NUR KHAULY
H0622010002

Telah dipertahankan di hadapan Panitia Ujian Magister pada tanggal 9 Juli 2024 dan dinyatakan telah memenuhi syarat kelulusan
ada

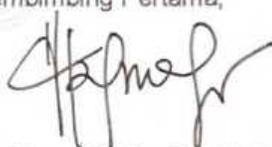
Program Studi Statistika
Departemen Statistika
Sekolah Pascasarjana
Universitas Hasanuddin
Makassar

Mengesahkan

Pembimbing Utama,

Pembimbing Pertama,


Prof. Dr. Dr. Georgina Maria Tinungki, M.Si.
NIP. 19620926 198702 2 001


Dr. Erna Tri Herdiani, S.Si., M.Si.
NIP. 19750429 200003 2 001

Ketua Program Studi
Magister Statistika

Dekan Fakultas MIPA
Universitas Hasanuddin


Dr. Erna Tri Herdiani, S.Si., M.Si.
NIP. 19750429 200003 2 001


Dr. Eng. Amiruddin, M.Si.
NIP. 19720515 199702 1 002



PERNYATAAN KEASLIAN TESIS DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, tesis berjudul "Perbandingan Metode *Random Forest* dan *Naive Bayes* pada Klasifikasi Perilaku Mahasiswa di LMS SPADA Indonesia" adalah benar karya saya dengan arahan dari pembimbing Sri Astuti Thamrin, S.Si., M.Stat., Ph.D sebagai Pembimbing Utama dan Dr. Nirwan, S.Si., M.Si. sebagai Pembimbing Pendamping. Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka tesis ini. Sebagian dari isi tesis ini telah dipublikasikan di Jurnal (*Mathematics and Statistics*) sebagai artikel dengan judul "*Aplication of Random Forest and Naive Bayes Method Classification to Student Behavior*". Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan tesis ini adalah karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut berdasarkan aturan yang berlaku.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya berupa tesis ini kepada Universitas Hasanuddin.

Makassar, 25 Juli 2024



MITA ASTUTI
H062201003



UCAPAN TERIMA KASIH

Segala Puji Syukur kepada **Allah SWT**, Tuhan Yang Maha Esa, atas berkat, bimbingan dan kasih karunia-Nya yang dilimpahkan kepada penulis, serta shalawat dan salam semoga selalu tercurahkan kepada baginda tercinta, Nabi yang paling dimuliakan, pemimpin orang-orang bertakwa, **Muhammad SAW** yang dinanti-nantikan syafaatnya di akhirat kelak. Limpahan doa kepada keluarga serta sahabat Rasulullah SAW. *Alhamdulillah*, berkat rahmat dan karunia serta mukzizat-Nya sehingga penulis dapat menyelesaikan tesis ini sebagai salah satu syarat akademik untuk meraih gelar Magister Sains pada Program Studi Magister Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Penulisan tugas akhir ini tentunya tidak lepas dari bantuan berbagai pihak baik moril maupun materil. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang tak terhingga untuk Ayah tercinta **Rahman B** yang tak kenal lelah mendoakan, memberikan dukungan, dan selalu melimpahkan cinta dan kasih sayangnya kepada penulis dan tugas akhir ini penulis persembahkan untuk ibunda tercinta Almarhumah **Harina**, semoga beliau bangga dengan perjuangan anaknya. Mereka menjadi motivasi terbesar penulis dalam menyelesaikan tugas akhir ini. Untuk saudara tercinta **Fatir Ma'rif** terima kasih telah memberikan semangat motivasi, dukungan dan doa yang diberikan kepada penulis.

Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan juga penulis ucapkan kepada:

1. Prof. Dr. Ir. Jamaluddin Jompa, M.Sc. selaku Rektor Universitas Hasanuddin.
2. Dr. Eng. Amiruddin, M.Si. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
3. Dr. Anna Islamiyati, S.Si., M.Si. selaku Ketua Departemen Statistika Fakultas MIPA Universitas Hasanuddin yang telah bersedia memberikan masukan-masukan, arahan, serta memberikan dukungan dalam penyelesaian tesis ini.
4. Dr. Erna Tri Herdiani, S.Si., M.Si. selaku Ketua Program Studi Magister Statistika Departemen Statistika Fakultas MIPA Universitas Hasanuddin yang menjadi salah satu tim penguji yang telah memberikan arahan dan dukungan dalam penyelesaian tesis ini.
5. Sri Astuti Thamrin, S.Si., M.Stat., Ph.D. selaku Pembimbing Utama yang sangat bersabar dan bersedia meluangkan banyak waktunya untuk membimbing penulis serta memberikan ilmu, dukungan, masukan motivasi, dan kemudahan kepada penulis dalam menyelesaikan tesis ini.
6. Dr. Nirwan, M.Si. selaku Pembimbing Pertama yang sangat bersabar dan bersedia meluangkan banyak waktunya untuk membimbing penulis serta memberikan ilmu, dukungan, masukan motivasi, dan kemudahan kepada penulis dalam penyelesaian tesis ini.
- Sanusi, S.Si., M.Si. selaku penguji penulis yang telah bersedia memberikan masukan-masukan dan arahan dalam penyusunan tesis.



