

Implementasi Algoritma Naïve Bayes Dalam Prediksi Kemungkinan Mahasiswa Mengikuti Perkuliahan

Faiza Renaldi

Jurusan Informatika, Fakultas Sains and Informatika
Universitas Jenderal Achmad Yani
faiza.renaldi@unjani.ac.id

Irma Santikarama

Jurusan Informatika, Fakultas Sains and Informatika
Universitas Jenderal Achmad Yani

Afmi Rurifando

Jurusan Informatika, Fakultas Sains and Informatika
Universitas Jenderal Achmad Yani

Imam Rofiki

Jurusan Tadris Matematika
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Abstrak— Penilaian kinerja mahasiswa dinilai dari beberapa element dan factor seperti kehadiran, tugas, kuis, dan ujian akhir. Factor yang mendasar dalam penilaian ialah kehadiran dimana mahasiswa sering kali menyepelekan terhadap masalah kehadiran yang menyebabkan tidak lulus dalam salah satu matakuliah. Untuk memprediksi mahasiswa dalam mengikuti kegiatan perkuliahan, diperlukan analisa yang cermat dengan mempertimbangkan banyak kriteria dan faktor. Kriteria yang dijadikan patokan dalam memprediksi mahasiswa antara lain waktu tempuh, waktu kegiatan pembelajaran, riwayat kegagalan kegiatan perkuliahan, waktu luang mahasiswa setelah kegiatan perkuliahan, dan juga kesehatan mahasiswa. Untuk mengolah semua kriteria tersebut diperlukan suatu sistem yang dapat memilih dan mengklasifikasikan kriteria untuk mahasiswa dalam pengambilan keputusan untuk mengikuti kegiatan perkuliahan. Algoritma Naive Bayes adalah pengklasifikasi kemungkinan sederhana yang menghitung sekumpulan peluang dengan menjumlahkan frekuensi dan kombinasi nilai dari kumpulan data yang diberikan. Penerapan metode Naive Bayes diharapkan mampu memprediksi kegiatan mahasiswa mengikuti perkuliahan. Langkah awal dalam penelitian ini ialah dengan melakukan Data Collection, Pre-Process, Pengolahan Data, Implementasi Naive Bayes dan Evaluasi. Dari 500 data riwayat kegiatan mahasiswa yang digunakan dalam pengujian dengan metode Naive Bayes, maka diperoleh persentase 66% untuk akurasi prediksi, dimana dari 100 data kegiatan perkuliahan mahasiswa yang diuji terdapat 66 data kegiatan pembelajaran mahasiswa yang berhasil diklasifikasikan dengan benar.

Kata kunci— Naive Bayes; Klasifikasi; Kegiatan Perkuliahan;

I. PENDAHULUAN

Algoritma Naive Bayes menggunakan teorema Bayes dan mengasumsikan semua atribut asumsi independensi antar prediktor yang diberikan oleh nilai dalam variabel kelas [1][2]. Contohnya, kita bisa menggunakan classification untuk meramalkan apakah cuaca pada hari tertentu akan ‘cerah’, ‘hujan’ atau ‘berawan’. Algoritma Naive Bayes sebagian besar digunakan dalam analisis sentimen, penyaringan spam, sistem rekomendasi, dll. Hanya membutuhkan sedikit data latih untuk menentukan estimasi parameter yang dibutuhkan dalam proses klasifikasi. Pada metode Naive Bayes data String konstan

