

## ***Perbandingan Metode Naive Bayes Dan Random Forest Untuk Menentukan Prestasi Belajar Siswa Pada Jurusan RPL (Studi Kasus SMK Swasta Siti Banun Sigambal)***

*Ramadani<sup>1</sup>, B.Herawan Hayadi<sup>2</sup>*

Email: ramadans.ordinary@gmail.com<sup>1</sup>, b.herawan.hayadi@gmail.com<sup>2</sup>

<sup>1,2</sup> Ilmu Komputer, Fakultas Teknik dan Informatika, Universitas Potensi Utama

---

### **ABSTRACTS**

Pendidikan merupakan salah satu faktor kemajuan dan kemandirian bangsa. Semakin maju pendidikan suatu bangsa, maka akan semakin maju dan mandiri bangsa tersebut. Prestasi belajar merupakan salah satu aspek yang paling penting dalam bidang pendidikan dan menjadi harapan semua pihak. Bagi pihak Sekolah Menengah Kejuruan (SMK) Swasta Siti Banun Sigambal prestasi belajar Siswa merupakan salah satu indikator efektif proses belajar mengajar, yang sekaligus dapat digunakan untuk meningkatkan citra Sekolah, khususnya Sekolah Menengah Kejuruan (SMK) Swasta Siti Banun Sigambal. Sekolah Menengah Kejuruan (SMK) Swasta Siti Banun Sigambal prestasi belajar yang dicapainya oleh siswa dilihat dari nilai rapor. Dalam mewujudkan salah satu misi dari Sekolah Menengah Kejuruan (SMK) yakni Mewujudkan Pembelajaran yang bermutu di Sekolah. Dalam menganalisis dan memprediksi kinerja siswa dalam belajar dilakukan teknik data mining diantaranya adalah Naïve Bayes dan Random Forest. Dalam penelitian ini menggunakan 10-fold cross validation untuk memprediksi tingkat error dari data. Dataset dibagi dua yaitu data training dan data testing. Dari penelitian yang telah dilakukan maka didapatkan Random Forest memiliki tingkat akurasi yang paling tinggi dengan bobot 95.12%. Fitur yang paling berpengaruh adalah sekolah dengan bobot 0.202. Berdasarkan hasil tersebut dapat disimpulkan bahwa Random Forest memiliki accuracy yang paling tinggi dari metode algoritma Naïve Bayes.

---

### **ARTICLE INFO**

**Article History:**

*Received*

*Revised*

*Accepted*

*Available online*

---

**Keywords:**

*K-Nearest Neighbor,*

*Random Forest*

## I. PENDAHULUAN

Sekolah Menengah Kejuruan (SMK) Swasta Siti Banun berdiri sejak tahun 2002 pertama kali jurusan Reakayasa Perangkat Lunak (RPL) ini berdiri pada tahun 2006. Salah satu yang menjadi faktor suatu Sekolah Menengah Kejuruan termasuk yang berkualitas atau tidak salah satunya dapat dilihat dari prestasi yang diraih. Sebagaimana salah tujuan akhir pembelajaran adalah prestasi belajar, prediksi prestasi belajar peserta didik menjadi urgent untuk dilakukan. Prediksi prestasi belajar peserta didik dapat dijadikan sebagai dasar untuk melakukan intervensi dini terhadap potensi kegagalan peserta didik untuk mencapai tujuan pembelajaran dan di saat yang sama dapat melakukan perubahan strategi pembelajaran agar dapat memfasilitasi keragaman peserta didik [1].

Prestasi belajar peserta didik yang dimaksud dalam artikel ini adalah *students performance*. Namun secara kontekstual, *students performance* tidak dapat direduksi hanya menjadi prestasi belajar, namun juga mencakup hasil belajar. Hasil belajar adalah wujud konkret dari potensi yang dimiliki peserta didik, sebuah luaran dari proses perubahan tingkah laku setelah mengikuti pembelajaran. Wujud konkret hasil belajar dapat dilihat dari pemahamannya terhadap ilmu yang dipelajari, kepiawaiannya mengolah informasi dan mengambil keputusan yang didasarkan atas pemikiran tertentu atau pun keterampilan motoriknya [2].

