

Turnitin LLCC

2213020242_APLIANA SRIYANTI BILI.docx

 TI-09

Document Details

Submission ID**trn:oid:::2945:394252072****Submission Date****Jun 19, 2026, 4:07 PM GMT+7****Download Date****Jun 19, 2026, 4:09 PM GMT+7****File Name****2213020242_APLIANA SRIYANTI BILI.docx.pdf****File Size****1.8 MB****80 Pages****13,029 Words****82,200 Characters**

29% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography

Top Sources

- 15%  Internet sources
- 14%  Publications
- 27%  Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Top Sources

- 15% Internet sources
- 14% Publications
- 27% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	repository.dinamika.ac.id	<1%
2	Internet	repository.usd.ac.id	<1%
3	Student papers	Universitas Islam Sumatera Utara on 2026-04-20	<1%
4	Student papers	Universitas Putera Batam on 2026-02-03	<1%
5	Student papers	UIN Sunan Gunung Djati Bandung on 2017-12-07	<1%
6	Internet	kc.umn.ac.id	<1%
7	Student papers	Universitas Islam Indonesia on 2026-06-15	<1%
8	Internet	core.ac.uk	<1%
9	Internet	repository.ub.ac.id	<1%
10	Student papers	Boston University on 2026-05-13	<1%
11	Internet	docplayer.info	<1%

12	Internet	repository.umsu.ac.id	<1%
13	Student papers	Universitas Muhammadiyah Sukabumi on 2026-01-13	<1%
14	Student papers	Fakultas Teknik on 2025-08-21	<1%
15	Student papers	Universitas Sanata Dharma on 2026-06-15	<1%
16	Internet	jurnal.dharmawangsa.ac.id	<1%
17	Internet	repository.ulb.ac.id	<1%
18	Student papers	Sriwijaya University on 2026-05-04	<1%
19	Student papers	Universitas Khairun on 2026-05-25	<1%
20	Publication	Achmad Hakim Qoirul Haq, Harminto Mulyo, Adi Sucipto. "Optimasi Kinerja Algor...	<1%
21	Publication	Riska Aini Putri, Muhammad Abdul Hadi, Fuad Maulana, Dzaky Akram, Nathasa S...	<1%
22	Student papers	Universitas Sangga Buana YPKP on 2026-02-21	<1%
23	Publication	Zidane Muhammad Azmi, Alvino Octaviano. "Prediksi Pemahaman Matematika Si...	<1%
24	Publication	Erika Adelia Pratiwi, Sita Muharni. "PENERAPAN ALGORITMA RANDOM FOREST U...	<1%
25	Student papers	UPN Veteran Jakarta on 2025-12-19	<1%

26	Student papers	Udayana University on 2026-01-08	<1%
27	Student papers	Universitas Budi Luhur on 2025-08-02	<1%
28	Student papers	Universitas Islam Negeri Raden Fatah on 2026-03-11	<1%
29	Internet	etheses.uin-malang.ac.id	<1%
30	Student papers	Telkom University on 2026-05-04	<1%
31	Student papers	Universitas Budi Luhur on 2026-01-31	<1%
32	Student papers	Universitas Sumatera Utara on 2026-05-18	<1%
33	Internet	123dok.com	<1%
34	Student papers	Universitas Brawijaya on 2017-02-23	<1%
35	Student papers	Universitas Islam Indonesia on 2026-05-12	<1%
36	Student papers	Universitas Terbuka on 2025-07-23	<1%
37	Internet	jurnal.polinema.ac.id	<1%
38	Publication	Evania Priyanto, Rina Firliana, M. Najibullah Muzaki. "SISTEM INFORMASI AKADE..."	<1%
39	Publication	Mhd Wedo, Gunadi Widi Nurcahyo, Rini Sovia. "Implementasi Metode Profile Mat..."	<1%

40	Student papers	Universitas Dinamika on 2026-06-11	<1%
41	Student papers	Fakultas Teknik on 2025-10-27	<1%
42	Student papers	Fakultas Teknologi Industri on 2026-05-15	<1%
43	Publication	Fredyo Eltanin Lumban Raja, Daniel Purba, Syifa Nur Rakhmah, Findi Ayu Sariasih...	<1%
44	Student papers	Universitas Negeri Surabaya on 2026-06-19	<1%
45	Student papers	Telkom University on 2026-01-21	<1%
46	Student papers	Telkom University on 2026-04-27	<1%
47	Student papers	Universitas Pendidikan Indonesia on 2025-08-11	<1%
48	Publication	Alvina Jelita Firdaus, Zaehol Fatah. "Disease Risk Prediction Using the Random Fo...	<1%
49	Publication	Rini Ariza, Muhamad Fatchan, Asep Suprianto. "Analisis Sentimen Opini Publik Te...	<1%
50	Student papers	Universitas Muhammadiyah Sukabumi on 2026-01-20	<1%
51	Internet	repository.umb.ac.id	<1%
52	Publication	Sonia Sarah Sitinjak, I Wayan Restu, Nyoman Dati Pertamina. "DAYA DUKUNG DAN I...	<1%
53	Student papers	Telkom University on 2026-04-26	<1%

54	Student papers	Universitas Muria Kudus on 2025-10-15	<1%
55	Publication	Azra'i Mahendra Imam, Aldestra Bagas Wardana, Jodi Armyanto, Moch. Raffi Dwi ...	<1%
56	Student papers	IAKN Palangka Raya on 2026-06-14	<1%
57	Publication	Rifky Permana, Febby Ariyanti Herdiana. "Analisis Klasifikasi Dan Prediksi Pola Pu...	<1%
58	Student papers	Sultan Agung Islamic University on 2018-10-02	<1%
59	Student papers	Universitas Islam Indonesia on 2026-06-15	<1%
60	Student papers	Universitas Putera Batam on 2026-02-02	<1%
61	Publication	Hidayat Hidayat, Musliadi KH, Indar Kusmanto, Munawirah Kadir, Kristian Kristia...	<1%
62	Student papers	Telkom University on 2026-05-06	<1%
63	Student papers	Universitas Muhammadiyah Makassar on 2026-06-07	<1%
64	Student papers	Universitas Pancasila on 2026-03-02	<1%
65	Student papers	Universitas Putera Batam on 2026-02-02	<1%
66	Internet	hayusakola.com	<1%
67	Internet	repository.amikom.ac.id	<1%

68	Student papers	Telkom University on 2026-03-27	<1%
69	Student papers	UIN Syarif Hidayatullah Jakarta on 2025-07-01	<1%
70	Student papers	UPN Veteran Yogyakarta on 2026-03-02	<1%
71	Student papers	Universitas Sanata Dharma on 2026-06-09	<1%
72	Student papers	Universitas Tarumanagara on 2025-12-19	<1%
73	Internet	etheses.iainponorogo.ac.id	<1%
74	Student papers	Institut Teknologi Sumatera on 2025-09-22	<1%
75	Publication	Muhammad Syaifur Rohman, Afrinaldi Afrinaldi, Ahmad Syauqani, Maya Safira. "...	<1%
76	Publication	Simpido Halawa. "IMPLEMENTASI METODE WEIGHTED PRODUCT DALAM SISTEM P...	<1%
77	Student papers	UPN Veteran Yogyakarta on 2026-06-03	<1%
78	Student papers	UPN Veteran Yogyakarta on 2026-06-04	<1%
79	Student papers	Universitas Islam Negeri Sumatera Utara on 2026-05-30	<1%
80	Student papers	Universitas Muhammadiyah Sukabumi on 2026-01-26	<1%
81	Student papers	Universitas Tarumanagara on 2025-12-12	<1%

82	Internet	daftarsekolah.net	<1%
83	Internet	repositori.unimma.ac.id	<1%
84	Student papers	FAKULTAS ILMU KOMPUTER on 2026-01-12	<1%
85	Publication	Fitriani Noer Jamilah. "IMPLEMENTASI DATA MINING UNTUK PREDIKSI DAN KLASI...	<1%
86	Publication	Ibran Simbolon, Putra Aditya, Estetika Br Purba. "Prediksi Performa Akademik Sis...	<1%
87	Student papers	Institut Agama Islam Negeri Curup on 2026-06-11	<1%
88	Publication	Syarifah Syarifah, Kurniadin Abd Latif. "PERANCANGAN E-LEARNING BERBASIS WE...	<1%
89	Publication	Syawaludin Author, Antonius Bagas Sunu W.A Author. "PERBANDINGAN MODEL E...	<1%
90	Student papers	Universitas Dian Nuswantoro on 2015-06-18	<1%
91	Student papers	Universitas Dinamika on 2026-06-11	<1%
92	Student papers	Universitas Putera Batam on 2026-02-03	<1%
93	Internet	ejournal.uin-suka.ac.id	<1%
94	Publication	Harliana Harliana, Tito Prabowo, Ady Alzhava Nuary. "Perbandingan Logistic Reg...	<1%
95	Publication	Kemas Muhammad Wahyu Hidayat. "Pengembangan Sistem Informasi Dalam Me...	<1%

96	Publication	Marwa Sulehu, Wisda Wisda, First Wanita, Markani Markani. "Optimasi Prediksi K...	<1%
97	Student papers	UIN Sunan Ampel Surabaya on 2019-02-10	<1%
98	Internet	digilib.uin-suka.ac.id	<1%
99	Internet	e-journal.hamzanwadi.ac.id	<1%
100	Internet	eprints.upj.ac.id	<1%
101	Internet	pkm.tunasbangsa.ac.id	<1%
102	Internet	www.mdpi.com	<1%
103	Student papers	Institut Bisnis dan Teknologi Indonesia (INSTIKI) on 2025-06-08	<1%
104	Student papers	UIN Raden Intan Lampung on 2026-05-18	<1%
105	Student papers	UPN Veteran Yogyakarta on 2025-12-22	<1%
106	Student papers	UPN Veteran Yogyakarta on 2026-05-05	<1%
107	Student papers	Universitas Brawijaya on 2018-12-13	<1%
108	Student papers	Universitas Muhammadiyah Prof. Dr. Hamka on 2026-04-07	<1%
109	Student papers	Universitas Muhammadiyah Sukabumi on 2025-07-12	<1%

110	Student papers	Universitas Sumatera Utara on 2026-06-05	<1%
111	Internet	ejurnal.methodist.ac.id	<1%
112	Publication	M Syafrizal Zain, Better Swengky, Bradika Almandin Wisesa, Vivin Mahat Putri. "G...	<1%
113	Publication	Septina Simanihuruk, Yasifati Hia. "Pengembangan E-Modul Menggunakan Flip P...	<1%
114	Student papers	UIN Raden Intan Lampung on 2026-02-12	<1%
115	Student papers	UPN Veteran Yogyakarta on 2026-03-02	<1%
116	Student papers	Universitas Bengkulu on 2025-10-17	<1%
117	Student papers	Universitas Budi Luhur on 2026-01-27	<1%
118	Student papers	Universitas Dian Nuswantoro on 2026-01-04	<1%
119	Student papers	Universitas Muhammadiyah Buton on 2026-06-05	<1%
120	Student papers	Universitas Mulawarman on 2021-01-12	<1%
121	Internet	adoc.pub	<1%
122	Internet	ejournal.unesa.ac.id	<1%
123	Internet	ejurnal.stmik-budidarma.ac.id	<1%

124	Internet	eprints3.upgris.ac.id	<1%
125	Internet	jurnal.polibatam.ac.id	<1%
126	Publication	Diana Sava Salsabila, Arimbi Puspitasari, Dwi Rolliawati. "Klasifikasi Jenis Sampah...	<1%
127	Publication	Revi Setiawan, Bayu Priyatna, Elfina Novalia, Baenil Huda. "Klasifikasi Tingkat Pe...	<1%
128	Student papers	UIN Sultan Syarif Kasim Riau on 2020-01-20	<1%
129	Student papers	UPN Veteran Yogyakarta on 2026-02-25	<1%
130	Student papers	Universitas Islam Riau on 2025-09-19	<1%
131	Student papers	Universitas Kristen Duta Wacana on 2026-05-07	<1%
132	Student papers	Universitas Putera Batam on 2026-01-29	<1%
133	Student papers	Universitas Tarumanagara on 2025-12-18	<1%
134	Student papers	University of Colorado, Denver on 2017-01-09	<1%
135	Internet	jurnal.ulb.ac.id	<1%
136	Internet	repository.uinbanten.ac.id	<1%
137	Publication	Alfendio Alif Faudisyah, Kristoko Dwi Hartomo, Hindriyanto Dwi Purnomo. "Detek...	<1%

138	Publication	Anggie Yolanda. "Prediksi Pergerakan Harga Saham Menggunakan Support Vect...	<1%
139	Student papers	Universitas Airlangga-1 on 2025-12-15	<1%
140	Student papers	Universitas Brawijaya on 2020-06-19	<1%
141	Student papers	Universitas Islam Negeri Raden Fatah on 2026-06-02	<1%
142	Student papers	Universitas Muria Kudus on 2018-09-10	<1%
143	Student papers	Universitas Putera Batam on 2026-02-02	<1%
144	Student papers	Universitas Tarumanagara on 2025-12-11	<1%
145	Internet	conference.ut.ac.id	<1%
146	Internet	eprints.uny.ac.id	<1%
147	Internet	journal.jisti.unipol.ac.id	<1%
148	Internet	journal.universitasbumigora.ac.id	<1%
149	Internet	jurnal.fkip.unmul.ac.id	<1%
150	Internet	lab.tik.pnj.ac.id	<1%
151	Student papers	poltekssn on 2025-08-19	<1%