dibedakan dengan data numerik kontinu, perbedaan ini akan terlihat pada saat menentukan nilai probabilitas setiap kriteria apakah kriteria dengan nilai data string atau kriteria dengan nilai data numerik [3]. (Bagus et al, 2017) menunjukkan bahwa Naive Bayes mampu mengklasifikasikan data testing dengan benar. Penentuan data latih dan data pengujian dapat mempengaruhi hasil pengujian, karena pola data latih akan digunakan sebagai aturan untuk menentukan kelas pada data pengujian [4]. Dengan algoritma Naive Bayes dapat mempermudah dalam pengambilan keputusan dan klasifikasi siswa SMA jurusan non-Linguistik dapat bersekolah di Fakultas Sastra, meskipun hal ini diimbangi dengan bakat dan juga nilai ujian nasional. beberapa data siswa Penelitian ini menegaskan bahwa algoritma Naive Bayes memiliki akurasi yang lebih baik dalam mengevaluasi kinerja akademik siswa. Dengan kata lain, algoritma Naive Bayes dapat digunakan sebagai model alternatif dalam kinerja evaluasi akademik mahasiswa [6]. Dengan menggunakan Algoritma Naive Bayes dan Algoritma Decision Tree, penelitian ini menemukan faktor yang paling berpengaruh dalam kelulusan yaitu indeks prestasi kumulatif pada semester ketiga dan keempat serta indeks kinerja semester mahasiswa pada semester pertama. Pola kelulusan dapat diketahui dari awal siswa masuk sekolah hingga tahun kedua [7]. Dalam penelitian ini dijelaskan metode klasifikasi yang didasarkan pada algoritma naive bayes dan digunakan untuk data mining Akademik. Hal ini sangat mendukung bagi mahasiswa dan dosen untuk evaluasi kinerja akademik. Yang merupakan suatu metode peringatan bagi siswa untuk meningkatkan kinerja belajar mereka [8].

Pada dunia pendidikan di Indonesia terdapat hal penting dalam evaluasi yaitu hasil dari pembelajaran yang dilakukan oleh siswa yang mana merupakan tujuan semua sekolah dan institusi pendidikan. Kinerja akademik dalam komputasi mahasiswa selama tahun pertama di universitas adalah berputar sudut di jalur akademik sarjana dan biasanya mengganggu Indexs Prestasi Komulatif (IPK) dalam metode penting. Penilaian elemen mahasiswa seperti tugas, kuis, ujian tengah semester dan akhir, praktikum dipelajari. Factor kehadiran

merupakan salah satu penilaian penting yang dapat mempengaruhi penilaian IPK mahasiswa. Dimana ketika element penilaian mahasiswa menunjukkan hasil yang baik tetapi factor kehadirann sangat buruk maka akan menyebabkan tidak lulus dalam salah satu mata kuliah dan terburuknya akan menyebabkan tidak lulusnya mahasiswa tersebut tepat waktu. menganalisis kinerja mahasiswa sarjana dengan bantuan metodologi data mining. Dalam penelitian ini, perhatian utama adalah pada dua faktor signifikan dari kinerja siswa. Keberhasilan pendidikan siswa yang dirancang untuk membantu siswa yang memiliki IPK minimum diakui oleh struktur aplikasi ini. Algoritma yang diusulkan menunjukkan tingkat akurasi sekitar 53% [9]. Prediksi prestasi akademik mahasiswa sangat penting bagi institusi pendidikan. Dalam penelitian ini, potensi siswa, kekurangan dan hobi mereka diakui. Prestasi mahasiswa tersebut didasarkan pada beberapa faktor. Beberapa pengklasifikasi seperti Naïve Bayes, MLP dan pohon keputusan J48 digunakan untuk tujuan klasifikasi [10].