Prestasi belajar adalah ukuran pencapaian peserta didik setelah mengikuti serangkaian kegiatan pembelajaran berupa skala penilaian [3]. Pemberian bobot atau skala penilaian untuk prestasi belajar menuntut disusunnya indikator penilaian, serta perlu dipastikan validitas dan reliabilitasnya. Nilai tersebutlah yang kemudian digunakan sebagai gambaran prestasi belajar dalam kurun waktu tertentu. Data peserta didik pada lembaga-lembaga pendidikan yang dapat diolah untuk melakukan prediksi prestasi belajar saat ini semakin banyak. Pesatnya digitalisasi dan perubahan pola pembelajaran dan tata kelola lembaga pendidikan dari yang sebelumnya berbasis luring menjadi berbasis daring turut mendorong melimpahnya data peserta didik dan bersifat sangat dinamis seperti data perilaku, jenis dan frekuensi aktivitas pembelajaran peserta didik yang sangat berguna untuk memprediksi prestasi belajar. Pemrograman eksplisit (*explicit programming*) tidak lagi memadai pada beberapa kasus untuk melakukan prediksi prestasi belajar peserta didik dalam kondisi data yang sangat besar dan dinamis ini. Dalam hal ini, pembelajaran mesin mengisi kekosongan tersebut [2].

Pada penelitian Ying Zhang dan Ving Zhang [4], menggunakan beberapa algoritma klasifikasi dalam data mining untuk mengetahui retensi (penyimpanan) mahasiswa menggunakan tiga algoritma dalam penelitiannya diantaranya *Navie Bayes*, *Support Vectore Mechine* dan *Decision Tree* untuk menganalisis prestasi akademik mahasiswa. Dalam penelitiannya didapat bahwa dari ketiga Algoritma Data Mining tersebut yang paling akurat adalah *Navie Bayes* dengan 89,5% dan yang kedua adalah *Support Mechine* dengan 83,5% dan yang terakhir adalah *Decision Tree* dengan 81,3% [4]. Dalam penelitian

mengatakan ketika dilakukan pengujian K Fold Cross validation sejumlah 10 Fold untuk dengan metode KNN pada data uji sejumlah 500 data, maka didapati nilai K terbaik adalah pada metode KNN adalah 5. Lebih lanjut [6] mengatakan sistem informasi data mining untuk siswa berpotensi akademik berbasis online dapat memberikan dukungan bagi orang tua dalam mengevaluasi potensi putra dan putrinya, dikarenakan kemudahan dalam melakukan akses informasi dan akurasi yang tinggi.

Panharesani dan Anugrah [7] dalam penelitian ini dilakukan klasifikasi waktu penyelesaian skripsi mahasiswa Program Studi Teknik Informatika Universitas Muhammadiyah Gresik dengan menggunakan metode Weighted Naïve Bayes. Dengan menambahkan bobot terhadap atribut kelas pada algoritma naïve bayes, maka ketepatan klasifikasi pada algoritma weighted naïve bayes tidak didasari hanya dengan probabilitas melainkan juga dipengaruhi oleh pembobotan fitur terhadap kelas. Hasil klasifikasi data latih mahasiswa teknik informatika Universitas Muhammadiyah Gresik dalam waktu penyelesaian skripsi didapatkan nilai probabilitas mahasiswa teknik informatika dalam menyelesaikan skripsi dengan waktu 1 semester, 2 semester, dan > 2 semester adalah 0.14286, 0.53968, dan 0.31746. Dan hasil pengujian evaluasi performa dengan menggunakan confusion matrix, didapatkan nilai accuracy 77,5%, precision kelas 1S 100%, precision kelas 2S 72,3%, precision kelas 3S 82,4%, recall kelas 1S 11,1%, recall kelas 2S 100%, recall kelas 3S 93,3%, specificity kelas 1S 100%, specificity kelas 2S 75%, specificity kelas 3S 88%.