8. Prof. Dr. Dr. Georgina Maria Tinungki, M.Si. selaku penguji penulis yang telah bersedia memberikan masukan-masukan dan arahan dalam penyusunan tesis.
9. Bapak dan Ibu Dosen serta Staf Departemen Statistika Fakultas MIPA Universitas Hasanuddin, yang dengan tulus ikhlas memberikan ilmu pengetahuan dan pengalaman yang dimilikinya selama perkuliahan berlangsung sehingga memberikan banyak manfaat bagi penulis untuk saat ini maupun di masa mendatang.
10. Bapak Darman dan Ibu Satria selaku orang tua kedua penulis, yang senantiasa mendoakan, memberikan dukungan, motivasi, serta selalu melimpahkan cinta dan kasih sayangnya kepada penulis.
11. Sahabat sekaligus saudara terbaik penulis Fani, Alfi, Ainun, Wana, Tata, Ainan dan wiwi terima kasih atas nasehat, dukungan yang luar biasa dan kebersamaanya yang telah dilalui bersama penulis dalam menyelesaikan tesis ini.
12. Seluruh teman-teman Mahasiswa Program Studi Magister Statistika 2020 Alfi Nur Khauly, Nurul Fadhilah, A. Nurhayati Latief, Ardiansyah Abubakar, Firman Saputra dan Bakry. Terima kasih atas dukungan luar biasa kepada penulis.
13. Teman-teman Mahasiswa Program Magister Statistika angkatan I sampai angkatan IX, terkhusus Muammar Ashari Abuspin, Missrianti, Khairunnisa Abdullah dan Mubasirotul Munawaroh, terima kasih atas bantuan yang luar biasa kepada penulis.
14. Semua pihak yang telah membantu dalam penyelesaian tesis ini yang tidak dapat penulis sebutkan satu per satu. Jazakumullah Khairan Katsiran.

Semoga Allah SWT memberikan pahala yang berlipat ganda atas segala kebaikan yang telah diberikan kepada penulis dan semoga penulisan tesis ini bermanfaat bagi perkembangan ilmu pengetahuan dan teknologi, khususnya dalam dunia statistika dan data sains.

Penulis,

Mita Astuti



ABSTRAK

MITA ASTUTI. **Perbandingan Metode *Random Forest* dan *Naive Bayes* pada Klasifikasi Perilaku Mahasiswa di LMS SPADA Indonesia** (dibimbing oleh Sri Astuti Thamrin, dan Nirwan).

Latar Belakang. *Educational Data Mining* (EDM) merupakan aturan yang digunakan untuk mengembangkan metode untuk mendapatkan informasi yang berguna dari dataset pendidikan dan dapat dijadikan sebagai solusi untuk dunia pendidikan. Beberapa metode pembelajaran terawasi pada *data mining* yang umum digunakan untuk melakukan klasifikasi data sistem pembelajaran diantaranya *Random Forest* dan *Naive Bayes*. **Tujuan.** Studi ini bertujuan mengklasifikasikan data yang berasal dari *Learning Management System* (LMS) di Sistem Pembelajaran dalam Jaringan (SPADA) Indonesia. Selain itu, tujuan lainnya untuk mendapatkan akurasi prediksi klasifikasi. **Metode.** Atribut yang digunakan dalam studi ini terdiri atas 7 variabel, 6 variabel atribut dan 1 variabel kelas dengan status (*Intuitive* atau *Sensing*). Sebanyak 109.807 data mahasiswa yang mengikuti LMS SPADA Indonesia. Pada studi ini digunakan proporsi data latih 80% (87.845 data) dan 20% data uji (21.962 data). Terdapat dua metode yang digunakan dalam studi ini yaitu metode *Random Forest* dan *Naive Bayes*. **Hasil.** Model *Random Forest* memperoleh nilai *bootstrap* dan *Tree Optimal* sebesar 200 pohon klasifikasi dan 2 pemilihan variabel acak dengan nilai *Error* terkecil 13,24%. Kinerja klasifikasi *Random Forest* memperoleh tingkat akurasi klasifikasi model sebesar 95%. Pada model *Naive Bayes*, nilai rata-rata dan standar deviasi digunakan untuk menentukan nilai densitas gaussian setiap kelas. Kinerja klasifikasi *Naive Bayes* memperoleh tingkat akurasi klasifikasi model sebesar 93%. **Kesimpulan.** Model *Random Forest* memperoleh nilai akurasi prediksi klasifikasi sebesar 95% dan metode *Naive Bayes* memperoleh nilai akurasi prediksi klasifikasi sebesar 93%.

Kata Kunci: Klasifikasi, *Random Forest*, *Naive Bayes*, LMS, SPADA Indonesia



ABSTRACK

MITA ASTUTI, **Comparison of Random Forest and Naive Bayes Methods for Classifying Student Behavior in LMS SPADA Indonesia** (Supervised by Sri Astuti Thamrin, dan Nirwan)

Background. Educational Data Mining (EDM) is a rule used to develop methods to obtain useful information from educational datasets and can be used as a solution for the world of education. Some supervised learning methods in data mining that are commonly used to classify learning system data include Random Forest and Naive Bayes. **Objective.** **Aims.** This study aims to classify data from the Learning Management System (LMS) in the Indonesian Online Learning System (SPADA). In addition, another goal is to obtain classification prediction accuracy. **Method.** The attributes used in this study consist of 7 variables, 6 attribute variables and 1 class variable with status (Intuitive or Sensing). A total of 109,807 student data participating in the SPADA Indonesia LMS. This study used a proportion of 80% training data (87,845 data) and 20% test data (21,962 data). There are two methods used in this study, namely the Random Forest and Naive Bayes methods. **Results.** The Random Forest model obtained bootstrap and Tree Optimal values of 200 classification trees and 2 random variable selections with the smallest Error value of 13.24%. The Random Forest classification performance obtained a model classification accuracy rate of 95%. In the Naive Bayes model, the average value and standard deviation are used to determine the gaussian density value of each class. The Naive Bayes classification performance obtained a model classification accuracy rate of 93%. **Conclusion.** The Random Forest model obtained a classification prediction accuracy value of 95% and the Naive Bayes method obtained a classification prediction accuracy value of 93%.