152	Internet	rcf-indonesia.org	<1%
153	Internet	repository.maranatha.edu	<1%
154	Internet	repository.pnj.ac.id	<1%
155	Publication	Baehaqi Wahyu Kurniawan, Pungkas Subarkah, Dinar Mustofa. "Perbandingan N...	<1%
156	Publication	Dedy Atmajaya, Annisa Febrianti, Herdianti Darwis. "Metode SVM dan Naive Baye...	<1%
157	Student papers	Fakultas Ekonomi Universitas Indonesia on 2015-11-23	<1%
158	Student papers	Fakultas Teknik on 2025-07-24	<1%
159	Student papers	Institut Teknologi Nasional Malang on 2021-02-15	<1%
160	Student papers	Sekolah Teknik Elektro & Informatika on 2025-07-23	<1%
161	Publication	Shania Hery Wattimury, Emanuella M C Wattimena, Helda Yunita Taihuttu. "Predi...	<1%
162	Student papers	UNIVERSITAS BUDI LUHUR on 2025-09-17	<1%
163	Student papers	Universitas Brawijaya on 2017-11-27	<1%
164	Student papers	Universitas Islam Riau on 2026-06-10	<1%
165	Student papers	Universitas Maritim Raja Ali Haji on 2024-07-25	<1%

166	Student papers	Universitas Prima Indonesia on 2025-06-24	<1%
167	Student papers	Universitas Tarumanagara on 2025-12-11	<1%
168	Internet	betaraubd.wordpress.com	<1%
169	Internet	digilib.unila.ac.id	<1%
170	Internet	doku.pub	<1%
171	Internet	ejurnal.lkpkaryaprima.id	<1%
172	Internet	eprints.utdi.ac.id	<1%
173	Internet	es.scribd.com	<1%
174	Internet	eudoxuspress.com	<1%
175	Student papers	itera on 2025-07-24	<1%
176	Internet	jurnal.una.ac.id	<1%
177	Internet	library.um.ac.id	<1%
178	Internet	min3merangin.mdrsh.id	<1%
179	Internet	repository.unissula.ac.id	<1%

180	Internet	www.coursehero.com	<1%
181	Internet	www.scribd.com	<1%
182	Publication	Andry Firdiansyah, Ibnu Al Ikrom, Moh. Khamdanni, Wahyu Cahyo Utomo. "Pema...	<1%
183	Publication	Diani Fitri Supriadi, Andin Eka Safitri. "Implementasi Sistem Evaluasi Perkembang...	<1%
184	Publication	Kiki Setiawan, Humam Mu'asyir. "Perbandingan Metode Naïve Bayes dan Support...	<1%
185	Publication	Muhammad Wisnu Nugroho. "Analisis Performa Algoritma Random Forest dalam...	<1%
186	Publication	Nur Jernih Delpi Karianita Zebua, Eliyunus Waruwu, Dedy Setiawan Zebua, Yupite...	<1%
187	Student papers	Politeknik Negeri Bandung on 2017-07-24	<1%
188	Publication	Prestian Ramadhan, Yuhandri, Jhon Veri. "Eksplorasi Algoritma Decision Tree unt...	<1%
189	Publication	Rini Sovia, Eka Praja Wiyata Mandala, Sitty Mardhiah. "Algoritma K-Means dalam ...	<1%
190	Student papers	STKIP Sumatera Barat on 2025-08-22	<1%
191	Student papers	STT PLN on 2021-08-29	<1%
192	Student papers	Sriwijaya University on 2021-03-16	<1%
193	Student papers	Surabaya University on 2016-06-10	<1%

194	Student papers	UIN Raden Intan Lampung on 2026-04-02	<1%
195	Student papers	Universitas Amikom on 2021-01-23	<1%
196	Student papers	Universitas Bengkulu on 2025-09-03	<1%
197	Student papers	Universitas Brawijaya on 2016-06-17	<1%
198	Student papers	Universitas Brawijaya on 2016-10-12	<1%
199	Student papers	Universitas Brawijaya on 2016-11-15	<1%
200	Student papers	Universitas Brawijaya on 2017-07-30	<1%
201	Student papers	Universitas Brawijaya on 2019-04-22	<1%
202	Student papers	Universitas Maritim Raja Ali Haji on 2025-01-12	<1%
203	Student papers	Universitas Muhammadiyah Sumatera Utara on 2025-08-21	<1%
204	Student papers	Universitas Muria Kudus on 2019-08-30	<1%
205	Student papers	Universitas Pancasila on 2026-03-04	<1%
206	Student papers	Universitas Putera Batam on 2020-12-01	<1%
207	Student papers	Universitas Putera Batam on 2026-01-29	<1%

208	Student papers	Universitas Sangga Buana YPKP on 2026-02-23	<1%
209	Student papers	Universitas Sumatera Utara on 2026-06-04	<1%
210	Student papers	Universitas Tarumanagara on 2025-12-11	<1%
211	Student papers	Universitas Tarumanagara on 2025-12-15	<1%
212	Student papers	Xavier University on 2023-09-07	<1%
213	Internet	anzdoc.com	<1%
214	Internet	dspace.uui.ac.id	<1%
215	Internet	eprint.stieww.ac.id	<1%
216	Internet	garuda.kemdikbud.go.id	<1%
217	Internet	jacis.pubmedia.id	<1%
218	Internet	journal.uniku.ac.id	<1%
219	Internet	jurnal.unpand.ac.id	<1%
220	Internet	ojs.unud.ac.id	<1%
221	Internet	repositori.uin-alauddin.ac.id	<1%

222	Internet	www.mtsnselatkapuas.sch.id	<1%
223	Internet	www.ojs.unkriswina.ac.id	<1%
224	Student papers	Fakultas Teknik on 2025-08-07	<1%
225	Student papers	Politeknik Negeri Bandung on 2025-06-13	<1%
226	Student papers	UPN Veteran Yogyakarta on 2025-10-01	<1%
227	Student papers	Universitas Negeri Jakarta on 2019-06-24	<1%
228	Student papers	Universitas Pamulang on 2026-06-09	<1%
229	Student papers	Universitas Putera Batam on 2026-02-03	<1%
230	Student papers	Universitas Sanata Dharma on 2026-06-09	<1%
231	Student papers	Universitas Sanata Dharma on 2026-06-15	<1%
232	Student papers	poltekssn on 2025-07-09	<1%
233	Student papers	UIN Syarif Hidayatullah Jakarta on 2023-04-04	<1%
234	Student papers	UPN Veteran Yogyakarta on 2026-05-05	<1%
235	Student papers	Universitas Islam Indonesia on 2019-11-27	<1%

236	Student papers	Universitas Islam Indonesia on 2026-04-20	<1%
237	Student papers	Universitas Jember on 2023-01-13	<1%
238	Student papers	Universitas Pamulang on 2025-07-20	<1%
239	Student papers	Universitas Pendidikan Indonesia on 2015-06-10	<1%
240	Student papers	Universitas Putera Batam on 2021-03-15	<1%
241	Student papers	Universitas Putera Batam on 2026-01-30	<1%
242	Student papers	Universitas Sumatera Utara on 2026-05-18	<1%
243	Student papers	Universitas Terbuka on 2026-02-06	<1%
244	Internet	zombiedoc.com	<1%

98

BAB I**PENDAHULUAN****A. Latar Belakang Masalah**

Perkembangan teknologi informasi memberikan manfaat besar di berbagai bidang, termasuk pendidikan. Kehadiran data yang semakin banyak di lingkungan sekolah, seperti data akademik dan non-akademik, seperti nilai rata-rata rapor akhir semester, kehadiran, sikap disiplin, prestasi lomba, serta partisipasi dalam kegiatan ekstrakurikuler, membuka kesempatan untuk menghasilkan informasi yang lebih lengkap. Salah satu contohnya adalah proses penentuan siswa berprestasi yang sebelumnya dilakukan secara manual oleh pihak sekolah. Selain itu, keterbatasan dalam menganalisis data, kemungkinan kesalahan saat mencatat, serta ketidaksejajaran standar penilaian antar pihak yang menilai. Hal ini semakin diperparah karena metode manual tidak efisien dalam mengolah data dalam jumlah besar.

218

SMP Negeri 4 Kediri adalah salah satu lembaga pendidikan yang terus berusaha meningkatkan kualitas proses belajar mengajar. Namun, pengelolaan data siswa yang semakin rumit setiap tahun menimbulkan berbagai masalah serius, terutama dalam mengidentifikasi siswa yang berprestasi. Penilaian prestasi siswa masih bergantung pada hitungan akademik dan non-akademik secara tradisional yang terbatas dan sering kali tidak konsisten antar periode, sehingga data yang tersedia tidak dapat dimanfaatkan secara optimal. Kondisi ini membuat hasil analisis menjadi kurang akurat dan tidak mampu menggambarkan perkembangan siswa secara menyeluruh. Akibatnya, beberapa siswa dengan kemampuan tinggi tidak teridentifikasi secara tepat dan penetapan siswa berprestasi menjadi kurang optimal.

Dalam bidang pembelajaran *machine learning*, algoritma klasifikasi memainkan peran penting dalam melakukan prediksi berdasarkan pola data

historis. Menurut (Nachouki & Abou Naaj, 2022), algoritma klasifikasi seperti *Random Forest* mampu membantu sistem memahami struktur data secara lebih mendalam, sehingga menghasilkan prediksi performa siswa yang lebih akurat dan dapat mendukung proses pengambilan keputusan akademik. Salah satu algoritma yang terbukti efektif dalam mengolah data yang beragam dan kompleks adalah *Random Forest*. (Manzali et al., 2024) menjelaskan bahwa *Random Forest* bekerja dengan membangun sejumlah pohon keputusan dan menggabungkan prediksi dari setiap pohon sehingga menghasilkan performa yang lebih stabil. Hal ini diperkuat oleh penelitian (Sofiyatus Zawiyah et al., 2025) serta (Huang & Wei, 2022) yang menunjukkan bahwa *Random Forest* mampu menangani data berdimensi tinggi, tahan terhadap noise, serta bekerja baik pada data siswa yang kompleks.

Penerapan *Random Forest* dalam memprediksi siswa berprestasi memberikan manfaat signifikan bagi dunia pendidikan. (Chen & Jin, (2024) menegaskan bahwa pendekatan *machine learning* berbasis *Random Forest* dapat memberikan informasi yang lebih akurat mengenai potensi siswa serta membantu institusi pendidikan melakukan intervensi secara tepat sasaran. Selain itu, penelitian oleh (Nastiti et al., 2025) dan (Saputri et al., 2025) menunjukkan bahwa model prediksi berbasis *Random Forest* dapat dimanfaatkan untuk mengevaluasi kinerja siswa dan mendukung optimalisasi proses pembelajaran secara berkelanjutan.

Selain manfaat praktis yang diperoleh, penelitian mengenai penerapan *Random Forest* dalam konteks pendidikan juga memberi kontribusi penting secara akademik. (Nastiti et al., 2025) serta (Chen & Jin, 2024) menegaskan bahwa penelitian terkait penggunaan algoritma *Random Forest* dalam prediksi performa siswa terus berkembang dan perlu diperluas, khususnya pada jenjang sekolah menengah pertama. Studi-studi tersebut dapat menjadi dasar bagi penelitian selanjutnya, misalnya dengan membandingkan efektivitas berbagai metode klasifikasi atau dengan menambahkan variabel baru untuk

meningkatkan akurasi evaluasi siswa berprestasi. Dengan demikian, penelitian ini tidak hanya memberikan solusi praktis bagi sekolah, tetapi juga memperkaya literatur dalam bidang *machine learning* dan pendidikan.

Penelitian mengenai penerapan *Random Forest* dalam sistem klasifikasi untuk memprediksi siswa berprestasi perlu dilakukan. Dengan menggabungkan teknologi pengolahan data dan kebutuhan sekolah akan sistem prediksi yang lebih efisien, penelitian ini diharapkan memberikan manfaat baik bagi pengembangan sistem informasi sekolah maupun membantu pengambilan keputusan yang lebih tepat dan terukur.

B. Identifikasi Masalah

Berdasarkan uraian latar belakang yang sudah dijelaskan, dapat diidentifikasi beberapa permasalahan terkait proses penentuan siswa berprestasi di sekolah, sebagai berikut:

1. Proses penentuan siswa berprestasi di SMP Negeri 4 Kediri masih dilakukan secara manual, sehingga memerlukan waktu yang lama dan rentan terhadap kesalahan.
2. Data siswa yang beragam dan rumit belum dimanfaatkan secara maksimal untuk membantu memprediksi prestasi mereka.
3. Sekolah belum memiliki sistem yang berbasis *machine learning* yang mampu mengklasifikasikan siswa secara tepat dan akurat.
4. Penilaian terhadap siswa berprestasi masih mengandung unsur subjektivitas, sehingga hasilnya tidak konsisten.

C. Rumusan Masalah

Berdasarkan masalah yang sudah diidentifikasi, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana cara merancang model klasifikasi yang mampu memprediksi prestasi siswa dengan lebih tepat berdasarkan data yang sudah digabungkan?

26

2. Bagaimana efektivitas algoritma *Random Forest* dalam meningkatkan ketepatan model prediksi prestasi siswa menggunakan data dari SMP Negeri 4 Kediri?