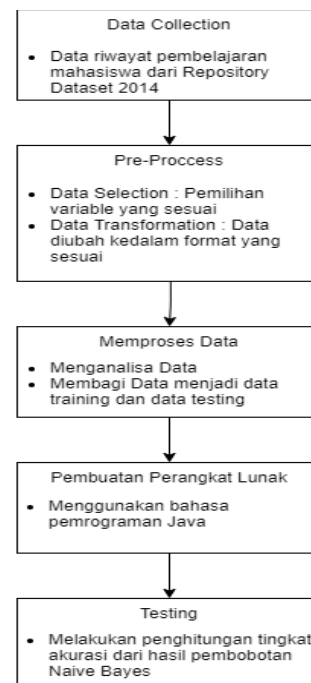
Oleh karena itu penelitian ini difokuskan pada penerapan algoritma Naïve Bayes dalam membantu mahasiswa memprediksi dan mengidentifikasi mengikuti kegiatan pembelajaran secara dini yang rentan tidak lulus pada mata kuliah karena terkait masalah kehadiran, sehingga kebijakan tertentu dapat dilakukan oleh institute. Dalam melakukan klasifikasi menggunakan algoritma Naïve Bayes atribut mahasiswa yang akan diolah menggunakan algoritma naïve bayes berupa jenis kelamin, usia mahasiswa, status tinggal mahasiswa, waktu tempuh yang diperlukan dari rumah ke institute, jumlah kegagalan dalam pelaksanaan perkuliahan, waktu luang mahasiswa setelah melakukan perkuliahan, keadaan kesehatan yang dialami oleh mahasiswa, dan juga jumlah ketidak hadiran mahasiswa dalam perkuliahan tersebut.

II. METODE

Tahapan yang dilakukan dalam penelitian ini untuk penggunaan metode algoritma Naïve Bayes dan tahapannya dapat dilihat pada Gambar 1.

A. Data Collection

Data didapatkan dari repository dataset terbuka (<http://archive.ics.uci.edu>) yang merupakan data umpan balik mahasiswa yang dinilai dari beberapa faktor internal dan eksternal seperti faktor waktu perjalanan dari rumah, lamanya kegiatan pekuliahan, kesehatan dan waktu luang mahasiswa, selain itu juga data dari sekolah (tingkat kehadiran) yang mana factor faktor itu berpengaruh pada hasil belajar. Data Latih dianalisis dengan algoritma klasifikasi dan tahap ini adalah disebut sebagai tahap pelatihan. Di sisi lain, data uji digunakan untuk memperkirakan keakuratan klasifikasi. Dengan cara ini memperkirakan kelas dari rekor yang baru diperoleh [11].



Gambar 1. Metode pada algoritma Naïve Bayes untuk memprediksi kemungkinan mahasiswa kuliah

B. Pre processing

Setelah data didapatkan kemudian dilakukan proses analisa terhadap data yang ada Pada penelitian ini data yang digunakan mempunyai 500 record data dan juga terdiri dari 12 variable [12]. Faktor seperti pendapatan keluarga, pekerjaan orang tua dan faktor konsumsi alkohol dihilangkan karena bersifat pribadi dan menjadikan data tidak konsisten. Pada Tabel 1 merupakan atribut data yang dinilai dalam Implementasi Algoritma Naïve Bayes Untuk Memprediksi Kemungkinan Mahasiswa Kuliah.

TABEL 1. ATRIBUT DATA YANG DINILAI

Atribut	Nilai
Jenis Kelamin	Jenis kelamin mahasiswa (Laki Laki, Perempuan)
Usia	Usia Mahasiswa (15 sampai 22 tahun)
Status Tinggal	Status tempat tinggal mahasiswa (Bersama orang tua atau Sendiri)
Internet	Memiliki layanan internet dirumah atau tempat tinggal sekarang (Iya, Tidak)
Kegiatan	Kegiatan Organisasi (Iya, Tidak)
Waktu Tempuh	Waktu tempuh yang dibutuhkan dari tempat tinggal ke sekolah (10 menit, 20 menit, 30 menit, atau >30 menit)
Waktu Perkuliahan	Waktu matakuliah dalam satu hari (2 jam, 2 jam, 6 jam, atau 8 jam)
Kegagalan	Jumlah kegagalan perkuliahan sebelumnya (tidak pernah, 1 kali, 2 kali, 3 kali dan 4 kali)
Waktu Luang	waktu luang setelah perkuliahan (sangat rendah, rendah, normal, tinggi dan sangat tinggi)
Kesehatan	status kesehatan saat ini (sangat buruk, buruk, normal, baik dan sangat baik)
Tidak hadir	Jumlah tidak hadir dalam perkuliahan (tidak pernah, 1 kali, 2 kali dan 3 kali)
kehadiran	Apakah mahasiswa akan mengikuti perkuliahan (Iya atau Tidak)