Keywords: Classification, Random Forest, Naive Bayes, Learning Mangement System, SPADA Indonesia



DAFTAR ISI

HALAMAN JUDUL.....	i
PERNYATAAN PENGAJUAN.....	ii
HALAMAN PENGESAHANTESIS	iii
PERNYATAAN KEASLIAN TESIS.....	iv
UCAPAN TERIMA KASIH.....	v
ABSTRAK.....	vi
ABSTRACK.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan masalah.....	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah.....	3
1.6 Kajian Teori.....	3
BAB II METODE PENELITIAN	12
2.1 Sumber Data	12
2.2 Metode Analisis	13
BAB III HASIL DAN PEMBAHASAN	15
3.1 Deskripsi Data	15
3.2 Penyaringan Data.....	15
3.3 Data Latih dan Data Uji.....	15
3.4 Metode Random Forest.....	16
3.5 Metode Naive Bayes	19
3.6 Perbandingan Model	24
BAB IV REKOMENDASI DAN SARAN	25
4.1 Kesimpulan.....	25
4.2 Saran.....	26
4.3 Daftar Pustaka.....	28



DAFTAR TABEL

Tabel 1 Ilustrasi <i>confusion matrix</i>	10
Tabel 2 Variabel Penelitian.....	12
Tabel 3 Deskripsi data di LMS SPADA Indonesia.....	15
Tabel 4 <i>Confusion Matrix</i> Menggunakan Data Uji.....	17
Tabel 5 Kinerja klasifikasi <i>Model Random Forest</i>	18
Tabel 6 <i>Confusion Matrix</i> Menggunakan Data Uji.....	18
Tabel 7 Kinerja Klasifikasi Model <i>Random Forest</i>	19
Tabel 8 Nilai Probabilitas Data Latih.....	19
Tabel 9 Mean dan Standar Deviasi.....	20
Tabel 10 Probabilitas Munculnya Fitur Forum <i>Post</i>	20
Tabel 11 Nilai Data Uji.....	21
Tabel 12 Nilai Densitas <i>Gaussian</i> Kelas <i>Intuitive</i> dan <i>Sensing</i>	21
Tabel 13 Deskripsi Data Uji.....	23
Tabel 14 <i>Confusion Matrix</i> Menggunakan Data Uji.....	23
Tabel 15 Kinerja Klasifikasi Model Naive Bayes.....	23
Tabel 16 Perbandingan Model Random Forest dan Naive Bayes.....	24



DAFTAR GAMBAR

Gambar 1. Ilustrasi Pembentukan Model <i>Random Forest</i>	8
Gambar 2. Plot <i>OOB Error</i> model <i>random forest</i>	14



BAB I

PENDAHULUAN

1.1 Latar Belakang

Aktivitas mahasiswa dapat digunakan untuk memprediksi kinerja mahasiswa pada suatu perguruan tinggi. Aktivitas mahasiswa diperoleh pada saat proses pembelajaran berlangsung, namun setelah ditetapkannya COVID-19, pembelajaran tidak lagi dilakukan secara luring melainkan secara daring. Salah satu sistem pembelajaran dalam jaringan yang dikeluarkan oleh Kementerian Pendidikan dan Kebudayaan Riset, dan Teknologi yaitu SPADA (Sistem Pembelajaran dalam Jaringan Indonesia).

Sistem pembelajaran dalam jaringan Indonesia (SPADA) merupakan program Direktorat Jenderal Pembelajaran dan Kemahasiswaan Kementerian Pendidikan, Riset, dan Teknologi yang bertujuan meningkatkan pemerataan terhadap pembelajaran yang bermutu di perguruan tinggi. SPADA Indonesia memberikan peluang bagi mahasiswa untuk mengikuti suatu mata kuliah dari perguruan tinggi lain dan hasil belajarnya dapat diakui oleh perguruan tinggi di mana mahasiswa tersebut terdaftar. LMS (*Learning Management System*) merupakan perangkat berbasis web yang digunakan untuk merancang, mengimplementasikan, dan mengevaluasi pembelajaran, memfasilitasi interaksi mahasiswa, memberikan umpan balik kinerja dan mengatur aktivitas mahasiswa (Mohd Kasim & Khalid, 2016). LSM SPADA Indonesia menyediakan *datalog* perilaku mahasiswa seperti jumlah revisi kuis, jumlah kunjungan tugas, jumlah konten yang dikunjungi, lama waktu pada konten, forum diskusi yang dikunjungi dan forum diskusi yang dibuat. Perilaku mahasiswa terekam oleh LSM dan tersimpan dalam skala besar yang berpotensi berguna bagi proses pendidikan.

Educational data mining (EDM) menurut Jovanovic dkk, (2012) merupakan pengembangan metode dalam memperoleh pengetahuan dari data yang berasal dari lingkungan pendidikan (tradisional atau pembelajaran jarak jauh). Teknologi informasi dan komunikasi dapat mempermudah dalam mengatasi permasalahan dalam dunia pendidikan dengan cepat sehingga lebih efisien, terutama dalam ranah komputasi. EDM merupakan salah satu pemanfaatan teknologi yang mendapat perhatian selama satu dekade terakhir (Purba, 2012). Dunia pendidikan menyimpan banyak *database* yang dapat dimanfaatkan. *Database* memuat informasi yang berpotensi untuk digali sehingga dapat menghasilkan pola-pola yang baru yang dapat digunakan untuk mengefisienkan proses yang ada dalam pendidikan. Menurut Bilal Zoric (2020), penggunaan metode klasifikasi dalam EDM digunakan dalam pengelompokan data berdasarkan keterikatan terhadap data sampel kemudian menargetkan atribut atau kelas yang dapat membantu dalam menganalisis data dan memprediksi hasil. Ada beberapa metode klasifikasi yang dapat digunakan dalam EDM seperti *neural network*, *random forest*, *decision tree*, *naive bayes*, dan *logistic regression* (Adekitan & Salau, 2019).