D. Batasan Masalah

87

Untuk mencegah pembahasan menjadi terlalu luas dan agar penelitian lebih terarah, batasan-batasan dalam penelitian ini ditentukan sebagai berikut:

50

1. Penelitian dilaksanakan di SMP Negeri 4 Kediri.
2. Data yang digunakan terbatas pada data akademik dan data non-akademik yang mendukung prestasi siswa, yang tersedia di sekolah tersebut, khususnya untuk siswa kelas VIII tahun 2025-2026
3. Metode klasifikasi yang digunakan berupa algoritma *Random Forest*, tanpa dibandingkan dengan metode lainnya.
4. Sistem yang dikembangkan berfungsi untuk memprediksi siswa berprestasi, tanpa mencakup rekomendasi tindak lanjut.
5. Evaluasi kinerja mode dilakukan dengan melihat akurasi dan metrik penilaian berdasarkan dataset yang disediakan oleh sekolah.
6. Kriteria siswa berprestasi dalam penelitian ini mencakup nilai rata-rata dari rapor akhir semester, kehadiran siswa, kedisiplinan, prestasi lomba, serta partisipasi dalam kegiatan ekstrakurikuler.

E. Tujuan Penelitian

160

Agar fokus pembahasan dan arah penelitian tetap jelas, tujuan penelitian ini dirumuskan sebagai berikut:

195

1. Membuat model klasifikasi berbasis machine learning yang dapat memprediksi siswa berprestasi dengan tingkat keakuratan yang lebih baik.
2. Menggunakan algoritma *Random Forest* dan menilai bagaimana performanya dalam memprediksi prestasi siswa kelas VIII di SMP Negeri 4 Kediri.

31

244

33

F. Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat baik secara teoritis maupun praktis sebagai berikut:

31

4

a. Manfaat Teoritis

Membantu memperkaya pengetahuan di bidang *machine learning*, terutama dalam penerapan algoritma *Random Forest* untuk mengklasifikasikan data pendidikan, serta menjadi acuan bagi penelitian selanjutnya terkait prediksi prestasi siswa, penerapan model klasifikasi pada data akademik dan non-akademik, serta analisis di bidang pendidikan.

221

b. Manfaat Praktis

- 1) Bagi Sekolah: Membantu sekolah mengenali siswa berprestasi secara jelas, cepat, dan tepat melalui sistem prediksi, sehingga mempermudah dan mempercepat identifikasi siswa berprestasi serta mendukung program penghargaan dan pembinaan.
- 2) Bagi Guru: Memudahkan guru dalam mengawasi perkembangan siswa melalui sistem prediksi berbasis web yang mudah diakses, serta membantu dalam proses pemantauan dan pembinaan berdasarkan data yang akurat.
- 3) Bagi Peneliti: Memberikan alat bantu analisis data untuk mendukung evaluasi dan pengambilan keputusan dalam pengembangan potensi siswa, serta menjadi dasar penelitian lanjutan sistem prediksi prestasi.

215

BAB II**LANDASAN TEORI**

38

A. Landasan Teori**1. Siswa Berprestasi**

Siswa berprestasi adalah peserta didik yang menunjukkan pencapaian unggul baik dalam aspek akademik maupun non-akademik melalui keterlibatan aktif dalam pembelajaran dan kegiatan ekstrakurikuler. Menurut (Siti Asiyah & Novebri, 2024), prestasi tersebut dapat ditingkatkan melalui manajemen peserta didik yang efektif, seperti pembinaan minat dan bakat, dukungan ekstrakurikuler, serta motivasi belajar yang berkelanjutan.

Menurut (Prassida & Maulida, 2023), siswa berprestasi dapat dinilai berdasarkan lima indikator utama, yaitu:

- a. Nilai rata-rata rapor,
- b. Kehadiran dan kedisiplinan siswa,
- c. Prestasi atau penghargaan yang diperoleh, serta
- d. Partisipasi dalam kegiatan ekstrakurikuler.

178

Kombinasi indikator akademik dan non-akademik menghasilkan penilaian yang lebih komprehensif terhadap kemampuan siswa secara menyeluruh. Temuan (Huang & Wei, 2022) juga mendukung hal tersebut, di mana aktivitas non-akademik seperti ekstrakurikuler terbukti memberikan pengaruh positif terhadap prestasi akademik siswa karena melalui kegiatan tersebut siswa belajar keterampilan sosial, tanggung jawab, dan kerja sama yang berkontribusi pada hasil belajar. Selain itu, penelitian (Chen & Jin, 2024) menunjukkan bahwa kehadiran dan kedisiplinan siswa memiliki hubungan signifikan dengan peningkatan prestasi akademik, sehingga kedua aspek ini menjadi faktor penting dalam mengevaluasi performa belajar siswa.

Berdasarkan temuan penelitian-penelitian tersebut, maka penelitian ini menggunakan empat indikator utama dalam menentukan siswa berprestasi, yaitu nilai rapor akhir semester, kehadiran dan kedisiplinan, prestasi lomba atau penghargaan, serta partisipasi dalam kegiatan ekstrakurikuler.

2. Klasifikasi Data

Klasifikasi merupakan salah satu teknik dalam *data mining* yang digunakan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan pola atau karakteristik yang ditemukan pada data. Menurut (Fang et al., 2024) dalam penelitian berjudul "*Prediction model of middle school student performance based on MBSO and MDBO-BP-AdaBoost*", klasifikasi adalah proses membangun model prediktif dengan mempelajari pola dari data historis untuk menentukan label pada data baru secara otomatis. Penelitian tersebut juga menunjukkan bahwa teknik klasifikasi sangat efektif digunakan untuk menganalisis performa akademik siswa tingkat sekolah menengah pertama (SMP), sehingga sesuai dengan konteks penelitian ini.

Langkah-langkah umum dalam proses klasifikasi data meliputi:

a. Pengumpulan data

Data akademik dan non-akademik siswa dikumpulkan dari catatan sekolah, meliputi nilai rapor akhir semester, kehadiran, sikap, prestasi lomba, dan partisipasi ekstrakurikuler. Variabel-variabel ini sejalan dengan yang digunakan oleh (Fang et al., 2024), yaitu kombinasi faktor akademik dan perilaku siswa untuk membangun model prediktif yang lebih akurat.

b. Praproses data (*preprocessing*)

Pada tahap ini, data dibersihkan dari nilai kosong, duplikasi, dan anomali, serta dilakukan normalisasi untuk memastikan seluruh fitur

berada pada skala yang seragam. Proses ini penting sebagaimana dijelaskan oleh (Fang et al., 2024), karena kualitas data sangat mempengaruhi kinerja model klasifikasi.

c. Pelabelan data (*labeling*)

Data siswa kemudian diberi label “berprestasi” dan “tidak berprestasi” berdasarkan indikator yang diadaptasi dari penelitian (Syah et al. 2023), seperti nilai rapor, kehadiran, sikap, prestasi, dan aktivitas ekstrakurikuler. Pendekatan pelabelan ini sejalan dengan konsep klasifikasi dalam penelitian (Fang et al., 2024), yang mengelompokkan siswa berdasarkan performa mereka.

d. Pelatihan mode (*training*)

Model klasifikasi dilatih menggunakan algoritma *machine learning* untuk mempelajari pola hubungan antarvariabel. Meskipun (Fang et al. 2024) menggunakan kombinasi *MBSO* dan *MDBO-BP-AdaBoost*, penelitian ini menggunakan algoritma *Random Forest*, yang juga termasuk metode *ensemble* dan telah terbukti efektif dalam memprediksi performa siswa.

e. Pengujian model (*testing*)

Tahap pengujian dilakukan untuk mengevaluasi performa model dalam memprediksi data baru. Sesuai dengan praktik yang dijelaskan oleh (Fang et al. 2024), evaluasi model dilakukan menggunakan metrik seperti *accuracy*, *precision*, *recall*, dan *F1-score* untuk memastikan stabilitas dan keandalan hasil klasifikasi.

3. Machine Learning

Machine Learning merupakan cabang dari *Artificial Intelligence* (AI) yang memungkinkan komputer belajar dari data untuk menghasilkan keputusan atau prediksi tanpa harus diprogram secara eksplisit. Menurut (Pratiwi et al., 2025) dalam Strategi Manajemen Pendidikan Berbasis *Machine Learning* untuk Prediksi Prestasi Siswa, *machine learning* bekerja dengan memanfaatkan algoritma matematis yang mampu

mengidentifikasi pola dari data yang kompleks sehingga dapat membangun model prediksi secara otomatis. Penelitian tersebut menunjukkan bahwa algoritma seperti *Random Forest* dan *SVM* dapat digunakan secara efektif untuk memprediksi prestasi akademik siswa dengan hasil validasi silang yang memperlihatkan bahwa *Random Forest* memiliki akurasi dan stabilitas yang lebih unggul pada data pendidikan multidimensi.

Machine Learning umumnya dibagi menjadi tiga kategori utama, yaitu sebagaimana dijelaskan dalam artikel (Shaveta, 2023) yang diterbitkan oleh, bahwa *machine learning* terdiri dari tiga pendekatan *fundamental*, yaitu *supervised learning*, *unsupervised learning*, dan *reinforcement learning*, yang masing-masing memiliki tugas dan fungsi berbeda dalam proses pemodelan data.

a. Supervised Learning

Pada pendekatan ini, data latih memiliki label yang sudah diketahui. Model belajar mengenali hubungan antara fitur *input* dan label *output* sehingga mampu melakukan prediksi pada data baru.

b. Unsupervised Learning

Data tidak memiliki label, dan algoritma bertugas menemukan struktur atau pola tersembunyi dalam data, misalnya melalui pengelompokan (*clustering*).

c. Reinforcement Learning

Pendekatan ini sistem belajar melalui mekanisme umpan balik berupa *reward* atau *punishment* sehingga model secara bertahap mengoptimalkan keputusan berdasarkan pengalaman.

Pendekatan ini sistem belajar melalui mekanisme umpan balik berupa *reward* atau *punishment* sehingga model secara bertahap mengoptimalkan keputusan berdasarkan pengalaman.

Penelitian ini menggunakan pendekatan *supervised learning* karena model perlu dilatih untuk mengklasifikasikan siswa ke dalam kategori “berprestasi” dan “tidak berprestasi”. Namun, data siswa yang diperoleh dari sekolah tidak memiliki label bawaan. Oleh karena itu, dilakukan proses pelabelan (*labeling*) terlebih dahulu berdasarkan empat indikator utama, yaitu nilai rapor, kedisiplinan, prestasi lomba, dan partisipasi ekstrakurikuler. Setelah data terlabeli dengan benar, model kemudian dilatih menggunakan algoritma *Random Forest* untuk menghasilkan prediksi yang akurat terkait status siswa berprestasi.

4. Algoritma *Random Forest*

Algoritma *Random Forest* merupakan metode *ensemble learning* yang terdiri dari sekumpulan *decision tree* yang dibangun secara acak. Setiap pohon keputusan memberikan prediksi, kemudian hasil dari seluruh pohon digabungkan melalui voting mayoritas untuk klasifikasi atau rata-rata untuk regresi. Pendekatan ini menghasilkan model yang lebih akurat dan stabil dibandingkan model tunggal.

Menurut (Yin et al., 2023) dalam *Information in Medicine Unlocked*, *Random Forest* memiliki kemampuan tinggi dalam mengelola data berdimensi besar dan menunjukkan akurasi prediksi di atas 90% dalam memodelkan performa akademik siswa. Penelitian tersebut juga menegaskan bahwa kombinasi banyak pohon membuat algoritma ini lebih tahan terhadap *noise* serta mampu memberikan hasil prediksi yang konsisten.

Langkah-Langkah Kerja Algoritma *Random Forest* Menurut (Iranzad & Liu, 2025).

a. Bootstrap Sampling

Dataset diambil secara acak dengan pengembalian (*bootstrapping*) untuk membentuk beberapa subset pelatihan.

b. Membangun Banyak *Decision Tree*

Setiap subset digunakan untuk membangun satu *decision tree* dengan pemilihan fitur secara acak (*random feature selection*) pada setiap node.

c. Proses Prediksi

Masing-masing pohon menghasilkan prediksi independen terhadap data uji.

d. Voting Mayoritas (*Averaging*)

- 1) Klasifikasi: hasil akhir diambil berdasarkan suara terbanyak (*majority vote*).
- 2) Regresi: hasil akhir merupakan rata-rata prediksi seluruh pohon.

Rumus Algoritma *Random Forest*:

Untuk regresi:

$$\hat{y} = \sum_{i=1}^N T_i(x) \dots \dots \dots (2.1)$$

Untuk klasifikasi:

$$\hat{y} = \text{mode}\big(T_1(x), T_2(x), \dots, T_N(x)\big) \dots \dots \dots (2.2)$$

Keterangan:

$\hat{y}$ = hasil prediksi akhir

N = jumlah pohon Keputusan

(x) = hasil prediksi dari pohon ke- i

Berdasarkan temuan (Yağcı, 2022) dan literatur lain:

- a. Mampu menangani data berdimensi besar dan variabel multikompleks.

- b. Tidak mudah *overfitting* karena melibatkan banyak pohon secara paralel.
- c. Menyediakan *feature importance*, membantu mengidentifikasi variabel paling berpengaruh terhadap hasil prediksi.
- d. Stabil dan akurat, cocok untuk dataset pendidikan yang mengandung variabilitas tinggi.

5. Evaluasi Klasifikasi Model

Evaluasi model klasifikasi dilakukan untuk menilai sejauh mana algoritma mampu menghasilkan prediksi yang akurat terhadap label data. Berdasarkan penelitian Menurut “(Fazel Hesar & Foing, n.d. 2023), berbagai algoritma klasifikasi seperti *Random Forest, Support Vector Machine, dan Logistic Regression* dievaluasi menggunakan metrik *accuracy, precision, recall, dan F1-score*. Penelitian tersebut menunjukkan bahwa metrik-metrik ini memberikan gambaran komprehensif mengenai performa model, terutama saat menangani data kompleks seperti *time series* dari misi Kepler. Dengan demikian, penggunaan *accuracy, precision, recall, dan F1-score* dalam penelitian ini sangat tepat sebagai acuan dalam menilai kualitas prediksi model klasifikasi.