C. Memproses Data

Setelah data terkumpul dilakukan analisis data untuk menyesuaikan proses data yang akan diolah dengan metode Naïve Bayes. Kemudian data akan dibagi menjadi dua bagian berdasarkan ukuran standar pembagian datasets yaitu 80% diantaranya akan digunakan untuk data training dan 20% untuk data testing. Yang artinya dari 500 data yang dimiliki 400 data akan digunakan menjadi data training dan 100 data akan dijadikan data uji.

D. Implementasi Naïve Bayes

Untuk menunjang dalam implementasi algoritma Naïve Bayes, penulis menggunakan bahasa pemrograman java, dimana bahasa ini sangat cocok untuk mengimplementasikan algoritma Naïve bayes, dimana algoritma ini sangat tergantung dari rumus rumus matematis mulai dari perhitungan, penjumlahan, pengurangan dan pembagian. Penggunaan bahasa pemrograman java di dunia nyata kebanyakan digunakan untuk menulis rumus rumus matematis dalam sebuah kode. (Enaro et al) menganalisis frekuensi kode rumus bahwa satu dari 700 baris kode dalam sampel mengimplementasikan rumus penjumlahan [13]. Untuk contoh proyek komputasi ilmiah, peneliti menemukan bahwa satu dari 100 baris kode menerapkan rumus penjumlahan atau perkalian.

III. HASIL DAN DISKUSI

Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut tidak bergantung secara kondisional jika diberi nilai keluaran. Pada metode Naive Bayes data String konstan dibedakan dengan data numerik kontinu, perbedaan ini akan terlihat saat menentukan nilai probabilitas dari setiap kriteria apakah itu kriteria dengan nilai data string atau kriteria dengan nilai data numeric [14][3][15]. Untuk menentukan data yang akan dianalisis dengan metode Naive Bayes maka langkah pertama yang dilakukan adalah membaca data latih.

A. Probabilitas Waktu Tempuh

Berdasarkan data mahasiswa berdasarkan data latih sebanyak 500 data terhadap waktu tempuh yang harus dilalui dari tempat tinggal menuju institute dengan waktu tempuh selama 10 menit memiliki kemungkinan mahasiswa tersebut akan mengikuti perkuliahan sangat besar dengan tingkat kemungkinan 0,45 untuk iya dan 0,34 untuk tidak, kemudian untuk waktu tempuh selama 20 menit tingkat kemungkinan mahasiswa tersebut iya sebesar 0,24 dan 0,21 untuk tidak, dan untuk waktu tempuh selama 30 menit sampai lebih dari 40 menit memiliki tingkat kemungkinan mahasiswa mengikuti perkuliahan sangat rendah sebesar 0,15 dan 0,16 untuk iya dan 0,23 dan 0,21 untuk tidak sesuai dengan **Error! Reference source not found.**

TABEL 2. PELUANG TERHADAP WAKTU TEMPUH MAHASISWA

Waktu Tempuh	Jumlah peluang Mahasiswa mengikuti kegiatan perkuliahan	
	Iya	Tidak
10 menit	145	60
20 menit	78	40
30 menit	47	40

40menit	50	40
jumlah	320	180

B. Probabilitas Waktu Perkuliahan

Berdasarkan data mahasiswa berdasarkan data latih sebanyak 500 data terhadap waktu perkuliahan dalam sehari yang harus diikuti oleh mahasiswa sangat berpengaruh dalam peluang mahasiswa mengikuti perkuliahan. Dimana peluang terbesar mahasiswa mengikuti perkuliahan pada waktu 4 jam dalam sehari dengan tingkat kemungkinan sebesar 0,43 untuk iya dan 0,40 untuk tidak, kemudian untuk waktu perkuliahan selama 2 jam dengan tingkat peluang sebesar 0,27 untuk iya dan 0,35 untuk tidak, selanjutnya untuk waktu perkuliahan selama 6 dan 8 jam dalam sehari memiliki peluang sebesar 0,18 dan 0,12 untuk iya dan 0,09 dan 0,17 untuk tidak, sesuai dengan **Error! Reference source not found.**