Decision tree adalah algoritma *Supervised Learning* yang digunakan untuk klasifikasi dengan cara membangun *Multiple Decision Tree* pada saat pelatihan kemudian mendapatkan beberapa prediksi kemudian dipilih satu prediksi terbaik (Sugianto, 2021). *Random forest* merupakan salah satu metode *decision tree* yang menggunakan konsep *bagging* yang secara umum melakukan pemilihan label

kelas terbanyak dari hasil beberapa model di mana model tersebut digabungkan sebagai pengklasifikasian utama dari model akhir yang dihasilkan (Aqsha dkk, 2021). Dalam penelitian Bulut & Yavuz (2019) yang menggunakan klasifikasi *Random Forest*, *Classification and Regression Tree*, dan *Support Vector Machine* untuk memprediksi tingkat kemahiran siswa pada tes literasi sains PISA Siswa tahun 2015 dan menunjukkan bahwa metode *Random Forest* merupakan algoritma dengan performa terbaik. Selain itu dengan menggunakan metode *Random Forest* tingkat kesalahan model tidak berubah setelah 100 *trees*, artinya model yang sama dapat diestimasi hanya dengan 100 *trees* untuk mendapatkan model akhir yang lebih efisien.

Naive Bayes merupakan salah satu algoritma yang dapat digunakan dalam klasifikasi karena mampu mengklasifikasikan data dengan sederhana dan menghitung probabilitas dengan menjumlahkan frekuensi dan nilai dari data yang ada. *Naive Bayes* berdasar kepada teorema *Bayes* yang digunakan untuk menghitung probabilitas dari tiap kelas dengan asumsi setiap kelas saling independen atau tidak saling tergantung (Sutoyo & Almaarif, 2020). *Naive Bayes* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam *database* yang besar (Muhamad dkk, 2017). Yaacob dkk, (2019) menggunakan 4 teknik klasifikasi yaitu *K Nearest neighbor*, *Naive Bayes*, *Decision tree* dan regresi logistik yang diterapkan untuk memprediksi prestasi siswa. Dalam penelitian Yacoob et al, *Naive Bayes* mengungguli algoritma lainnya dengan akurasi sebesar 85,4%. Dalam penelitian Mueen dkk, (2016) juga menggunakan algoritma klasifikasi *data mining Naive Bayes*, *Multilayer Perception* dan *Decision Tree* (C4.5) untuk memprediksi prestasi akademik siswa. Klasifikasi *Naive Bayes* mengungguli dua pengklasifikasi lainnya dengan mencapai akurasi prediksi keseluruhan sebesar 86%. Kinerja model klasifikasi diukur dengan mengevaluasi ketepatan keputusan klasifikasi menggunakan *Confusion Matrix*.

Salah satu masalah yang diperoleh dari penelitian (Jovanovic dkk, 2012) adalah kurangnya data untuk analisis yang menyeluruh yang menargetkan banyak lembaga pembelajaran jarak jauh. Berbeda dengan LMS SPADA Indonesia yang menyediakan *datalog* perilaku mahasiswa dalam jumlah yang besar dan dari beberapa perguruan tinggi di Indonesia sehingga perlu dilakukan analisis menggunakan metode *Random Forest* dan *Naive Bayes* untuk memprediksi kinerja mahasiswa. *Random Forest* digunakan karena tingkat akurasi dari metode ini lebih tinggi dari metode lainnya, di mana *Random Forest* akan membuat pohon klasifikasi dengan jumlah yang banyak kemudian digabungkan untuk mencari tingkat akurasi yang tinggi (Yahya, 2018), sedangkan *Naive Bayes* digunakan karena pengkategorian dilakukan berdasarkan probabilitas yang sederhana dan dirancang untuk menggunakan asumsi bahwa antar satu kelas dengan kelas yang lain tidak saling tergantung, serta hasil klasifikasi yang diperoleh memiliki nilai akurasi yang tinggi (Berry & Browne, 2006).



menjadi acuan dalam menentukan model yang baik yaitu dengan membandingkan hasil dari model yang digunakan. Salah satu hal yang mempengaruhi akurasi yang tinggi yaitu jumlah data yang besar, pembagian data, dan jumlah kelas mayor dan kelas minor pada data (Kuswanto, 2019). Semakin banyak data yang digunakan dalam membangun model klasifikasi, maka

nilai keakuratan model menjadi lebih tinggi. Semakin konsisten kelas data latih yang digunakan maka semakin akurat pula model yang dihasilkan (Santosa, 2007). Sebanyak 109.807 data yang mencakup aktivitas mahasiswa dalam SPADA Indonesia. Klasifikasi pada perilaku mahasiswa di SPADA Indonesia bertujuan untuk memprediksi kelas dari objek yang akan digunakan, dimana objek yang digunakan tidak diketahui kelasnya. Klasifikasi pada perilaku mahasiswa dapat membantu tenaga pendidik dalam menyelesaikan permasalahan yang dihadapi, seperti membantu dalam melakukan penilaian terhadap mahasiswa terkhusus untuk pembelajaran dalam jaringan.

1.2 Rumusan masalah

Berdasarkan latar belakang, maka dapat dirumuskan permasalahan sebagai berikut:

1. Bagaimana akurasi prediksi klasifikasi menggunakan metode *Random Forest* ?
2. Bagaimana akurasi prediksi klasifikasi menggunakan *Naive Bayes* ?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mendapatkan akurasi prediksi klasifikasi menggunakan metode *Random Forest*.
2. Mendapatkan akurasi prediksi klasifikasi menggunakan metode *Naive Bayes*.

1.4 Manfaat Penelitian

Manfaat penelitian pada penelitian ini adalah sebagai berikut:

1. Hasil dari penelitian ini diharapkan data menjadi solusi terhadap persoalan pendidikan di perguruan tinggi.
2. Hasil dari penelitian ini diharapkan dapat memberikan informasi tentang akurasi prediksi klasifikasi *Random Forest* dan *Naive Bayes* ketika diterapkan pada data SPADA Indonesia.

1.5 Batasan Masalah

Batasan masalah dalam penelitian ini dibatasi pada hasil keakuratan klasifikasi dalam menggunakan metode *random Forest* dan *Naive Bayes*. Data yang digunakan dalam penelitian ini adalah aktivitas mahasiswa di SPADA Indonesia yang diperoleh dari <https://spada.kemdikbud.go.id>.