Berdasarkan penelitian (Waqia Wania et al., n.d. 2025) metrik evaluasi berupa *accuracy, precision, recall, dan F1-score* terbukti efektif digunakan untuk menilai kinerja algoritma *Random Forest* dalam prediksi performa siswa.

a. Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots (2.3)$$

Keterangan:

TP (True Positive) = prediksi positif yang benar

1

TN (True Negative) = prediksi negatif yang benar

FP (False Positive) = prediksi positif yang salah

FN (False Negative) = prediksi negatif yang salah

b. *Precision*

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots (2.4)$$

c. *Recall*

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots (2.5)$$

d. *F1-Score*

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \dots\dots\dots (2.6)$$

B. Kajian Pustaka

Penelitian sebelumnya menjadi acuan penting dalam mengevaluasi bagaimana baiknya metode klasifikasi yang digunakan di bidang pendidikan. Dengan menganalisis kajian ini, kita bisa tahu seberapa luas algoritma *Random Forest* dan metode lainnya diterapkan serta hasil atau performa yang berhasil dicapai.

Penelitian Afandi et al., n.d. (2022) memprediksi kebutuhan stok barang menggunakan kombinasi algoritma Apriori dan metode *Moving Average*. Apriori digunakan untuk menemukan pola pembelian, sedangkan *Moving Average* untuk memprediksi stok berdasarkan 365 data transaksi tahun 2021. Hasilnya menunjukkan produk semen segitiga, besi begel 6×12, dan 8×12 memiliki penjualan tertinggi dengan akurasi prediksi 63%–81%. Penelitian ini relevan sebagai pembandingan karena menunjukkan efektivitas teknik prediksi berbasis data historis.

Penelitian yang dilakukan oleh Laschka et al., (2024) mengembangkan sistem prediksi minat penjualan jaket menggunakan algoritma *Naive Bayes*. Penelitian ini bertujuan membantu pelaku usaha dalam mengidentifikasi produk dengan minat tertinggi berdasarkan data historis seperti model, corak,

186

26

29
50
92

bahan, dan harga. Hasil penelitian menunjukkan bahwa metode *Naive Bayes* mampu memberikan akurasi prediksi yang baik sehingga dapat dijadikan alat bantu pengambilan keputusan. Meskipun konteksnya bukan pendidikan, penelitian ini relevan karena sama-sama menerapkan algoritma klasifikasi berbasis *data mining*.

138
161
184
103
214

Penelitian oleh Saputro & Swanjaya, n.d. (2023) menerapkan metode *Neural Network* untuk memprediksi harga saham menggunakan data numerik harian seperti harga pembukaan, penutupan, volume transaksi, dan *net foreign flow*. Hasil penelitian memperlihatkan bahwa algoritma jaringan saraf tiruan mampu mempelajari pola fluktuasi data secara efektif dan menghasilkan prediksi yang stabil. Relevansi penelitian ini terletak pada penggunaan teknik *machine learning* dalam analisis prediktif, serupa dengan penelitian ini yang menggunakan algoritma *Random Forest* meskipun objek yang diteliti berbeda.

217
8
61

Penelitian yang dilakukan oleh Nuralia & Prabowo, (2023) dalam artikel berjudul “Implementasi *Naive Bayes Classifier* Dalam Memprediksi Kelulusan Mahasiswa” yang diterbitkan pada jurnal JACIS memanfaatkan algoritma *Naive Bayes Classification* untuk memprediksi kelulusan mahasiswa berdasarkan data akademik. Penelitian tersebut menguji tiga skenario pembagian data (80:20, 50:50, dan 20:80) dan menunjukkan bahwa skenario 80:20 menghasilkan akurasi tertinggi. Hasil penelitian menunjukkan bahwa *Naive Bayes* mampu memprediksi kelulusan dengan akurasi yang cukup baik. Temuan ini menjadi pembanding bagi penelitian yang menggunakan *Random Forest* karena metode tersebut mampu menangani variabel kompleks dan interaksi antarfitur untuk menghasilkan prediksi yang lebih optimal.

207
191

Penelitian yang dilakukan oleh Saraswati et al., (2023) yang dipublikasikan melalui *ResearchGate* memanfaatkan kombinasi algoritma Apriori dan metode *Single Moving Average* untuk memprediksi kebutuhan stok barang pada perusahaan *retail*. Algoritma Apriori digunakan untuk

229 mengkali pola transaksi yang sering muncul, sementara metode *Single Moving Average* berfungsi menghasilkan prediksi permintaan berdasarkan data historis penjualan. Hasil penelitian menunjukkan bahwa integrasi kedua metode tersebut mampu memberikan prediksi stok yang lebih akurat dan mendukung efisiensi proses pengambilan keputusan dalam manajemen persediaan. Meskipun penelitian ini berada dalam ranah bisnis, relevansinya terletak pada penerapan teknik prediktif berbasis *data mining*, yang sejalan dengan penelitian ini dalam memanfaatkan metode prediksi untuk menentukan siswa berprestasi.

84 Penelitian yang dilakukan Mrg et al., (2024) memanfaatkan algoritma *Random Forest* untuk mengklasifikasikan siswa terbaik pada tingkat Madrasah Aliyah. Penelitian tersebut mengolah berbagai indikator penilaian, seperti nilai akademik, nilai agama, etika, kemampuan kepemimpinan, dan keikutsertaan siswa dalam kegiatan ekstrakurikuler. Hasil penelitian menunjukkan bahwa *Random Forest* menghasilkan klasifikasi yang stabil dan objektif dengan akurasi 0,96. Temuan ini memperkuat keandalan *Random Forest* dalam bidang pendidikan serta relevan sebagai pembanding dalam penelitian identifikasi siswa berprestasi.

20 Penelitian yang dilakukan oleh Nachouki & Abou Naaj, (2022) menggunakan algoritma *Random Forest* untuk memprediksi performa mahasiswa dalam rangka meningkatkan bimbingan akademik. Data perilaku belajar dan nilai akademik dianalisis untuk menghasilkan model prediksi dengan akurasi tinggi. Hasilnya menunjukkan bahwa *Random Forest* lebih unggul dibandingkan beberapa algoritma lainnya, sehingga relevan sebagai dasar metodologis dalam penelitian ini.

96 Penelitian yang dilakukan oleh Hemasri & Kiran, (2022) meneliti prediksi hasil belajar siswa menggunakan algoritma *Random Forest* dengan mempertimbangkan variabel seperti nilai ujian, kegiatan sekolah, dan catatan disiplin. Hasilnya menunjukkan bahwa kombinasi antara data akademik dan

19

non-akademik memberikan hasil prediksi yang lebih akurat. Penelitian ini sangat relevan karena mendukung penggunaan variabel multidimensi dalam menentukan siswa berprestasi

47

Penelitian yang di lakukan Riskiyono & Mahdiana, (2024) menunjukkan bahwa algoritma *Random Forest* memiliki potensi besar dalam memprediksi kelulusan mahasiswa. Studi ini menganalisis faktor akademik, seperti kehadiran, IPK, dan nilai mata kuliah inti. Hasilnya menunjukkan bahwa *Random Forest* mampu memprediksi kelulusan tepat waktu dengan akurasi yang baik serta mendukung pengambilan keputusan dan intervensi dini di bidang akademik.

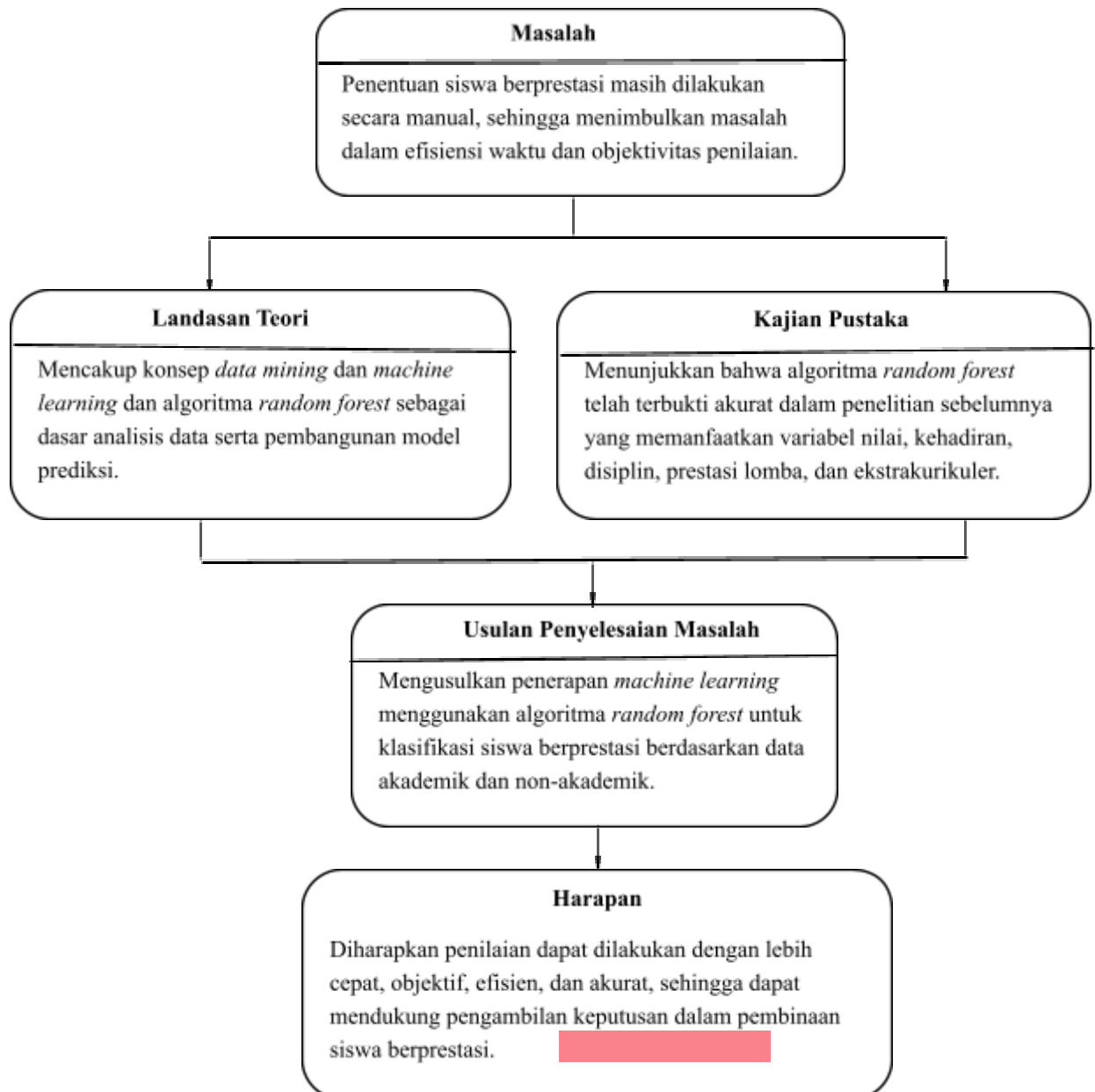
182

47

Penelitian yang di lakukan Firdiansyah et al., (2024) membahas pemanfaatan *data mining* untuk memprediksi kelulusan mahasiswa pada mata kuliah tertentu dengan menggunakan algoritma klasifikasi berbasis nilai prasyarat dan nilai ujian. Hasil penelitian menunjukkan bahwa model mampu memprediksi kelulusan dengan akurasi yang baik dan mendukung dosen dalam menentukan strategi pembelajaran. Studi ini juga menegaskan efektivitas data mining dalam menganalisis performa akademik mahasiswa.

C. Kerangka Berpikir

Kerangka berpikir ini menggambarkan proses penelitian yang dimulai dari permasalahan yang terjadi, teori serta hasil penelitian sebelumnya sebagai dasar, hingga solusi yang ditawarkan dan harapan yang ingin dicapai.



Gambar 2. 1 Kerangka Berpikir

100

176

Kerangka berpikir pada penelitian ini berangkat dari permasalahan bahwa proses penentuan siswa berprestasi masih dilakukan secara manual, sehingga menimbulkan kendala dalam hal efisiensi waktu dan objektivitas penilaian. Untuk menjawab permasalahan tersebut, penelitian menggunakan landasan teori mengenai *data mining*, *machine learning*, serta algoritma *random forest* sebagai dasar analisis dan pembangunan model prediksi. Selain itu, kajian pustaka menunjukkan bahwa algoritma *random forest* telah terbukti akurat dalam penelitian sebelumnya, terutama ketika digunakan pada data nilai, kehadiran, kedisiplinan, prestasi lomba, dan aktivitas ekstrakurikuler.

15

Berdasarkan teori dan bukti penelitian terdahulu, penelitian ini mengusulkan penerapan algoritma *random forest* sebagai solusi untuk mengklasifikasikan siswa berprestasi secara otomatis berdasarkan data akademik dan non-akademik. Penerapan metode ini diharapkan mampu meningkatkan kualitas proses penilaian, baik dari segi kecepatan, objektivitas, efisiensi, maupun akurasi. Pada akhirnya, sistem klasifikasi berbasis *machine learning* ini diharapkan dapat membantu pihak sekolah dalam mengambil keputusan pembinaan siswa berprestasi secara lebih tepat dan terukur.

128

BAB III

METODE PENELITIAN

A. Desain Penelitian

Desain penelitian adalah rencana atau strategi yang menyeluruh digunakan untuk menjawab pertanyaan penelitian. Penelitian ini termasuk dalam kategori desain penelitian pengembangan karena berfokus pada proses perancangan, pengembangan, dan pengujian sistem klasifikasi berbasis Algoritma *Random Forest* untuk memprediksi siswa berprestasi di SMP Negeri 4 Kediri.