Fajar Mu'Alim [9] Dalam penelitian ini dilakukan penjurusan siswa dengan cara menghitung nilai siswa menggunakan perhitungan manual. Dengan adanya kemajuan teknologi pihak Madrasah Aliyah Negeri (MAN) Sintang dapat melakukan penjurusan siswa melalui sebuah sistem klasifikasi penjurusan siswa. Klasifikasi penjurusan siswa pada penelitian ini menggunakan metode Random Forest. Metode ini digunakan untuk mendapatkan hasil voting sebagai penentuan akhir pada hasil klasifikasi penjurusan siswa. Hasil pengujian aplikasi klasifikasi penjurusan siswa di Madrasah Aliyah Negeri (MAN) Sintang menggunakan 210 data latih dan 90 data uji mendapatkan nilai akurasi pengujian sebesar 94,38%.

Jadi masalah yang dapat dirumuskan yaitu bagaimana perbandingan Algoritma Naïve Bayes dan Random Forest terhadap prestasi belajar siswa Sekolah Menengah Kejuruan (SMK) Swasta Siti Banun Sigambal, Jurusan Rekayasa Perangkat Lunak (RPL) berdasarkan variabel demografi yaitu: (suku, pembiayaan, tempat tinggal, umur, jenis kelamin) dan Nilai Rapor. Tujuan dari penelitian ini yaitu mengetahui perbandingan dan nilai akurasi dari Algoritma Naïve Bayes dan Random Forest.

## II. METODE PENELITIAN

Pada penelitian ini penulis memilih proses model CRISP-DM (Cross Industry standard process for data mining) [8] seperti pada gambar II.1. Model ini menjadi dasar teori dalam metode penelitian ini. Adapun langkah-langkah dalam penelitian ini yaitu :

1. Fase Business Understanding, data yang diperoleh dari Siswa Sekolah Menengah Kejuruan (SMK) Swasta Siti Banun, Jurusan Rekayasa Perangkat Lunak (RPL) selama ini belum pernah dilakukan pengambilan data tentang biaya sekolah terhadap data tersebut dan belum dimanfaatkan dalam menentukan prediksi prestasi akademik mahasiswa menggunakan metode Data Mining. Oleh karena itu penelitian ini akan menggali data tersebut dengan menggunakan Algoritma Data Mining.
2. Fase Data Understanding, data yang akan diambil dari Wakil Kepala pada bagian Kurikulum dan data Siswa Sekolah Menengah Kejuruan (SMK) Swasta Siti Banun, Jurusan Rekayasa Perangkat Lunak (RPL) adalah suku, sekolah asal, tempat tinggal, umur dan jenis kelamin.
3. Fase Data Preparation, Dari data Siswa Sekolah Menengah Kejuruan (SMK) Swasta Siti Banun, Jurusan Rekayasa Perangkat Lunak (RPL) maka akan dilakukan teknik data preparation agar kualitas data diperoleh lebih baik dengan cara :
  - a. Data Validation, untuk mengidentifikasi dan menghapus data yang ganjil, data yang tidak konsisten dan data yang tidak lengkap.
  - b. Data Integration and Transformation : untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penelitian ini bernilai kategorikal untuk model klasifikasi.
  - c. Data Size Reduction dan Discretization : untuk memperoleh data set dengan jumlah atribut dan record yang lebih sedikit tetapi bersifat informatif. Di dalam data training yang digunakan dalam penelitian ini, dilakukan seleksi atribut dan penghapusan data duplikasi.
4. Fase Modeling, pada tahap ini data di proses dan diklasifikasikan sehingga menghasilkan sejumlah aturan dengan menggunakan algoritma data mining seperti penelitian [7]
5. Evaluation Phase, pada fase ini dilakukan pengujian data terhadap model klasifikasi dengan algoritma data mining [11].