1.6 Kajian Teori

1.6.1 *Educational Data Mining*

EDM (*Educational Data Mining*) adalah disiplin yang muncul dari penerapan *data mining* yang mengembangkan metode untuk mendapatkan informasi unik yang berasal dari data pendidikan. Informasi unik tersebut kemudian dievaluasi agar dapat digunakan



permasalahan-permasalahan dalam dunia pendidikan seperti siswa, memprediksi kegagalan sekolah dan sebagainya (Hidayat a pendidikan dapat menggunakan *data mining* untuk menemukan *database* yang dikenal sebagai EDM. *Educational Data Mining* pendidikan dalam hal (Kamath dkk, 2016):

Perbandingan kinerja siswa antara hasil ujian sebelumnya dengan mereka saat memperkirakan kinerja akademik siswa

- b. Deteksi subjek yang sangat terkait dalam silabus
- c. Penemuan pengetahuan tentang prestasi akademik
- d. Klasifikasi prestasi akademik siswa menurut gaya belajar
- e. Menemukan persamaan dan perbedaan antar sekolah

Data mining disebut juga *knowledge discovery* karena terdapat pengambilan pola pada data besar yang kemudian diproses dan menghasilkan informasi yang penting. terdapat beberapa istilah lain yang memiliki makna sama dengan data mining, yaitu *knowledge discovery in databases* (KDD). Istilah *data mining* dan *knowledge discovery in databases* sering digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu kumpulan data yang besar. Proses KDD menurut Kamath & Kamat, (2016) secara garis besar terdiri dari:

1. *Data Selection*

Pemilihan (seleksi) data dan sekumpulan data operasional, dilakukan sebelum tahap penggalian informasi. Data hasil seleksi yang akan digunakan untuk proses *data mining*, disimpan dalam suatu berkas terpisah dari basis data operasional.

2. *Data Preparation*

Data Preparation merupakan hal yang penting dalam proses *Data Mining*. Ilmuwan data menghabiskan 60% waktu mereka untuk membersihkan dan mengatur data. Mengumpulkan set data menempati urutan kedua dengan 19% dari waktu mereka, yang berarti para ilmuwan data menghabiskan sekitar 80% dari waktu mereka untuk menyiapkan dan mengolah data untuk analisis (Gil Press, 2016). *Data Mining* menciptakan model dan beroperasi secara eksklusif pada kumpulan data. persiapan untuk *Data Mining* melibatkan variabel secara individual serta mengunci kumpulan data secara keseluruhan. Sebelum menerapkan data algoritma *Data Mining*, penting untuk melakukan beberapa tugas pra-pemrosesan seperti pembersihan data, integrasi, transformasi, dan diskritisasi (Widyahastuti & Tjhin, 2018).

3. Transformasi Data

Transformasi data merupakan proses konversi data sebelum melakukan proses *data mining*. Melakukan transformasi data bertujuan agar data tersebut sesuai ketika dilakukan proses *data mining*. Langkah-langkah transformasi data yaitu pengumpulan (*capturing*), memilih (*verifying*), mengelompokkan (*classifying*), penyusunan (*sorting*), meringkas (*summarizing*), perhitungan (*calculating*), dan penyimpanan (*storing*).

4. *Data Mining*

Pada tahap ini proses mencari pola atau informasi yang menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu seperti metode *random forest* dan *naive bayes*. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.



ola informasi yang diperoleh akan ditampilkan dalam bentuk yang i oleh pihak yang berkepentingan. Serta melakukan pemeriksaan u informasi yang diperoleh bertentangan dengan fakta atau ida sebelumnya.

1.6.2 Klasifikasi

Ada berbagai jenis tugas EDM yang populer diantaranya klasifikasi, prediksi, aturan asosiasi, pengelompokan, deteksi pencilan, dan lain-lain. Menurut (Kamath dkk, 2016) tujuan dari metode klasifikasi adalah untuk mengklasifikasikan data set ke kelas atau ke kelompok yang telah ditentukan. Metode ini merupakan teknik terawasi dan disebut juga pembelajaran langsung karena kelas-kelasnya telah ditentukan sebelumnya, kemudian mengekstraksi pola pada data target.

Classification and Regression Tree (CART) merupakan Teknik pembelajaran *decision tree* non parametrik yang menghasilkan klasifikasi dan pohon regresi, dimana variabel dependen adalah masing-masing terdiri dari variabel kategorik dan numerik. *Decision tree* membuat partisi hierarkis data yang menghubungkan partisi yang berbeda pada tingkat simpul akar (*root node*) sampai simpul terakhir (*leaf node*). Partisi hierarkis pada setiap simpul dibuat dengan menggunakan kriteria *split*. Kriteria *split* dapat menggunakan kondisi atau predikat pada atribut tunggal, atau berisi kondisi pada beberapa atribut. yang pertama disebut pemisah univariat, sedangkan yang terakhir disebut sebagai pemisah multivariat. Diskriminasi antara kelas yang berbeda dalam simpul (*node*) dimaksimalkan menggunakan *Gini Impurity*. *Gini Impurity* atau pengukuran ketidakmurnian gini adalah suatu metode yang digunakan dalam menentukan pohon keputusan dalam menentukan pemisahan optimal dari simpul akar, dan pemisahan berikutnya. Indeks gini adalah ukuran sederhana dari distribusi yang koefisiennya antara 0 hingga 1. Jika kelas pada atribut menghasilkan kelas target yang berbeda pada node, maka kelas tersebut yang terpilih untuk di *split*. Langkah-langkah *Decision Tree* adalah sebagai berikut (Kuswanto, 2019):

1. Pemilihan (*Classifier*)

Setiap penyekatan tergantung pada nilai yang hanya berasal dari satu peubah penjelas. Untuk peubah kategorik penyekatan yang berasal dari semua kemungkinan penyekatan berdasarkan terbentuknya dua anak gugus yang saling lepas (*disjoint*). Jika peubah X_j merupakan peubah kategorik nominal dengan L kategori, maka ada $2^{L-1} - 1$ penyekatan, sedangkan jika peubah kategorik ordinal, maka ada $L - 1$ penyekatan yang mungkin.