Selain itu, penelitian ini juga memiliki unsur eksperimental karena melakukan pengujian terhadap kinerja model dengan mengubah variabel independen, seperti fitur data dan parameter algoritma, untuk melihat dampaknya terhadap hasil klasifikasi, yaitu variabel dependen.

1. Jenis Penelitian (*Research Type*)

Jenis penelitian yang digunakan adalah penelitian pengembangan dengan pendekatan kuantitatif eksperimental. Penelitian ini dilakukan untuk:

- Mengembangkan sistem klasifikasi berbasis *machine learning* yang mampu memprediksi siswa berprestasi.
- Menguji tingkat akurasi algoritma *Random Forest* sebagai metode klasifikasi terhadap data akademik dan non-akademik siswa.

Desain ini menggabungkan tahapan pengembangan perangkat lunak dengan eksperimen terhadap model klasifikasi.

2. Desain Pengembangan

Proses pengembangan yang dilakukan dalam penelitian ini mengacu pada langkah-langkah dalam rekayasa perangkat lunak. Berikut penjelasan masing-masing tahapan:

- a. Analisis Kebutuhan: Mengidentifikasi kebutuhan dari pengguna, yaitu guru dan pihak sekolah, serta menentukan data yang digunakan dalam proses klasifikasi, seperti nilai rapor, kehadiran, kedisiplinan, prestasi dalam lomba, serta partisipasi dalam kegiatan ekstrakurikuler.
- b. Perancangan Sistem: Merancang struktur sistem secara keseluruhan dan menentukan alur kerja untuk model klasifikasi yang menggunakan algoritma *Random Forest*, termasuk dalam merancang tampilan antarmuka pengguna.
- c. Implementasi Sistem: Mengembangkan algoritma *Random Forest* dengan menggunakan bahasa pemrograman *Python* dan beberapa pustaka pendukung seperti *Scikit-learn*.
- d. Pengujian dan Evaluasi: Melakukan pengujian terhadap model berdasarkan dataset siswa. Hasil pengujian dinilai menggunakan metrik *accuracy, precision, recall*, dan *F1-score* untuk mengevaluasi kinerja model prediksi.

3. Desain Eksperimen

Penelitian ini juga menggunakan pendekatan eksperimen kuantitatif karena dalam penelitian ini dilakukan pengujian algoritma dengan memperhatikan pengaruh beberapa variabel terhadap hasil prediksi. Desain eksperimen yang digunakan adalah *post-test only control group design*, yaitu membandingkan hasil klasifikasi sistem setelah dilakukan pelatihan model dengan data yang sudah disiapkan.

- a. Variabel Independen
 - 1) Parameter algoritma *Random Forest* (jumlah pohon, kedalaman pohon, dan fitur yang digunakan).
 - 2) Jenis data (kombinasi data akademik dan non-akademik).
- b. Variabel Dependen
 - 1) Akurasi model prediksi.
 - 2) Nilai *precision, recall*, dan *F1-score*.
- c. Variabel Kontrol

- 1) *Dataset* siswa yang digunakan sama untuk seluruh eksperimen.
- 2) Lingkungan pengujian dan tools yang digunakan tetap (*Python*, *Scikit-learn*).

4. Studi Kasus

Penelitian ini menggunakan studi kasus di SMP Negeri 4 Kediri, yang menjadi lokasi pengambilan data siswa. Data diperoleh dari catatan akademik dan non-akademik tahun ajaran 2025–2026.

5. Metode Pengumpulan Data

Metode pengumpulan data dilakukan melalui:

- a) Wawancara: dengan pihak sekolah dan guru untuk memahami kriteria siswa berprestasi.
- b) Observasi: terhadap sistem penilaian dan pengelolaan data siswa di sekolah.
- c) Dokumentasi: pengumpulan data nilai rapor, kehadiran, kedisiplinan, prestasi lomba, dan kegiatan ekstrakurikuler

6. Analisis Data

Analisis data dilakukan secara kuantitatif dengan tahapan berikut:

- a) *Praproses Data (Preprocessing)*: Meliputi pembersihan data (*cleaning*), normalisasi, dan pelabelan (*labeling*) data siswa berprestasi dan tidak berprestasi
- b) *Pelatihan Model (Training)*: Data latih digunakan untuk membangun model klasifikasi *Random Forest*.
- c) *Pengujian Model (Testing)*: Data uji digunakan untuk mengevaluasi performa model menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.
- d) *Evaluasi Hasil*: Menilai efektivitas algoritma *Random Forest* dalam memprediksi siswa berprestasi dan membandingkan hasil prediksi dengan data aktual sekolah.

B. Instrumen Penelitian

Instrumen penelitian adalah alat atau cara yang digunakan untuk mengumpulkan data dan mengukur hasil dari penelitian tersebut. Dalam penelitian ini, instrumen yang digunakan mencakup perangkat keras, perangkat lunak, dataset, serta alat evaluasi hasil.

1. Perangkat Keras (*Hardware*)

Perangkat keras digunakan untuk menjalankan sistem klasifikasi dan melakukan proses pelatihan model *machine learning*. Spesifikasi perangkat keras yang digunakan adalah sebagai berikut:

- a) Prosesor: AMD Ryzen 3
- b) RAM: 8 GB DDR4
- c) Penyimpanan: SSD 256 GB
- d) GPU: AMD Radeon (TM) Graphics
- e) Sistem Komputer: Laptop berbasis arsitektur x64-based PC

Spesifikasi tersebut dipilih untuk memastikan proses pelatihan model *Random Forest* dan pengolahan dataset dapat berjalan optimal tanpa kendala performa

2. Perangkat Lunak (*Software*)

Perangkat lunak berperan penting dalam membangun, melatih, dan menguji model klasifikasi *Random Forest*. Software yang digunakan dalam penelitian ini meliputi:

- a. Sistem Operasi: Windows 11
- b. Bahasa Pemrograman: Python 3.12.1
- c. Library yang di gunakan:
 - 1) *scikit-learn*: untuk implementasi algoritma *Random Forest*
 - 2) *Pandas* dan *NumPy*: untuk pengolahan dan pembersihan data
 - 3) *Matplotlib* dan *Seaborn*: untuk visualisasi hasil evaluasi
- d. Tools Pendukung:
 - 1) *Jupyter Notebook*: untuk pelatihan dan pengujian model.

2) **Visual Studio Code (VS Code)**: untuk pengembangan antarmuka sistem berbasis web

3) **XAMPP**: untuk manajemen basis data *MySQL*

3. **Dataset**

Dataset yang digunakan dalam penelitian ini merupakan data asli siswa yang memuat beberapa variabel penilaian, yaitu nilai rata-rata rapor, kehadiran, kedisiplinan, prestasi lomba, dan keaktifan ekstrakurikuler. Data tersebut digunakan sebagai dasar dalam proses klasifikasi siswa berprestasi menggunakan algoritma *Random Forest*. Pada Tabel 3.1 ditampilkan 10 data siswa sebagai contoh dataset penelitian yang diambil dari dataset asli. Label Berprestasi (B) dan Tidak Berprestasi (TB) diperoleh berdasarkan hasil konversi nilai setiap variabel ke dalam skor numerik sesuai rubrik penilaian yang telah ditentukan.

Tabel 3. 1 *Dataset* berdasarkan data asli

Nama siswa	Nilai rapor	Kehadiran	Kedisiplinan	Prestasi lomba	Ekstrakurikuler
Ahmad S. F.	82	90	2	Kota	Aktif
Alfan A.	88	92	1	Nasional	Sangat aktif
Alysha A. P.	83	86	0	Kota	Aktif
Angger B. P.	78	85	2	Kota	Tidak aktif
Arhan A.	95	78	2	Kota	Aktif
Aulia M. S.	85	90	0	Provinsi	Aktif

4. Kuesioner dan Wawancara
Metode wawancara dilakukan untuk memperoleh informasi secara

Nama siswa	Nilai rapor	Kehadiran	Kedisiplinan	Prestasi lomba	Ekstensi
Azizah A. S.	84	85	2	Nasional	Sangat aktif
Chalita H. P.	95	95	0	Kota	Aktif
Cinta D. S.	92	86	0	Nasional	Aktif
Eksa R. P.	80	82	2	Kota	Aktif

langsung mengenai kriteria yang digunakan pihak sekolah dalam menentukan siswa berprestasi. Wawancara dilaksanakan dengan melibatkan guru dan staf akademik yang memahami proses penilaian serta perkembangan akademik siswa. Informasi yang dikumpulkan mencakup indikator penilaian, mekanisme pengolahan nilai, tingkat kehadiran, sikap, kedisiplinan, serta faktor lain yang menjadi pertimbangan dalam menetapkan siswa berprestasi. Beberapa pertanyaan yang diajukan kepada guru dan staf akademik meliputi:

- Indikator apa saja yang digunakan dalam menentukan siswa berprestasi.
- Bagaimana proses penilaian dilakukan saat ini, apakah secara manual atau secara sistematis.
- Faktor-faktor non-akademik apa yang paling berpengaruh terhadap prestasi siswa.

5. Analisis Data

Untuk mengevaluasi hasil penelitian, digunakan beberapa alat ukur (*evaluation metrics*) dan pengujian sistem, yaitu:

- Pengujian Fungsional:** Dilakukan untuk memastikan sistem bekerja sesuai dengan rancangan, menggunakan *Black Box Testing* (mengamati keluaran dari fungsi *input* tertentu tanpa memeriksa *code internal*).

- 2) Pengujian Non-Fungsional: Dilakukan untuk menguji performa model klasifikasi menggunakan metrik evaluasi berikut: *Accuracy, Precision, Recall, F1-Score, Confusion Matrix*.

6. Penerapan Instrumen

- a) Perangkat keras dan lunak digunakan untuk melatih model, menjalankan eksperimen, serta menampilkan hasil klasifikasi melalui sistem berbasis web.
- b) Dataset diproses menggunakan *Python dan Scikit-learn*, kemudian dibagi menjadi data latih dan data uji untuk validasi model.

C. Tempat dan Jadwal Penelitian

1. Tempat Penelitian

Tempat penelitian ditetapkan berdasarkan kebutuhan penelitian untuk memperoleh data akademik dan non-akademik siswa yang akan digunakan dalam pemodelan klasifikasi menggunakan algoritma *Random Forest*. Lokasi penelitian dipilih di SMP Negeri 4 Kediri Jl. Penanggungan No. 6, Kelurahan Bandar Lor, Kecamatan Mojoroto, Kota Kediri, Provinsi Jawa Timur. Jalan: Jl. Penanggungan No. 6.

Pemilihan lokasi ini didasarkan pada beberapa pertimbangan berikut:

- a) Ketersediaan data yang relevan, yaitu data akademik dan non-akademik siswa kelas VIII tahun ajaran 2025–2026, yang menjadi dasar pembangunan model prediksi siswa berprestasi.
- b) Kesesuaian dengan tujuan penelitian, yaitu mengembangkan sistem klasifikasi yang dapat membantu sekolah mengidentifikasi siswa berprestasi secara objektif.
- c) Dukungan institusi, berupa izin pengambilan data dan kerja sama dalam proses wawancara serta dokumentasi.
- d) Implementasi sistem, karena sistem prediksi yang dibangun dapat diujicobakan langsung di sekolah untuk melihat manfaat praktisnya.

2. Jadwal Penelitian

Masa penelitian berlangsung selama 6 bulan, dimulai pada Januari hingga Juni 2026. Jadwal penelitian disusun berdasarkan tahapan metodologi penelitian yang telah dijelaskan pada Bab III dan diselaraskan dengan alur

No	Kegiatan	Bulan ke 1				Bulan ke 2				Bulan ke 3				Bulan ke 4				Bulan ke 5				Bulan ke 6			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Studi Literatur & Analisis Kebutuhan																								
2	Pengumpulan Dataset																								
3	implementasi Sistem Prediksi																								
4	Pengujian & Evaluasi Sistem																								
5	Laporan Akhir & Revisi																								

kerja penelitian yang memanfaatkan algoritma Random Forest untuk klasifikasi siswa berprestasi

Tabel 3. 2 Jadwal Penelitian

D. Desain Penelitian atau Subjek Penelitian

Analisis kebutuhan sistem bertujuan untuk menjelaskan secara rinci apa yang diharapkan dari sistem prediksi siswa berprestasi yang akan dibangun. Sistem ini berfungsi mengolah data akademik dan non-akademik siswa menggunakan algoritma *Random Forest* sehingga menghasilkan prediksi yang objektif, akurat, dan dapat membantu pihak sekolah dalam proses identifikasi siswa berprestasi.

1. Kebutuhan fungsional

Kebutuhan fungsional adalah fitur-fitur utama yang harus dimiliki sistem agar dapat berfungsi sesuai tujuan. Fitur tersebut meliputi:

- a) Sistem dapat menerima input data siswa berupa nilai rapor, kehadiran, kedisiplinan, prestasi lomba, dan partisipasi ekstrakurikuler.
- b) Sistem dapat melakukan *preprocessing* data (normalisasi, pembersihan data, dan labeling).
- c) Sistem dapat melakukan pelatihan model klasifikasi menggunakan algoritma *Random Forest*.
- d) Sistem dapat memprediksi kategori prestasi siswa (berprestasi / tidak berprestasi).
- e) Sistem dapat menampilkan hasil prediksi secara jelas dalam bentuk tabel atau grafik.

- f) Sistem dapat menyimpan hasil prediksi untuk proses evaluasi dan pelaporan.