TABEL 3. PELUANG TERHADAP WAKTU PERKULIAHAN DALAM 1 HARI

Waktu Perkuliahan	Peluang Mahasiswa mengikuti kegiatan perkuliahan	
	Ya	Tidak
2 jam	87	60
4 jam	138	73
6 jam	57	17
8 jam	38	30
jumlah	320	180

C. Probabilitas Kegagalan Perkuliahan

Berdasarkan data mahasiswa berdasarkan data latih sebanyak 500 data terhadap kegagalan pelaksanaan perkuliahan sebelumnya memiliki pengaruh kecil terhadap ketidak ikutan mahasiswa dalam mengikuti kegiatan perkuliahan, dimana ketika tidak pernah terjadinya kegagalan perkuliahan sebelumnya tingkat kemungkinan mahasiswa mengikuti perkuliahan sebesar 0,20 untuk iya dan 0,16 untuk tidak, diikuti dengan 1 kali kegagalan perkuliahan sebelumnya memiliki peluang mahasiswa mengikuti perkuliahan sebesar 0,21 untuk iya dan 0,20 untuk tidak. selanjutnya terjadinya kegagalan perkuliahan sebanyak 2 dan 3 kali memiliki tingkat kemungkinan mahasiswa mengikuti perkuliahan sebesar 0,26 dan 0,34 untuk iya dan 0,26 dan 0,38 untuk tidak, sesuai dengan **Error! Reference source not found.**

TABEL 4. PELUANG TERHADAP KEGAGALAN PERKULIAHAN

Kegagalan perkuliahan	Jumlah peluang Mahasiswa mengikuti kegiatan perkuliahan	
	Ya	Tidak
Tidak Pernah	62	29
1 kali	65	38
2 kali	80	46
3 kali	109	71
jumlah	316	184

D. Probabilitas Waktu Luang

Berdasarkan data mahasiswa berdasarkan data latih sebanyak 500 data terhadap waktu luang setelah kegiatan perkuliahan sangat tinggi pada keadaan normal dengan tingkat

kemungkinan mahasiswa tersebut mengikuti perkuliahan sebesar 0,40 untuk iya dan 0,41 untuk tidak. kemudian diikuti waktu luang tinggi sebesar 0,29 untuk iya dan 0,26 untuk tidak. Kemudian untuk waktu luang rendah dengan tingkat kemungkinan sebesar 0,15 untuk iya dan 0,18 untuk tidak diikuti oleh waktu luang sangat tinggi dan sangat rendah dengan masing masing kemungkinan sebesar 0,11 dan 0,05 untuk iya dan 0,09 dan 0,06 untuk tidak, seperti pada **Error! Reference source not found.**

```
Mencari Kemungkinan? (y/n)
y
Lamanya Waktu Tempuh :40 menit
Lamanya Waktu Perkuliahan :4 jam
Jumlah Kegagalan Perkuliahan :1 kali
Keadaan Waktu Luang :tinggi
Kesehatan :baik

Hasil Perhitungan Probabilitas
Kemungkinan Mahasiswa Kuliah : 61.15%
Kemungkinan Mahasiswa Tidak Kuliah : 38.85%
```

Gambar 2. Hasil dari data uji pertama

Pada Gambar 2 terlihat untuk data pertama memiliki kemungkinan mahasiswa tersebut mengikuti perkuliahan sebesar 61% terlihat dan kemungkinan tidak mengikuti perkuliahan sebesar 38%.