Data yang diklasifikasikan dalam *Decision Tree* menggunakan *Gini Impurity* yang umum digunakan untuk memecahkan permasalahan pohon klasifikasi. Data yang diklasifikasikan dalam *decision tree* harus diketahui nilai *Gini Impurity* setiap kelas karena dalam penentuan simpul akar, simpul dalam, dan simpul terakhir didasarkan pada penyekatan *Goodness of Split* kelas yang ada. Misalkan 1 adalah semua pengamatan pada node t kelas ke- j dan 0 untuk yang lainnya. Besarnya suatu sampel pada Gini Impurity sebagai berikut:



$$- p(j|t) \quad (1)$$

lai *Gini Impurity* masing-masing atribut. *Gini Impurity* pada node t persamaan sebagai berikut:

$$\begin{aligned}
 i(t) &= \sum_{j=1}^k p(j|t)(1 - p(j|t)), \\
 &= \sum_{j=1}^k p(j|t) - \sum_{j=1}^k p^2(j|t), \\
 i(t) &= 1 - \sum_{j=1}^k p^2(j|t). \tag{2}
 \end{aligned}$$

Evaluasi penyekatan menggunakan *goodness of split* $\varphi(s, t)$ dari penyekat s pada simpul t . Fungsi penyekatan yang didefinisikan sebagai penurunan heterogenitas adalah:

$$\varphi(s, t) = i(t) - p_l i(t_l) - p_r i(t_r). \tag{3}$$

Pemilah yang menghasilkan (s, t) yang lebih tinggi merupakan pemilah yang lebih baik karena pemilah tersebut memungkinkan untuk mereduksi nilai impuritas secara lebih signifikan. Jika simpul yang dihasilkan merupakan kelas yang tidak homogen maka prosedur yang sama akan diulangi sampai pohon.

$$\varphi(s^*, t) = \max_{s \in S} \varphi(s, t). \tag{4}$$

di mana:

- $i(t)$: fungsi heterogenitas pada simpul t
- p_l : proporsi pengamatan simpul kiri
- p_r : proporsi pengamatan menuju simpul kanan
- $i(t_l)$: fungsi heterogenitas pada simpul anak kiri
- $i(t_r)$: fungsi heterogenitas pada simpul anak kanan

Setelah simpul utama diperoleh, penentuan simpul anak dilakukan dengan menggunakan cara yang sama.

2. Penentuan simpul akhir

suatu simpul t akan menjadi simpul akhir atau tidak, akan dipilih Kembali bila pada simpul t tidak terdapat penurunan keheterogenan secara berarti atau adanya Batasan minimum n seperti hanya terdapat satu pengamatan pada tiap simpul anak. Jumlah kasus minimum dalam suatu terminal akhir umumnya adalah 5, dan apabila hal itu terpenuhi maka pengembangan pohon dihentikan.

3. Penandaan label kelas



1 kelas pada simpul akhir dilakukan berdasarkan aturan jumlah as simpul akhir t adalah j_0 yang memberi nilai dugaan kesalahan pul t terbesar. Proses pembentukan pohon klasifikasi berhenti saat pengamatan dalam tiap-tiap simpul anak atau adanya Batasan pengamatan dalam tiap simpul anak identik, dan adanya Batasan dalam pohon maksimal.

$$p(j_0|t) = \max_j p(j|t) = \max_j \frac{N_j(t)}{N(t)}. \quad (5)$$

Dimana:

- $p(j|t)$: proporsi kelas j pada simpul
 $N_j(t)$: jumlah pengamatan kelas j pada simpul t
 $N(t)$: jumlah pengamatan pada simpul t

1.6.3 Random Forest

Random Forest digunakan untuk membangun pohon keputusan yang terdiri dari *root node*, *internal node*, dan *leaf node* dengan mengambil atribut dan data secara acak sesuai ketentuan yang diberlakukan. *Root node* merupakan simpul yang terletak paling atas atau biasa disebut sebagai akar dari *Decision Tree*. *Internal node* adalah simpul percabangan yang mempunyai *output* minimal dua dan hanya ada satu *input*. Sedangkan *leaf node* atau terminal *node* merupakan simpul terakhir yang hanya memiliki satu *input* dan tidak mempunyai *output* (Siburian & Mulyana, 2018).

Algoritma *Random Forest* menurut (Tibshirani & Friedman, 2008):

Input:

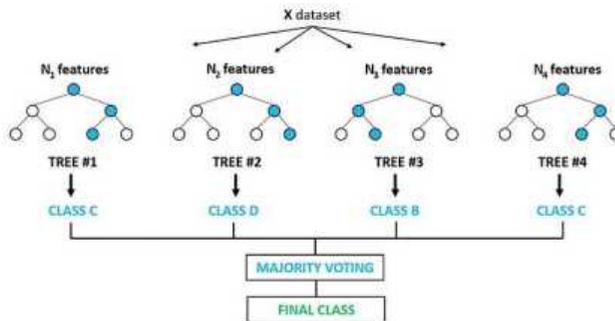
1. Data D dengan n sampel $\{X_i, y_i\}, i = 1, \dots, n$, dimana n adalah sampel keseluruhan, X_i dan p adalah jumlah fitur pada data.
2. Menentukan nilai B dan m , dimana B adalah jumlah pohon dan m adalah jumlah variabel yang akan dipilih secara acak saat pembuatan pohon.