2. Kebutuhan non-fungsional

Kebutuhan non-fungsional menggambarkan kualitas sistem yang harus dipenuhi, yaitu:

- a) Keamanan: Data siswa harus terlindungi dan hanya dapat diakses oleh pihak berwenang.
- b) Kinerja: Sistem harus mampu memproses data siswa dalam jumlah besar tanpa mengurangi kecepatan dan akurasi prediksi.
- c) Keandalan: Sistem harus berfungsi dengan baik tanpa error selama proses pengolahan data dan prediksi.
- d) *Usability* (Kemudahan penggunaan): Antarmuka sistem harus mudah digunakan oleh guru dan staf sekolah.

3. Kebutuhan bisnis

Sistem prediksi siswa berprestasi mampu:

- a) Membantu sekolah mengidentifikasi siswa yang memiliki potensi tinggi secara lebih objektif.
- b) Mendukung proses pengambilan keputusan dalam pembinaan dan penghargaan siswa.
- c) Mengoptimalkan pemanfaatan data akademik dan non-akademik yang selama ini belum dikelola secara sistematis.

4. Kebutuhan Pengguna

Pengguna sistem terdiri dari:

- a) Guru / Wali Kelas: Menginput data siswa dan melihat hasil prediksi.
- b) Bagian Kurikulum: Menggunakan hasil prediksi untuk pengambilan keputusan.
- c) Admin Sistem: Mengelola data, memelihara sistem, dan memastikan model berjalan dengan baik.

Kebutuhan pengguna meliputi:

- a) Sistem mudah diakses kapan saja
- b) Hasil prediksi mudah dipahami
- c) Tampilan menarik, interaktif, dan informatif

5. Kebutuhan Data

a. Definisi Variabel Penelitian

1) Nilai rapor

Nilai rapor merupakan nilai rata-rata rapor akhir semester yang diperoleh siswa berdasarkan seluruh mata pelajaran yang tercantum dalam rapor. Nilai ini digunakan untuk menggambarkan tingkat pencapaian akademik siswa selama satu periode pembelajaran dan diklasifikasikan ke dalam beberapa kriteria penilaian. Nilai rapor 85–100 dikategorikan Sangat Baik dengan nilai 4, nilai 75–84 dikategorikan Baik dengan nilai 3, nilai 65–74 dikategorikan Cukup dengan nilai 2, sedangkan nilai di bawah 65 dikategorikan Kurang dengan nilai 1.

2) Kehadiran

Kehadiran siswa adalah tingkat kehadiran siswa dalam mengikuti kegiatan pembelajaran selama satu semester yang diukur dalam bentuk persentase terhadap total hari efektif belajar. Tingkat kehadiran tersebut dikelompokkan ke dalam beberapa kriteria penilaian berdasarkan persentase kehadiran siswa. Kehadiran $\geq 95\%$ dikategorikan Sangat Baik dengan nilai 4, kehadiran 90%–94% dikategorikan Baik dengan nilai 3, kehadiran 80%–89% dikategorikan Cukup dengan nilai 2, dan kehadiran $< 80\%$ dikategorikan Kurang dengan nilai 1.

3) Kedisiplinan

Kedisiplinan siswa adalah tingkat kepatuhan siswa terhadap tata tertib sekolah yang diukur berdasarkan jumlah pelanggaran yang dilakukan selama satu semester. Jumlah pelanggaran tersebut diklasifikasikan ke dalam beberapa kriteria penilaian sesuai tingkat

222

157

227

kedisiplinan siswa. Siswa dengan 0 pelanggaran dikategorikan Sangat Baik dengan nilai 4, siswa dengan 1-2 pelanggaran dikategorikan Baik dengan nilai 3, siswa dengan 3-4 pelanggaran dikategorikan Cukup dengan nilai 2, sedangkan siswa dengan lebih dari 4 pelanggaran dikategorikan Kurang dengan nilai 1.

4) Prestasi Lomba

Prestasi lomba adalah capaian siswa dalam perlombaan akademik maupun nonakademik yang dinilai berdasarkan tingkat lomba tertinggi dan peringkat juara yang pernah diraih. Prestasi pada tingkat nasional dengan peringkat Juara (1, 2, 3) dikategorikan Sangat Baik dengan nilai 4, prestasi pada tingkat provinsi dengan peringkat Juara (1, 2, 3) dikategorikan Baik dengan nilai 3, dan prestasi pada tingkat kabupaten/kota dengan peringkat Juara (1, 2, 3) dikategorikan Cukup dengan nilai 2. Siswa yang tidak memiliki prestasi lomba dikategorikan Kurang dengan nilai 1.

5) Keaktifan kegiatan ekstrakurikuler

Keaktifan kegiatan ekstrakurikuler dinilai berdasarkan partisipasi siswa selama satu semester, yaitu sekitar 6 bulan atau $\pm 18-20$ minggu efektif, yang dilihat dari kehadiran dan keterlibatan siswa dalam kegiatan mingguan. Siswa dikategorikan Sangat Baik dengan nilai 4 apabila selalu hadir dan aktif dalam kegiatan ekstrakurikuler setiap minggu selama satu semester, dikategorikan Baik dengan nilai 3 apabila rutin hadir minimal satu kali dalam satu minggu, dikategorikan Cukup dengan nilai 2 apabila kehadirannya tidak rutin, dan dikategorikan Kurang dengan nilai 1 apabila jarang atau tidak pernah mengikuti kegiatan ekstrakurikuler

Setelah dilakukan pendefinisian setiap variabel penelitian, tahap selanjutnya adalah penyusunan rubrik penilaian. Rubrik

≡ 52

≡ 52

≡ 52

penilaian digunakan untuk mengonversi data mentah yang masih berbentuk skala awal ke dalam bentuk kategori penilaian dan kode numerik. Proses ini bertujuan untuk menyeragamkan data sehingga dapat digunakan sebagai input dalam algoritma *Random Forest* pada tahap pengolahan data selanjutnya.

Tabel 3. 3 Rubrik Penilaian

Nama Variabel	Skala Awal	Kriteria	Kategori
Nilai rapor	Nilai (0-100)	85-100 75-84 65-74 < 65	SB (Sangat Baik) B (Baik) C (Cukup) K (Kurang)
Kehadiran	Persentase (%)	≥ 95% 90% – 94% 80% – 89% < 80%	SB (Sangat Baik) B (Baik) C (Cukup) K (Kurang)
Kedisiplinan	Jumlah pelanggaran	0 1-2 3-4 > 4	SB (Sangat Baik) B (Baik) C (Cukup) K (Kurang)
Prestasi lomba	Tingkat lomba	Nasional Provinsi Kabupaten/ Kota Tidak ada	SB (Sangat Baik) B (Baik) C (Cukup) K (Kurang)
Ekstrakurikuler	Tingkat keaktifan	Sangat aktif Aktif Cukup aktif Tidak aktif	SB (Sangat Baik) B (Baik) C (Cukup) K (Kurang)

b. Data Master

1) Data siswa
(nama, kelas,
nomor induk
siswa)

2) Data

ekstrakurikuler

c. Data Analitik

1) Hasil proses klasifikasi *Random Forest*

2) Metrik evaluasi model (*accuracy, precision, recall, F1-score*)

Semua data yang digunakan merupakan data terstruktur dengan format tabel yang dapat diolah oleh algoritma *random forest*

E. Prosedur Penelitian

Bagian ini menjelaskan prosedur penelitian yang diterapkan dalam penelitian, sesuai dengan desain penelitian dan jadwal yang telah ditetapkan sebelumnya. Prosedur penelitian merupakan serangkaian aktivitas sistematis yang dilakukan untuk mencapai tujuan penelitian, yaitu membangun model klasifikasi siswa berprestasi menggunakan algoritma *Random Forest* dan mengimplementasikannya dalam bentuk sistem prediksi.

Dalam penelitian ini, prosedur penelitian mengacu pada metode *Software Development Life Cycle (SDLC) model Waterfall*, karena proses pengembangan dilakukan secara bertahap, terstruktur, dan setiap tahap harus diselesaikan sebelum masuk ke tahap berikutnya. Adapun tahapan prosedur penelitian adalah sebagai berikut:

1. Analisis Kebutuhan (*Requirement Analysis*)

Tahap ini berfokus pada pengumpulan data dan analisis kebutuhan sistem.

Kegiatan yang dilakukan meliputi:

- Mengidentifikasi kebutuhan fungsional dan non-fungsional sistem.
- Mengumpulkan data akademik dan non-akademik siswa SMP Negeri 4 Kediri, seperti nilai rapor, kehadiran, kedisiplinan, prestasi lomba, serta partisipasi ekstrakurikuler.
- Menentukan variabel prediktor (*independen*) dan label (*dependen*) untuk model klasifikasi.
- Menentukan kebutuhan pengguna seperti guru, wali kelas, dan pihak kurikulum.

2. Perancangan Sistem dan Model (*Design*)

Tahap perancangan dilakukan untuk mendesain alur kerja sistem prediksi dan model *machine learning*. Kegiatan yang dilakukan:

- a. Mendesain arsitektur sistem yang mencakup proses *input* data, *preprocessing*, *training* model, evaluasi, dan *output* hasil prediksi.
- b. Mendesain struktur *database* untuk menyimpan data siswa dan hasil prediksi.
- c. Mendesain antarmuka pengguna (UI) agar mudah digunakan oleh guru dan staf sekolah.
- d. Mendesain model klasifikasi berbasis algoritma *Random Forest*, termasuk pemilihan parameter dan konfigurasi model.

Output tahap ini berupa rancangan sistem, *flowchart*, diagram alur proses, serta desain model *machine learning*.

3. Implementasi Sistem dan Model (*Implementation*)

Pada tahap ini dilakukan pembangunan sistem dan implementasi algoritma *Random Forest*. Kegiatan yang dilakukan:

- a. Melakukan *preprocessing data*: pembersihan data, normalisasi, *encoding*, dan pembagian data menjadi data latih dan data uji.
- b. Mengimplementasikan model *Random Forest* menggunakan *Python* dan *library machine learning* (misalnya *scikit-learn*).
- c. Mengembangkan aplikasi prediksi siswa berprestasi sesuai desain pada tahap sebelumnya.
- d. Mengintegrasikan model dengan sistem aplikasi.

Tahap ini menghasilkan sistem prediksi dan model *machine learning* yang siap diuji.

4. Pengujian Sistem dan Model (*Testing*)

Tahap pengujian dilakukan untuk mengevaluasi performa model dan memastikan sistem bekerja sesuai tujuan. Kegiatan meliputi:

- a. Melakukan pengujian model dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.
- b. Menguji sistem menggunakan *black box testing* untuk memastikan setiap fungsi berjalan dengan baik.
- c. Menguji keakuratan prediksi sistem menggunakan data uji.
- d. Melakukan evaluasi kesesuaian sistem dengan kebutuhan pengguna.

5. Pemeliharaan dan Evaluasi (*Maintenance & Evaluation*)

Tahap ini dilakukan untuk memastikan sistem dapat digunakan secara berkelanjutan. Kegiatan meliputi:

- a. Memperbaiki bug atau kesalahan sistem yang ditemukan setelah pengujian.
- b. Melakukan penyempurnaan model jika diperlukan, seperti *tuning* parameter atau penambahan dataset.
- c. Mengevaluasi manfaat sistem bagi sekolah dalam proses identifikasi siswa berprestasi.
- d. Tahap ini memastikan sistem tetap relevan dan akurat saat digunakan dalam lingkungan sekolah.

D. Teknik Analisis Data

Teknik analisis data dan perancangan sistem pada penelitian ini menggunakan pendekatan *Structured System Analysis and Design (SSAD)* karena sistem berfokus pada alur proses dan aliran data dalam pengolahan data akademik dan non-akademik siswa untuk menghasilkan prediksi siswa berprestasi. Pendekatan ini dimodelkan secara terstruktur menggunakan *Data Flow Diagram (DFD)*, diagram dekomposisi, *flowchart*, serta *Entity Relationship Diagram (ERD)* sebagai perancangan struktur dan relasi data sistem.

1. Desain Sistem

- a. Arsitektur Sistem (*System Architecture*)

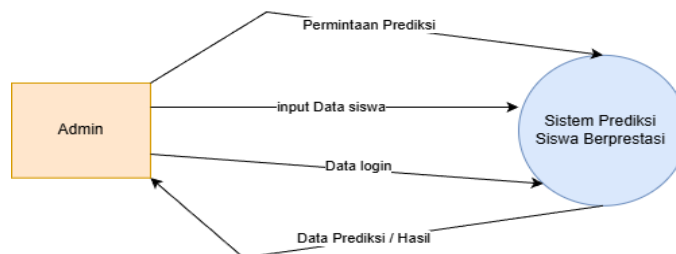
Arsitektur sistem menjelaskan alur proses mulai dari pengumpulan data siswa, pengolahan, pelatihan model *Random Forest*, hingga menghasilkan *output* berupa label prediksi “Berprestasi” atau “Tidak Berprestasi”.