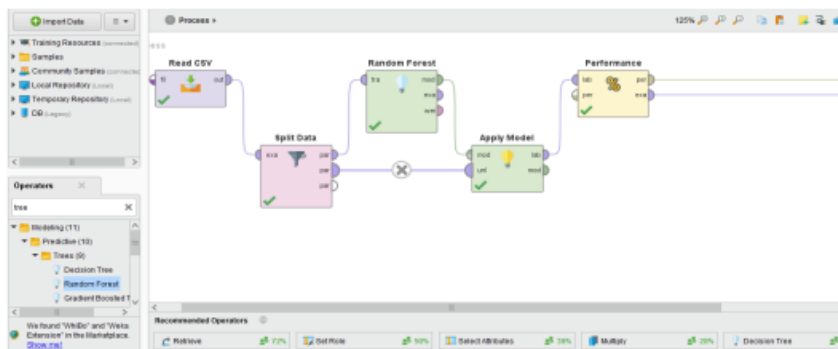
## III. HASIL DAN PEMBAHASAN

Populasi siswa jurusan Rekayasa Perangkat Lunak (RPL) Sekolah Menengah Kejuruan (SMK) Swasta Siti Banun berjumlah 180 siswa, kelas X (Sepuluh) berjumlah 36 siswa, kelas XI (Sebelas) berjumlah 72 siswa dan kelas XII (Duabelas) berjumlah 72 siswa. Dalam penelitian ini hanya mengambil data siswa pada jurusan Rekayasa Perangkat Lunak (RPL) dari kelas X, XI dan XII. Karena data yang diambil semester II tahun ajaran 2020/2021. Sampel yang digunakan dalam penelitian ini berjumlah 180 siswa dan data yang diambil adalah suku, atribut sekolah (SMP dan sederajat) umur saat masuk sekolah, nilai rapor, tempat tinggal, biaya sekolah, penghasilan orang tua.

Data menunjukkan atribut suku bangsa yang paling banyak berasal dari suku Jawa dengan jumlah 123 mahasiswa, atribut sekolah (SMP dan sederajat) yang paling banyak siswa tamatan Madrasah Tsanawiyah (MTS) dengan jumlah 166 siswa, biaya sekolah yang paling banyak membiayai adalah orangtua dengan

jumlah 180 siswa, penghasilan orangtua selama 1 bulan yang paling banyak penghasilan orangtua adalah 1-2 juta dengan jumlah 165 mahasiswa, tempat tinggal selama sekolah yang paling banyak tinggal bersama orang tua dengan jumlah 168 mahasiswa, umur saat masuk sekolah yang paling banyak adalah umur 15 tahun dengan jumlah 172 siswa, Nilai Raport dengan total nilai >80,00 dengan jumlah 139 siswa.

### 3.1 Algoritma Naive Bayes



Gambar IV.1 Implementasi Algoritma Naïve Bayes

Tabel 1. Cross Validation Naive Bayes

	true High	true Normal	true Low	class precision
pred.High	93	32	3	72.66%
pred.Normal	17	14	4	40.00%
pred.Low	1	0	0	0.00%
class recall	83.78%	30.43%	0.00%	

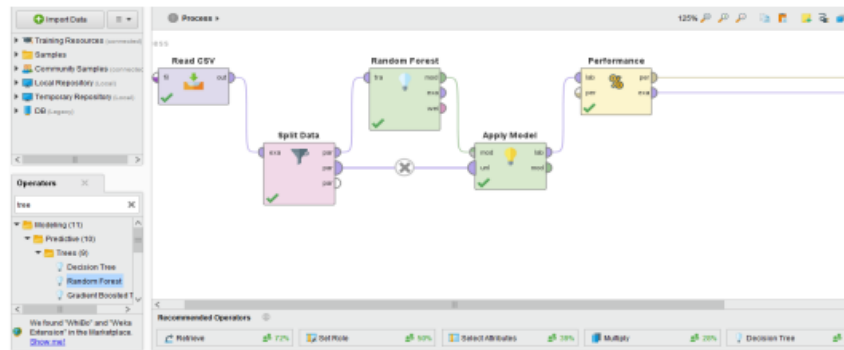
Pada tabel 1 Cross validation Naive Bayes. Dalam perhitungan tabel menggunakan confusion matrix, Class precision : pred.High 72.66%, pred.Normal 40.00% dan pred.Low 0.00%. Class recall : true High 83.78%, true Normal 30.43% dan true Low 0.00%.

Tabel 2 Independent Naive Bayes

	true High	true Normal	true Low	class precision
pred.High	91	34	4	70.54%
pred.Normal	20	10	3	30.30%
pred.Low	0	2	0	0.00%
class recall	80.98%	21.74%	0.00%	

Pada tabel III.2 Independent Naive Bayes. Dalam perhitungan tabel menggunakan confusion matrix, Class precision : pred.High 72.73%, pred.Normal 43.24% dan pred.Low 0.00%. Class recall : true High 79.28%, true Normal 34.78% dan true Low 0.00%.