Prosedur:

1. $b = 1$ sampai B
 - a. Dilakukan pengambilan secara acak sampel *Bootstrap* $Z^* \in D$ dengan ukuran n dari data latih.
 - b. Buat pohon (*tree*) T_b dari data sampel *bootstrap* dengan cara melakukan Langkah sebagai berikut secara rekursif untuk setiap simpul dari pohon (*tree*), sampai ukuran minimal simpul dicapai atau hingga data sudah tidak bisa dibagi lagi.
 - i. Pilih m variabel secara acak.
 - ii. Ambil variabel dengan penyekatan terbaik diantara m variabel.
 - iii. Menentukan simpul akhir
2. keluaran dari model adalah *Random Forest* $\{T_B\}_1^B$, dimana $\{T_B\}_1^B$ adalah kumpulan dari pohon-pohon yang telah dibangun, dimulai dari pohon pertama sampai pohon B .
 1 prediksi pada titik baru dengan cara memilih label kelas terbanyak yang telah dibentuk.



Random Forest Classifier



Gambar 1. Ilustrasi pembentukan model *Random Forest*

1.6.4 Naive Bayes

Naive bayes adalah suatu pengklasifikasian probabilistik sederhana yang melakukan perhitungan terhadap sekumpulan probabilistik dengan penjumlahan frekuensi dan gabungan nilai dari data set (Huda dkk, 2020). *Naive bayes* memerlukan data pelatihan untuk mengestimasi parameter yang dibutuhkan untuk klasifikasi. Selain itu, *Naive Bayes* mampu menangani nilai yang hilang dengan mengabaikan atribut selama perhitungan estimasi peluang. Kemudian hasil dari model klasifikasi tersebut akan dibandingkan (Mayadewi & Rosely, 2015).

Klasifikasi *Naive Bayes* menurut Leung (2007) bekerja sebagai berikut:

1. Misalkan T adalah himpunan sampel pelatihan, masing-masing dengan label kelasnya. Ada kelas k yang terdiri dari C_1, C_2, \dots, C_k . Setiap sampel diwakili oleh vektor n -dimensi, $X = \{x_1, x_2, \dots, x_n\}$, menggambarkan n nilai terukur dari n atribut.
2. Menentukan probabilitas posterior *maximum* untuk memperoleh $P(C_i|X)$, X termasuk ke dalam kelas C_i jika dan hanya jika:

$$P(C_i|X) > P(C_j|X) \quad \text{untuk } 1 \leq j \leq m, \quad j \neq i, \quad (6)$$

dimana $P(C_i|X)$ dimaksimalkan disebut hipotesis posterior maksimum. Teorema bayes:

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad (7)$$



n untuk setiap kelas, sehingga hanya perlu memaksimalkan jika probabilitas kelas apriori $P(C_i)$ tidak diketahui, maka bahwa kelas-kelas tersebut memiliki peluang yang sama yaitu $\dots = P(C_k)$, sehingga perlu memaksimalkan $P(X|C_k)$ jika tidak $\langle n P(X|C_i)P(C_i)$. Perhatikan bahwa probabilitas kelas apriori dapat dengan:

$$P(C_i) = \frac{\text{freq}P(C_i, T)}{|T|} \quad (8)$$

4. Diasumsikan bahwa nilai-nilai atribut adalah independen bersyarat satu sama lain, dengan label kelas sampel:

$$P(X|C_i) \approx \prod_{k=1}^n P(x_k|C_i) \quad (9)$$

Probabilitas $P(x_1|C_i), P(x_2|C_i), \dots, P(x_n|C_i)$ dapat dengan mudah diestimasi dari training set. Nilai x_k mengacu pada nilai atribut A_k untuk sampel X .

- Jika A_k adalah kategorik, maka $P(x_k|C_i)$ adalah banyaknya sampel kelas C_i di T yang memiliki nilai x_k untuk atribut A_k , dibagi dengan frekuensi C_i di T .
- Jika A_k bernilai kontinu, maka diasumsikan bahwa nilai memiliki distribusi *Gaussian* dengan mean (μ) dan standar deviasi (σ) didefinisikan dengan

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (10)$$

dimana:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (X_i - \mu)^2}{n}}, \quad (11)$$

dan

$$\mu = \frac{\sum_{i=1}^n X_i}{n}, \quad (12)$$

sehingga:

$$p(x_k|C_i) = g(x_k, \mu_{C_i}, \sigma_{C_i}), \quad (13)$$

menghitung μ_{C_i} dan σ_{C_i} yang merupakan mean dan standar deviasi dari nilai atribut A_k untuk sampel training dari kelas C_i .

5. Untuk memprediksi label kelas dari X , $P(X|C_i)P(C_i)$ dievaluasi untuk setiap kelas C_i . Klasifikasi memprediksi bahwa label kelas dari X adalah C_i jika dan hanya jika kelas tersebut memaksimalkan $P(X|C_i)P(C_i)$.

1.6.5 Kinerja Klasifikasi

Evaluasi akurasi model klasifikasi menggunakan metode *confusion matrix* untuk mengetahui apakah model yang digunakan baik atau tidak. Berikut ilustrasi *confusion matrix* menurut (Sastrawan dkk, 2019).



Tabel 1. Ilustrasi *confusion matrix*

Kelas	Aktual	Aktual
	Positif	Negatif
Prediksi Positif	TP (Benar)	FP (Salah)
Prediksi Negatif	FN (Salah)	TN (Benar)

TP (*True positive*) masuk kategori benar karena data diprediksi positif, kemudian setelah diidentifikasi data bernilai positif sehingga berhasil diprediksi dengan benar. FP (*False positive*) masuk kategori salah karena data diprediksi positif, kemudian setelah diidentifikasi, data bernilai negatif sehingga tidak berhasil memprediksi dengan benar (salah). FN (*False negative*) masuk kategori salah karena data diprediksi negatif, kemudian setelah diidentifikasi, data bernilai positif sehingga tidak berhasil memprediksi dengan benar (salah). TN (*True negative*) masuk kategori benar karena data diprediksi negatif, kemudian setelah diidentifikasi, data bernilai negatif, sehingga data berhasil diprediksi dengan benar.

Berdasarkan nilai TP, TN, FP, dan FN, diperoleh nilai *Accuracy*, *Precision*, *Recall* (Olson & Delen, 2008).