TABEL 5. PELUANG TERHADAP WAKTU LUANG SETELAH PERKULIAHAN

Waktu Luang	Jumlah peluang Mahasiswa mengikuti kegiatan perkuliahan	
	Ya	Tidak
Sangat Rendah	16	11
Rendah	48	32
Normal	126	72
Tinggi	94	42
Sangat Tinggi	34	17
Total	318	182

E. Probabilitas Kesehatan

Berdasarkan data mahasiswa berdasarkan data latih sebanyak 500 data terhadap kesehatan yang sedang dialami oleh mahasiswa sangat berpengaruh besar dalam menilai mahasiswa tersebut akan mengikuti perkuliahan. Dengan keadaan mahasiswa tersebut sangat baik maka kemungkinan untuk mengikuti perkuliahan sangat tinggi dengan tingkat kemungkinan sebesar 0,39 untuk iya dan 0,40 untuk tidak. sedangkan untuk kesehatan mulai dari baik hingga sangat buruk tingkat kemungkinan untuk mahasiswa mengikuti pelajaran sangat kecil. Dimana kesehatan baik memiliki tingkat kemungkinan sebesar 0,18 untuk iya dan 0,13 untuk tidak, kemudian keadaan normal sebesar 0,17 untuk iya dan 0,24 untuk tidak dilanjutkan dengan keadaan buruk dan sangat buruk sebesar 0,12 dan 0,14 untuk iya dan 0,10 dan 0,13 untuk tidak. sesuai dengan **Error! Reference source not found.**

```
Mencari Kemungkinan? (y/n)
y
Lamanya Waktu Tempuh :20 menit
Lamanya Waktu Perkuliahan :6 jam
Jumlah Kegagalan Perkuliahan :tidak pernah
Keadaan Waktu Luang :rendah
Kesehatan :buruk

Hasil Perhitungan Probabilitas
Kemungkinan Mahasiswa Kuliah : 64.75%
Kemungkinan Mahasiswa Tidak Kuliah : 35.25%
```

Gambar 3. Hasil dari data uji kedua

Pada Gambar 3 terlihat untuk data kedua memiliki kemungkinan mahasiswa tersebut mengikuti perkuliahan sebesar 64% dan tidak mengikuti perkuliahan sebesar 35%.

TABEL 6. PELUANG TERHADAP KESEHATAN MAHASISWA

Kesehatan	Jumlah peluang Mahasiswa mengikuti kegiatan perkuliahan	
	Ya	Tidak
Sangat Buruk	41	41
Buruk	37	36
Normal	55	56
Baik	60	56
Sangat Baik	124	74
Total	319	181

F. Pengujian pemrograman java

Dari hasil percobaan dengan menggunakan data uji pada pemrograman java seperti pada **Error! Reference source not found.** dan **Error! Reference source not found.**

IV. KESIMPULAN

Dalam tujuan mengevaluasi penulis melakukan eksperimen dengan menggunakan tool WEKA. Di bawah "Opsipengujian", validasi silang 10 kali lipat dipilih sebagai proses evaluasi [16]. Berdasarkan data mahasiswa yang digunakan sebagai data latih, metode Naive Bayes berhasil mengklasifikasikan data yang di uji dengan tingkat keakurasian berada pada kisaran 65% Tingkat akurasi yang dapat dipengaruhi oleh factor-faktor non-teknis (suasana hati, romantic, kegiatan organisasi, kompetisi) sehingga menyebabkan tingkat akurasi yang dihasilkan sangat kecil dari yang diharapkan. Perangkat lunak yang dibangun pada penelitian ini dapat dijadikan tumpuan dalam mengimplementasikan metode Naive Bayes dengan menggunakan bahasa pemrograman java. Juga program ini dapat dikembangkan kembali dengan menggunakan pemrograman java berbasis GUI dan dengan mengkombinasikan metode Naive Bayes dengan metode lainnya untuk menghasilkan tingkat akurasi yang lebih tinggi lagi.