### 3.2 Algoritma Random Forest



Gambar. 2 Implementasi Algoritma Random Forest

Tabel 3 Cross Validation Random Forest

	true High	true Normal	true Low	class precision
pred.High	89	32	4	71.20%
pred.Normal	21	13	3	35.14%
pred.Low	1	1	0	0.00%
class recall	80.18%	28.26%	0.00%	

Pada tabel .3 Cross validation Random Forest. Dalam perhitungan tabel menggunakan confusion matrix, Class precision : pred.High 71.20%, pred.Normal 35.14% dan pred.Low 0.00%. Class recall : true High 80.18%, true Normal 28.26% dan true Low 0.00%.

Tabel .4 Independent Random Forest

	true High	true Normal	true Low	class precision
pred.High	27	1	0	96.43%
pred.Normal	1	10	0	90.91%
pred.Low	0	0	2	100.00%
class recall	96.43%	90.91%	100.00%	

Pada tabel.4 *Independent Random Forest*. Dalam perhitungan tabel menggunakan confusion matrix, Class precision : pred.High 96.43% pred.Normal 90.91% dan pred.Low 100.00%. Class recall : true High 96.43%, true Normal 90.91% dan true Low 100.00%.

### 3.3. Perbandingan Cross Validation

Tabel III.5 Cross Validation

Cross Validadation	Accurac y	Recall	Presisi
Naive Bayes	65.24%	38.07%	37.55%
Random Forest	62.20%	36.15%	35.45%

Pada tabel.5 Cross Validattion Accuracy yang paling tinggi ialah Naive Bayes dengan bobot 65.24%, karena naive bayes memiliki data yang lebih banyak pred.High (93), pred.Normal (17) dan pred.low (1). Recall yang paling tinggi ialah Random Forest dengan bobot 95.78%. Dan Presisi yang paling tinggi ialah Random Forest dengan bobot 95.78%. Random forest memiliki acuracy 62.20%, pred.high (89), pred.normal (21), dan pred.low (1).

### 3.4. Perbandingan Independent

Tabel III.6 Independent

Independent	Accuracy	Recall	Presisi
Naive Bayes	63.41%	38.02 %	38.66 %
Random Forest	95.12%	95.78 %	95.78 %

Pada tabel III.6 Independent Accuracy yang paling tinggi adalah Random Forest dengan bobot 95,12%, karena memiliki data yang lebih sedikit pred.high (27), pred.normal (1), dan (0). Recall yang paling tinggi adalah Random Forest dengan bobot 95.78%. Dan Presisi yang paling tinggi adalah Random Forest dengan bobot 95.78%. Naive bayes memiliki acuracy yang lebih rendah 63.41%, pred.high (88), pred.normal (20) dan pred.low (3).

## IV. Kesimpulan

Kesimpulan yang dapat diambil dari penelitian tentang perbandingan algoritma Naïve Bayes dan Algoritma Random Forest terhadap prestasi belajar siswa Sekolah Menengah Kejuruan (SMK) Swasta Siti Banun, jurusan Rekayasa Perangkat Lunak (RPL) ditinjau dari aspek Kurikulum dan ekonomi adalah sebagai berikut :

1. Penelitian ini menggunakan 10-fold cross validation untuk memprediksi tingkat error data data. Dataset dibagi dua yaitu data training dan data testing. Cross validation dengan accuracy yang paling tinggi adalah naive bayes 65.24%, random forest 62.20%.
2. Independent dengan accuracy yang paling tinggi adalah random forest 95.12%, naive bayes 63.41%
3. Dalam penelitian ini ada tujuh atribut yang digunakan dalam penelitian ini, ada tiga atribut dengan bobot yang paling tinggi adalah sekolah (SMP dan sederajat) : 0.209, tempat tinggal : 0.200, dan jenis kelamin : 0.199.
4. Dalam penelitian ini menggunakan dua algoritma : naive bayes dan random forest. Algoritma dengan bobot yang paling tinggi adalah random forest dengan accuracy 95.12%.
- 5.

## V. Daftar Pustaka

- [1] Helal, S., Li, J., Liu, L., Ebrahimie, E., Dawson, S., Murray, D.J., dan Long, Q. 2017. Predicting academic performance by considering student heterogeneity. Knowledge-Based Syst., 161 (Dec. 2017), 134–146. DOI=<http://doi.acm.org/10.1016/j.knosys.2018.07.042>.