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)} \quad (14)$$

$$Precision = \frac{(TP)}{(TP + FP)} \quad (15)$$

$$Recall = \frac{(TP)}{(TP + FN)} \quad (16)$$

Nilai akurasi menggambarkan seberapa akurat sistem dapat mengklasifikasikan data secara benar dengan keseluruhan data. Nilai presisi menggambarkan jumlah data kategori positif yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasi positif. *Recall* menunjukkan berapa persen data kategori positif yang terklasifikasi dengan benar oleh sistem (Rasila & Ristian, 2019)

1.6.6 SPADA Indonesia

Learning Management System (LMS) adalah aplikasi perangkat lunak atau teknologi berbasis web yang digunakan untuk merencanakan, mengimplementasikan, dan menilai proses pembelajaran tertentu (Alias & Zainuddin, 2005). SPADA Indonesia berbasis LMS, menyediakan pilihan bagi pengajar dalam membuat dan menyampaikan aktivitas pelajar, dan menilai kinerja pelajar secara *online* serta bagi mahasiswa dalam menggunakan fitur interaktif seperti diskusi, video, dan forum diskusi.



Indonesia telah dikelompokkan menjadi 8 label berdasarkan model *Learning Model* (FLSM). FLSM merupakan salah satu model yang mengklasifikasikan gaya belajar seorang pelajar. *Felder-silverman* gaya belajar seseorang secara lebih rinci. Membedakan preferensi

menjadi empat dimensi diantaranya: (1) pemrosesan informasi yang aktif dan reflektif, (2) jenis informasi yang disukai pelajar: sensing atau intuitive, (3) saluran sensorik yang paling efektif: visual atau verbal, (4) perkembangan pelajar dalam memahami: sekuensial atau global. Dalam penelitian ini akan gunakan satu dimensi untuk mengklasifikasi data yaitu jenis informasi yang disukai (*sensing* dan *intuitive*).



BAB II METODE PENELITIAN

2.1 Sumber Data

Data yang digunakan dalam penelitian ini yaitu data sekunder yang diperoleh dari *Learning Management System (LMS)* pada sistem pembelajaran daring (SPADA). SPADA Indonesia merupakan program Direktorat Jenderal Pembelajaran dan Kemahasiswaan Kementerian Riset, Teknologi dan Pendidikan Tinggi yang memiliki tujuan untuk meningkatkan pemerataan akses terhadap pembelajaran yang bermutu di perguruan tinggi. Adapun aplikasi yang digunakan dalam pengolahan data yaitu menggunakan *software R-studio version 4.2.0*. Adapun variabel yang digunakan pada penelitian ini adalah:

Tabel 2. Variabel Penelitian

Variabel	Nama Variabel	Skala
X_1	Jumlah revisi	Rasio
X_2	Jumlah kunjungan tugas	Rasio
X_3	Jumlah konten yang dikunjungi	Rasio
X_4	Lama waktu pada konten	Rasio
X_5	Forum diskusi yang dikunjungi	Rasio
X_6	Forum diskusi yang dibuat	Nominal
Y	dimensi model pembelajaran Felder-Silverman	Nominal



2.2 Metode Analisis

Penelitian ini berfokus pada pengembangan model *educational data mining* dengan membandingkan klasifikasi *random forest* dan *naive bayes*. Adapun metode analisis disusun sebagai berikut:

3.2.1 Tahap Persiapan Data

1. Identifikasi masalah yang terkumpul dalam basis data yang tersedia di LMS pada sistem pembelajaran daring (SPADA).
2. Melakukan seleksi atau pemilihan data yang relevan dengan analisis yang dilakukan.
3. *Preparation* data dan transformasi data yang terdiri dari:
 - a. Pembersihan data diterapkan untuk menghilangkan *noise* dan memperbaiki data yang tidak konsisten.
 - b. Reduksi data digunakan untuk mengurangi ukuran data. Misalnya menggabungkan, menghilangkan fitur yang berlebihan atau mengelompokkan transformasi.
4. Transformasi data yang telah dipilih agar data tersebut sesuai dengan proses *data mining*.

3.2.2 Tahap Pembagian Data

Membagi data ke dalam data latih dan data uji.

3.2.3 Tahap Pemodelan

1. *Random Forest*

- a. Membangun model *tree* menggunakan data latih.
- b. Membentuk data *bootstrap* dengan pemilihan sampel berulang (*Replacement*/pengembalian).
- c. Menentukan penyekat yang mungkin untuk setiap variabel penjelas berdasarkan aturan penyekatan berdasarkan persamaan 2 dan Persamaan 3.
- d. Memilih secara acak m variabel penjelas yang akan digunakan.
- e. Memilih penyekat terbaik dari m variabel penjelas yang dipilih secara acak sesuai persamaan 4.
- f. Menentukan simpul akhir.
- g. Lakukan langkah (b) sampai (e) secara berulang hingga menghasilkan B pohon klasifikasi yang diinginkan.

2. *Naive Bayes*

- c. Membangun model menggunakan data latih
- d. Data latih digunakan untuk membuat klasifikasi
- e. Menghitung $P(C_i)$ untuk setiap kelas,
- f. Menghitung $P(X|C_i)$ untuk setiap kriteria dan setiap kelas berdasarkan n , apabila data kontinu maka menggunakan persamaan 10. an nilai $P(C_i|X)$ dengan cara mengalikan $P(X|C_i)$ dengan $P(C_i)$ an $P(X|C_i)$ yang lebih besar berdasarkan persamaan 9 kemudian kesimpulan



3.2.4 Evaluasi Model

1. Melakukan prediksi dengan data uji dengan model klasifikasi yang diperoleh dari tahap pemodelan *Random Forest* dan *Naive bayes*.
2. Evaluasi metode klasifikasi menggunakan *confusion matrix* dengan menghitung nilai *accuracy*, *precision*, dan *recall* pada persamaan 14, persamaan 15, dan persamaan 16.