DAFTAR PUSTAKA

- [1] S. Veni and A. Srinivasan, "Defect classification using naïve bayes classification," *Int. J. Appl. Eng. Res.*, vol. 12, no. 22, pp. 12693–12700, 2017.
- [2] H. Zhang, L. Jiang, and L. Yu, "Attribute and instance weighted naive Bayes," *Pattern Recognit.*, vol. 111, 2021, doi: 10.1016/j.patcog.2020.107674.

- [3] S. Veni and A. Srinivasan, "Efficient analysis of defects using naïve bayes classification and impact on effort and cost of the project," *Int. J. Appl. Eng. Res.*, vol. 12, no. 21, pp. 10546–10550, 2017.
- [4] I. Bagus, A. Peling, I. N. Arnawan, I. P. A. Arthawan, and I. G. N. Janardana, "Implementation of Data Mining To Predict Period of Students Study Using Naive Bayes Algorithm," vol. 2, no. 1, pp. 53–57, 2017.
- [5] U. Pujiyanto, E. N. Azizah, and A. S. Damayanti, "Naive Bayes Using to Predict Students ' Academic Performance at Faculty of Literature," pp. 163–169, 2017.
- [6] N. Dengen, E. Budiman, M. Wati, and U. Hairah, "Student Academic Evaluation using Naive Bayes Classifier Algorithm," *2018 2nd East Indones. Conf. Comput. Inf. Technol.*, pp. 104–107, 2018.
- [7] A. Wibowo, D. Manongga, and H. D. Purnomo, "The Utilization of Naive Bayes and C . 45 in Predicting the Timeliness of Students ' Graduation," vol. 7, no. 1, pp. 99–112, 2020.
- [8] F. Razaque *et al.*, "Using naïve bayes algorithm to students' bachelor academic performances analysis," *4th IEEE Int. Conf. Eng. Technol. Appl. Sci. ICETAS 2017*, vol. 2018-Janua, pp. 1–5, 2018, doi: 10.1109/ICETAS.2017.8277884.
- [9] R. Asif, A. Merceron, S. A. Ali, and N. G. Haider, "Analyzing undergraduate students' performance using educational data mining," *Comput. Educ.*, vol. 113, 2017, doi: 10.1016/j.compedu.2017.05.007.
- [10] S. Roy and A. Garg, "Predicting academic performance of student using classification techniques," in *2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics, UPCON 2017*, 2017, vol. 2018-January, doi: 10.1109/UPCON.2017.8251112.
- [11] M. Bowles, "Machine Learning in Python: Essential Techniques for Predictive Analysis," *CEUR Workshop Proceedings*, vol. 1542, pp. 33–36, 2015.
- [12] A. O. Enaro and S. Chakraborty, "Feature selection algorithms for predicting students academic performance using data mining techniques," *Int. J. Sci. Technol. Res.*, vol. 9, no. 4, 2020.
- [13] O. Moseler, F. Lemmer, S. Baltes, and S. Diehl, "On the diversity and frequency of code related to mathematical formulas in real-world Java projects," *J. Syst. Softw.*, vol. 172, p. 110863, 2021, doi: 10.1016/j.jss.2020.110863.
- [14] M. Karim and R. M. Rahman, "Decision Tree and Naive Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing," *J. Softw. Eng. Appl.*, vol. 06, no. 04, pp. 196–206, 2013, doi: 10.4236/jsea.2013.64025.
- [15] F. Paquin, J. Rivnay, A. Salleo, N. Stingelin, and C. Silva, "Multi-phase semicrystalline microstructures drive exciton dissociation in neat plastic semiconductors," *J. Mater. Chem. C*, vol. 3, pp. 10715–10722, 2015, doi: 10.1039/b000000x.
- [16] C. Anuradha and T. Velmurugan, "Feature selection techniques to analyse student academic performance using naïve bayes classifier," *3rd Int. Conf. Small Mediu. Bus.*, pp. 345–350, 2016.