Komponen arsitektur meliputi:

- 1) Input Data: nilai rapor, kehadiran, kedisiplinan, prestasi lomba, dan ekstrakurikuler.
- 2) Pra-pemrosesan Data: normalisasi, *encoding*, pembersihan data.
- 3) Model *Machine Learning: Random Forest Classifier*.
- 4) Evaluasi Model: *confusion matrix*, *accuracy*, *precision*, *recall*, *F1-score*.
- 5) *Output*: Hasil prediksi siswa berprestasi.

b. DFD level 0

DFD Level 0 menunjukkan gambaran umum Sistem Prediksi Siswa Berprestasi yang berinteraksi dengan Admin sebagai entitas eksternal. Admin

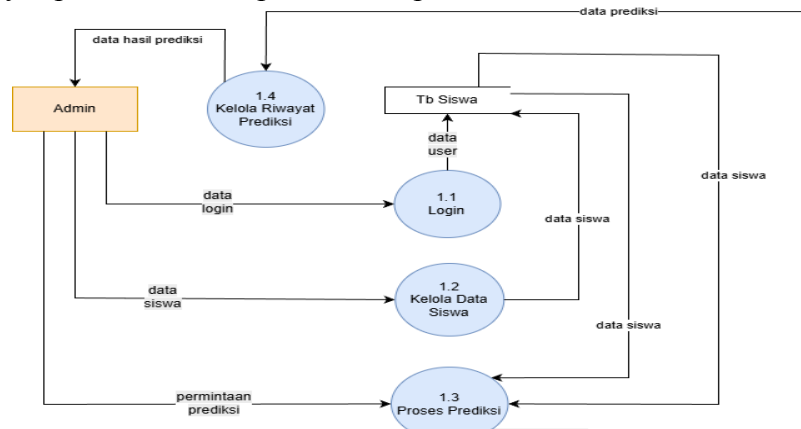


memasukkan data login, data siswa, serta permintaan prediksi ke dalam sistem. Sistem kemudian mengolah data tersebut dan menghasilkan *output* berupa hasil prediksi siswa berprestasi yang disampaikan kembali kepada Admin.

Gambar 3. 1 DFD level 0

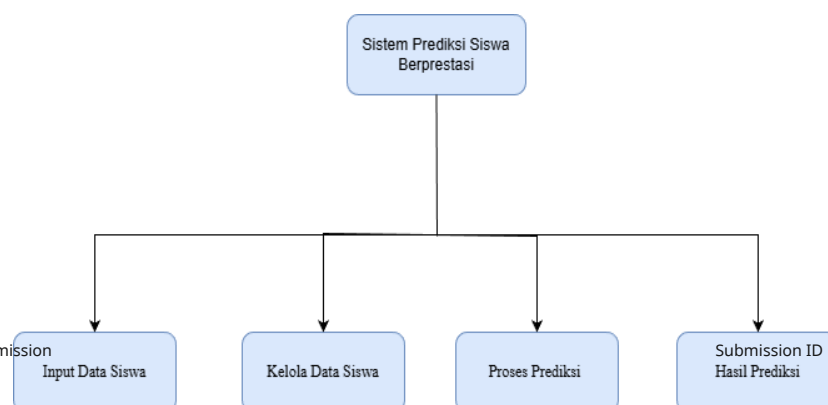
c. DFD level 1

DFD Level 1 merupakan penguraian dari DFD Level 0 yang menjelaskan proses utama Sistem Prediksi Siswa Berprestasi menjadi beberapa subproses. Sistem terdiri dari proses login (1.1) yang memverifikasi data pengguna, proses pengelolaan data siswa (1.2) untuk menyimpan dan memperbarui data siswa, proses prediksi (1.3) yang mengolah data siswa berdasarkan permintaan prediksi dari admin, serta proses pengelolaan riwayat prediksi (1.4) untuk menyimpan dan menampilkan hasil prediksi.



Gambar 3. 2 DFD level 1

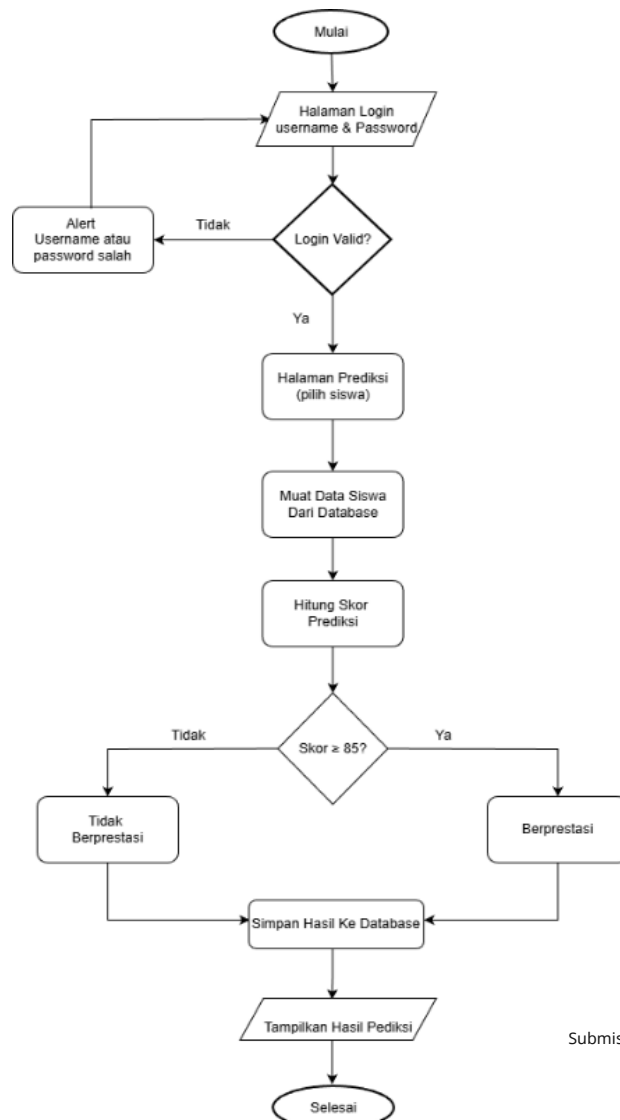
Diagram tersebut menggambarkan arsitektur sistem prediksi siswa berprestasi. Proses dimulai dari tahap *input* data siswa sebagai data awal sistem. Data kemudian melalui tahap *preprocessing* untuk menyiapkan data yang layak diproses. Selanjutnya data diolah menggunakan model *Random Forest* pada tahap pemodelan. Proses prediksi dilakukan dan hasilnya ditampilkan sebagai *output* sistem



Gambar 3. 3 Diagram Dekomposisi

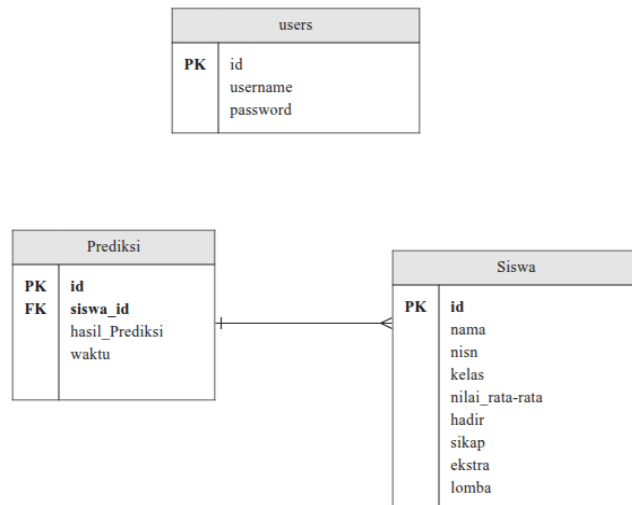
d. Flowchart

Flowchart sistem prediksi siswa berprestasi dimulai dari proses login pengguna. Setelah berhasil *login*, sistem mengambil data siswa, menghitung skor prediksi, dan mengklasifikasikan siswa sebagai berprestasi jika skor ≥ 85 atau tidak berprestasi jika kurang dari 85.



Gambar 3. 4 Flowchart Sistem Prediksi

- e. Desain Database Entity Relationship Diagram (ERD) pada sistem prediksi siswa berprestasi terdiri dari tiga entitas, yaitu Siswa, Prediksi, dan Users. Entitas Siswa menyimpan data akademik dan non-akademik siswa, sedangkan entitas. Prediksi menyimpan hasil klasifikasi yang berelasi **one-to-many (1:N)** dengan entitas Siswa, **di mana satu siswa dapat memiliki** beberapa riwayat prediksi. Entitas Users berfungsi untuk autentikasi pengguna sistem dan tidak memiliki relasi langsung dengan entitas lain.

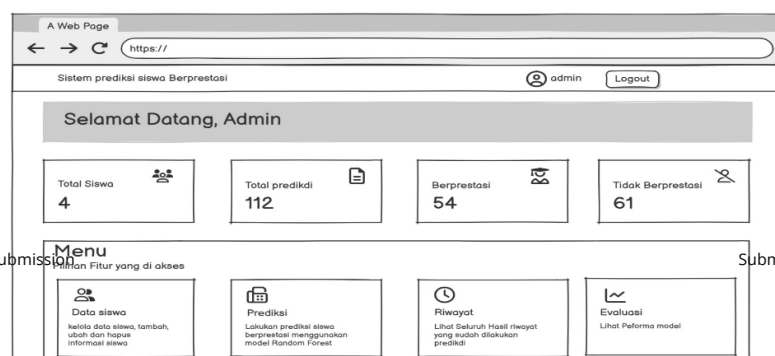


Gambar 3. 5 Entity Relationship Diagram

2. Desain Antarmuka Pengguna (UI)

Sistem prediksi akan ditampilkan dalam bentuk web sederhana.

Desain antarmuka mencakup:



Gambar 3. 6 Halaman Dashboard

Form *dashboard* sistem prediksi siswa berprestasi memungkinkan admin mengelola data siswa, menjalankan prediksi menggunakan *Random Forest*, meninjau riwayat prediksi, dan melihat performa model, sehingga proses pengelolaan dan penilaian prestasi siswa menjadi cepat, jelas, dan terstruktur.

The screenshot shows a web browser window with the URL 'https://'. The page title is 'Sistem prediksi siswa Berprestasi'. In the top right corner, there is a user profile icon labeled 'admin' and a 'Logout' button. The main heading is 'Tambah Data Siswa' with the subtext 'Masukkan data siswa'. The form contains two columns of input fields: NISN, Nama, Kelas, Nilai Rapor, Kehadiran, Kedisiplinan, Prestasi Lomba, and Ekstrakurikuler. At the bottom of the form, there are two buttons: 'Simpan' (Save) and 'Kembali' (Back).

Gambar 3. 7 Form tambah data siswa

Form “Tambah Data Siswa” digunakan untuk memasukkan informasi siswa baru ke dalam sistem. Admin dapat mengisi NISN, nama, kelas, nilai rapor, kehadiran, kedisiplinan, prestasi lomba, dan keaktifan ekstrakurikuler, lalu menekan tombol Simpan untuk menyimpan data ke database. Form ini memastikan data siswa lengkap tersedia sebelum dijalankan pada halaman prediksi.

The screenshot shows a web browser window with the URL 'https://'. The page title is 'Sistem prediksi siswa Berprestasi'. In the top right corner, there is a user profile icon labeled 'admin' and a 'Logout' button. The main heading is 'Prediksi Siswa Berprestasi'. Below the heading, there is a section titled 'Form Prediksi' with the subtext 'Pilih Siswa Yang Ingin Di Prediksi'. It features a search bar labeled 'Pilih Siswa' with a 'search' button and a 'Prediksi sekarang' button.

Gambar 3. 8 Halaman Prediksi

Halaman “Prediksi Siswa Berprestasi” digunakan untuk menjalankan model Random Forest dan menentukan kategori siswa. Admin dapat memilih siswa secara individu melalui kolom pencarian atau memilih semua siswa sekaligus, kemudian menekan tombol Prediksi sekarang untuk melihat hasil klasifikasi apakah siswa berprestasi atau tidak.

3. Proses Penyelesaian Masalah

Pada bagian ini dijelaskan proses penyelesaian masalah menggunakan algoritma *Random Forest* untuk mengklasifikasikan siswa ke dalam kategori Berprestasi dan Tidak Berprestasi.

a. Data Input dari Dataset Asli

Data input yang digunakan dalam penelitian ini berasal dari dataset asli siswa. Pada bagian ini ditampilkan 10 data siswa sebagai contoh data yang digunakan untuk menjelaskan proses perhitungan algoritma *Random Forest*.

Tabel 3. 4 *Dataset* siswa

Keterangan :
Berprestasi (B): 6
Tidak Berprestasi: (TB): 4
b. Konversi Data ke Skor Numerik
Sebelum diproses

Nama siswa	Nilai rapor	Kehadiran	Kedisiplinan	Prestasi lomba	Ekstra
Ahmad S. F.	82	90	2	Kota	Aktif
Alfan A.	88	92	1	Nasional	Sangat aktif
Alysha A. P.	83	86	0	Kota	Aktif
Angger B. P.	78	85	2	Kota	Tidak aktif
Arhan A.	95	78	2	Kota	Aktif
Aulia M. S.	85	90	0	Provinsi	Aktif
Azizah A. S.	84	85	2	Nasional	Sangat aktif
Chalita H. P.	95	95	0	Kota	Aktif
Cinta D. S.	92	86	0	Nasional	Aktif
Eksa R. P.	80	82	2	Kota	Aktif

menggunakan algoritma *Random Forest*, data asli dikonversi terlebih dahulu ke dalam bentuk numerik. Proses konversi dilakukan berdasarkan rubrik penilaian yang telah ditentukan.

Tabel 3. 5 Rubrik Konversi Data ke Skor Numerik

Variabel	Kriteria	Skor
Nilai rapor	85–100	4
Nilai rapor	75–84	3
Nilai rapor	65–74	2
Nilai rapor	< 65	1
Kehadiran	≥ 95%	4
Kehadiran	90%–94%	3
Kehadiran	80%–89%	2
Kehadiran	< 80%	1
Kedisiplinan	0 pelanggaran	4
Kedisiplinan	1–2 pelanggaran	3
Kedisiplinan	3–4 pelanggaran	2
Kedisiplinan	> 4 pelanggaran	1
Prestasi lomba	Nasional	4
Prestasi lomba	Provinsi	3
Prestasi lomba	Kabupaten/Kota	2
Prestasi lomba	Tidak ada	1
Ekstrakurikuler	Sangat aktif	4
Ekstrakurikuler	Aktif	3
Ekstrakurikuler	Cukup aktif	2

Ekstrakurikuler	Tidak aktif	1
-----------------	-------------	---

Rumus total skor:

$$Total\ Skor = Skor\ Rapor + Skor\ Kehadiran + Skor\ Kedisiplinan + Skor\ Prestasi + Skor\ E$$

Kriteria penentuan label:

Jika Total Skor ≥ 15 , maka Label = B *Jika Total Skor < 15 , maka Label = TB*

c. Hasil Koversi Data

Berdasarkan rubrik penilaian, data asli kemudian dikonversi menjadi skor numerik. Hasil konversi data ditampilkan pada tabel berikut.