- [2] Rajagukguk, S.A. (2021) 'Tinjauan Pustaka Sistematis: Prediksi Prestasi Belajar Peserta Didik Dengan Algoritma Pembelajaran Mesin', *Jurnal SNATi*, 1.
- [3] Dimiyati dan Mudjiyono. 2009. *Belajar dan Pembelajaran*. Rineka Cipta, Jakarta.
- [4] Y. Zhang, S. Oussena, T. Clark and H. Kim, "Use Data Mining To Improve Student Retention In Higher Education – A Case Study," 2014.
- [5] M. Khlil., Kusriani., Henderei. Penerapan Metode K Nearest Neighbord Dalam Proses Seleksi Penerima Beasiswa, "Seminar Nasional Sistem Informasi dan Teknologi Informasi", p. 13-18, 2018
- [6] D. Setiyadi., A. Nurdin. "Data Mining Potensi Akademik Siswa Berbasis Online". *Jurnal Sisfotenika*. ISSN: 2460-5344, Vol. 2, No. 1, 2012
- [7] Panharsi, G.A. and Anugrah, I.G. (2022) 'Klasifikasi Waktu Penyelesaian Skripsi Mahasiswa Menggunakan Metode Weighted Naive Bayes', *INDEXIA: Informatic and Computational Intelligent Journal*, 4(1), pp. 33-46.
- [8] D. Feblian and D. U. Daihani, "Implementasi Model Crisp-Dm Untuk Menentukan Sales Pipeline Pada Pt X," *Jurnal Teknik Industri* ISSN: 1411-6340, p. 5, 2015.
- [9] Fajar Mu'Alim, R.H. (2022) 'Implementasi Metode Random Forest Untuk Penjurusan', *Jurnal JUPITER*, 14(1), pp. 116-125.
- [10] D. Sofi, "Perbandingan: Prediksi Prestasi Belajar Mahasiswa Menggunakan Teknik Data Mining (Study Kasus Fasilkom Unsika)," *Konferensi Nasional Sistem Informasi 2014*, p. 118, 2014.
- [11] A. R. Khadafy, "Penerapan Naive Bayes untuk Mengurangi Data Noise pada Klasifikasi Multi Kelas dengan Decision Tree," *Konferensi Nasional Sistem Informasi 2014*, vol. 2, p. 136, 2015.
- [12] Rusdianto, D.dkk.2020. 'Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Mengetahui Pola Peminjaman Buku Di Perpustakaan Universitas Jurnal Sistem Informasi, J-SIKA Volume 02 Nomor 02, Desember 202', *Jurnal Sistem Informasi*, 02(02), pp. 1-10.
- [13] Takdirillah, R. (2020) 'Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Sebagai Pendukung Informasi Strategi Penjualan', *Edumatic: Jurnal Pendidikan Informatika*, 4(1), pp. 37-46. Available at: <https://doi.org/10.29408/edumatic.v4i1.2081>
- [14] Yunita, D., Rosyani, P. and Amalia, R. (2018) 'Analisa Prestasi Siswa Berdasarkan Kedisiplinan, Nilai Hasil Belajar, Sosial Ekonomi dan Aktivitas Organisasi Menggunakan Algoritma Naïve Bayes', *Jurnal Informatika Universitas Pamulang*, 3(4), p. 209. Available at: <https://doi.org/10.32493/informatika.v3i4.2032>.



- [15] J. Han, M. Kamber, and J. Pei, Data mining: Concepts and Techniques (3rd ed.). Morgan Kaufmann, 2012
- [16] Primajaya, A. and Sari, B.N. (2018) 'Random Forest Algorithm for Prediction of Precipitation', Indonesian Journal of Artificial Intelligence and Data Mining, 1(1), p. 27. Available at: <https://doi.org/10.24014/ijaidm.viii.4903>.
- [17] C. Ferri, J. Hernández-Orallo, R. Modroiu. "An Experimental Comparison of Performance Measures for Classification". Pattern Recognitipns Letters. 2009; 30(1): 27-38
- [18] Larose, Daniel T, Data Mining Methods and Models. Hoboken New Jersey : Jhon Wiley & Sons, Inc, 2006.