Tabel 3. 6 Hasil Konversi Data Sampel

	Nama siswa	Nilai rapor	Kehadiran	Kedisiplinan	Prestasi lomba	Ekstra	T
Distribusi data pada contoh tersebut adalah:	Ahmad S. F.	3	3	3	2	3	14
	Alfan A.	4	3	3	4	4	18
	Alysha A. P.	3	2	2	4	3	14
	Angger B. P.	3	2	3	2	1	11
Berprestasi (B)=5	Arhan A.	4	1	3	2	3	13
	Aulia M. S.	4	3	4	3	2	17
Siswa Tidak Berprestasi (TB) = 5 siswa	Azizah A. S.	3	2	3	4	4	16
	Chalita H. P.	4	4	3	2	3	16
	Cinta D. S.	4	2	4	4	3	17
	Eksa R. P.	3	2	3	2	3	13

d. Rumus

Random Forest yang di gunakan

Algoritma *Random Forest* bekerja dengan membentuk beberapa *decision tree*. Setiap *decision tree* akan melakukan pemisahan data berdasarkan fitur tertentu. Pemilihan fitur terbaik dapat dilakukan dengan menghitung nilai *Gini Index*. Semakin kecil nilai *Gini*, maka semakin baik pemisahan data yang dihasilkan.

1. Rumus *Gini Index*

$$Gini(S) = 1 - \sum_{i=1}^k p_i^2$$

Keterangan :

$$Gini(S)$$

= nilai ketidakmurnian data pada node

$$p_i$$

= proporsi data pada kelas ke-i

$$k$$

= jumlah kelas

Pada penelitian ini terdapat dua kelas, yaitu:

B = Berprestasi

TB = Tidak Berprestasi

2. Rumus Gini Split

$$Gini_{split} = \frac{|S_L|}{|S|} Gini(S_L) + \frac{|S_R|}{|S|} Gini(S_R)$$

Keterangan:

$$S_L$$

= data pada cabang kiri

$$S_R$$

= data pada cabang kanan

$$S$$

= total data pada node

3. Rumus Voting Mayoritas

$$\hat{y} = \text{mode}(h_1(x), h_2(x), h_3(x), \dots, h_n(x))$$

Keterangan:

$\hat{y}$

= hasil prediksi akhir

$h_n(x_{0029})$

= hasil prediksi dari *decision tree* ke- n

e. Pembagian Data Training dan Testing

Pada contoh perhitungan manual ini, data sampel dibagi menjadi data *training* dan data *testing*. Data *training* digunakan untuk membentuk aturan *decision tree*, sedangkan data *testing* digunakan untuk melihat hasil prediksi dari model

Pembagian data yang digunakan adalah:

Data *training* = 8 data

Data *testing* = 2 data

Tabel 3. 7 Data Training

Nama siswa	Nilai rapor	Kehadiran	Kedisplinan	Prestasi lomba	Ekstra
Ahmad S. F.	3	3	3	2	3
Alfan A.	4	3	3	4	4
Alysha A. P.	3	2	2	4	3
Nama siswa	Nilai rapor	Kehadiran	Kedisplina	Prestasi lomba	Ekstra
Arhan A.	4	1	3	2	3
Cinta D.	4	2	4	4	3
Aulia M. S. S.	4	3	4	3	2
Azizah A. S.	3	2	3	4	4
Eksa R. P.	3	2	3	2	3
Chalita H. P.	4	4	3	2	3

Distribusi data *training*:
B = 4 siswa
TB = 4 siswa

Tabel 3. 8 Data Testing

f. Perhitungan

Gini Node Root

Pada data *training* terdapat:

B = 4

$$TB = 4$$

$$\text{Total data} = 8$$

Maka:

$$P_B = \frac{4}{8} = 0.5$$

$$P_{TB} = \frac{4}{8} = 0.5$$

$$Gini_{root} = 1 - (P_B^2 + P_{TB}^2)$$

$$Gini_{root} = 1 - (0.5^2 + 0.5^2)$$

$$Gini_{root} = 1 - (0.25 + 0.25)$$

$$Gini_{root} = 0.50$$

Jadi, nilai *Gini* pada *node* awal adalah:

$$Gini_{root} = 0.50$$

g. Pembentukan *Decision Tree* 1

Pada *Tree* 1, fitur acak yang digunakan adalah:

1. Prestasi lomba
2. Kehadiran

Split pertama menggunakan fitur:

$$Prestasi \geq 3$$

Tabel 3. 9 Data Cabang Kanan pada Split $Prestasi \geq 3$

Nama siswa	Prestasi	Label
Alfan A.	4	B
Aulia M. S.	3	B
Azizah A. S.	4	B

Jumlah data:

$$B=3$$

$$TB=0$$

$$Gini_R = 1 - \left(\left(\frac{3}{3} \right)^2 + \left(\frac{0}{3} \right)^2 \right)$$

$$Gini_R = 1 - (1 + 0)$$

$$Gini_R = 0$$

Tabel 3. 10 Data Cabang Ke kiri pada Split Prestasi ≥ 3

Nama siswa	Prestasi	Label
Ahmad S. F.	2	TB
Alysha A. P.	2	TB
Angger B. P.	2	TB
Arhan A.	2	TB
Chalita H. P.	2	B

Jumlah data

$$B = 1$$

$$TB = 4$$

$$Gini_L = 1 - \left(\left(\frac{1}{5} \right)^2 + \left(\frac{4}{5} \right)^2 \right)$$

$$Gini_L = 1 - (0.04 + 0.64)$$

$$Gini_L = 0.32$$

Maka nilai *Gini Split* adalah:

$$Gini_{split} = \frac{5}{8}(0.32) + \frac{3}{8}(0)$$

$$Gini_{split} = 0.20 + 0$$

$$Gini_{split} = 0.20$$

Karena nilai:

$$Gini_{split} = 0.20$$

lebih kecil dari:

$$Gini_{root} = 0.50$$

maka fitur prestasi lomba dapat digunakan sebagai pemisah data pada *Tree* 1.

h. Split Lanjutan pada *Tree* 1

Pada cabang kiri, yaitu data dengan:

$$Prestasi < 3$$

dilakukan split lanjutan menggunakan fitur kehadiran:

$$Kehadiran \geq 4$$

Tabel 3. 11 Data Cabang kanan: Kehadiran ≥ 4

Nama siswa	Kehadiran	Label
Chalita H. P.	4	B

Karena seluruh data pada cabang ini memiliki label B, maka:

$$Gini = 0$$

Tabel 3. 12 Data Cabang kiri: Kehadiran < 4

Nama siswa	Kehadiran	Label
Ahmad S. F.	3	TB
Alysha A. P.	2	TB

Angger B. P.	2	TB
Arhan A.	1	TB

Karena seluruh data pada cabang ini memiliki label TB, maka:

$$Gini = 0$$

Aturan Tree 1 yang terbentuk adalah:

- 1) Jika prestasi lomba ≥ 3 , maka hasil prediksi = B.
- 2) Jika prestasi lomba < 3 dan kehadiran ≥ 4 , maka hasil prediksi = B.
- 3) Jika prestasi lomba < 3 dan kehadiran < 4 , maka hasil prediksi = TB.

i. Pembentukan *Tree 2* dan *Tree 3* Secara Ringkas

Random Forest tidak hanya membentuk satu *decision tree*, tetapi beberapa *decision tree*. Pada contoh ini digunakan tiga *tree* untuk menjelaskan proses *voting* mayoritas.

Tree 2

Fitur acak yang digunakan:

- 1) Ekstrakurikuler
- 2) Nilai rapor
- 3) Kedisiplinan

Aturan sederhana Tree 2:

- 1) Jika ekstrakurikuler = 4, maka hasil prediksi = B.
- 2) Jika ekstrakurikuler < 4 , nilai rapor = 4, dan kedisiplinan = 4, maka hasil prediksi = B.
- 3) Selain kondisi tersebut, hasil prediksi = TB.

Tree 3

Fitur acak yang digunakan:

- 1) Nilai rapor
- 2) Prestasi lomba

Aturan sederhana *Tree* 3:

- 1) Jika nilai rapor = 4 dan prestasi lomba ≥ 3 , maka hasil prediksi = B.
- 2) Selain kondisi tersebut, hasil prediksi = TB.

j. Prediksi Data *Testing* Menggunakan *Voting* Mayoritas

Data hasil konversi:

Tabel 3. 13 Data Data *testing* 1: Cinta Denisa Sahra Aulia

Nama siswa	Nilai rapor	Kehadiran	Kedisiplinan	Prestasi lomba	Ek
Cinta D. S.	4	2	4	4	3
		Hasil			

Tabel 3. 14
Data Hasil prediksi dari
setiap *tree*

Tree 1	Prediksi
Tree 1	B
Tree 2	B
Tree 3	B

Voting mayoritas:

$$\hat{y} = \text{mode}(B, B, B)$$

$$\hat{y} = B$$

Jadi, Cinta D. S. diprediksi sebagai:

Berprestasi (B)

Tabel 3. 15 Data *testing* 2: Eksa Risky Purnomo

Nama siswa	Nilai rapor	Kehadiran	Kedisiplinan	Prestasi lomba	Ek
Eksa R. P.	3	2	3	2	3

Tabel 3. 16 Data Hasil prediksi dari setiap *tree*

Tree 1	Hasil Prediksi
Tree 1	TB
Tree 2	TB
Tree 3	TB

Voting mayoritas:

$$\hat{y} = \text{mode}(TB, TB, TB)$$

$$\hat{y} = TB$$

Jadi, Eksa R. P. diprediksi sebagai:

Tidak Berprestasi (TB)

k. Kesimpulan Proses Penyelesaian Masalah

Berdasarkan proses penyelesaian masalah menggunakan algoritma Random Forest, klasifikasi siswa dilakukan melalui beberapa tahapan, yaitu konversi data asli ke bentuk skor numerik, pembagian data *training* dan *testing*, perhitungan *Gini Index*, pembentukan beberapa *decision tree*, serta penentuan hasil akhir menggunakan *voting* mayoritas.

Dengan demikian, algoritma *Random Forest* dapat digunakan untuk membantu proses klasifikasi siswa berprestasi berdasarkan data akademik dan non-akademik secara lebih objektif, terukur, dan akurat

BAB IV

HASIL DAN PEMBAHASAN

A. Hasil dan Penelitian

Penelitian ini menghasilkan sebuah sistem prediksi siswa berprestasi berbasis web dengan menerapkan algoritma *Random Forest Classifier*. Sistem ini dibuat untuk membantu pihak sekolah dalam mengklasifikasikan siswa ke dalam kategori Berprestasi dan Tidak Berprestasi berdasarkan data akademik dan non-akademik siswa.

Data yang digunakan dalam sistem meliputi nilai rapor, kehadiran, kedisiplinan, prestasi lomba, dan ekstrakurikuler. Variabel tersebut digunakan sebagai fitur input dalam proses prediksi. Sistem dikembangkan menggunakan bahasa pemrograman *Python*, *framework Flask*, database *MySQL*, serta beberapa *library* pendukung seperti *pandas*, *scikit-learn*, *joblib*, dan *mysql.connector*.

Model *Random Forest* dilatih menggunakan data dari tabel *data_latih*. Setelah proses pelatihan selesai, model disimpan dalam file *model_rf.pkl*, sedangkan informasi hasil evaluasi model disimpan dalam file *model_info.json*. File model tersebut kemudian digunakan oleh sistem untuk melakukan prediksi terhadap data siswa yang tersimpan di database.

1. Implementasi Desain Sistem

Implementasi desain sistem merupakan proses penerapan rancangan yang telah dijelaskan pada Bab III ke dalam bentuk sistem berbasis web. Sistem prediksi siswa berprestasi terdiri dari beberapa modul utama, yaitu modul *login*, *dashboard*, *data siswa*, *input data siswa*, *import data CSV*, *preprocessing* dan konversi *data*, prediksi *siswa*, prediksi semua siswa, riwayat prediksi, dan evaluasi model.

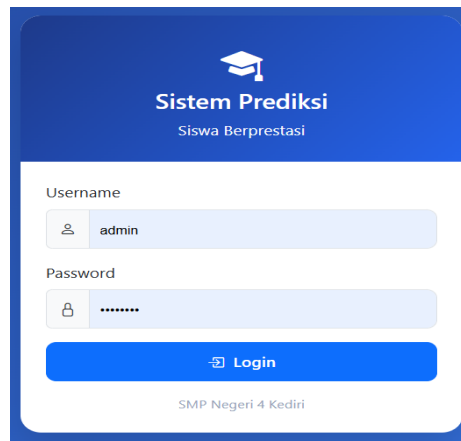
a) Implementasi Modul *Login*

Modul *login* merupakan halaman awal yang digunakan untuk membatasi akses pengguna ke dalam sistem. Pengguna harus memasukkan *username* dan *password* yang sesuai dengan data pada tabel *users*.

Pada sistem yang dibuat, proses autentikasi dilakukan dengan cara memeriksa data pengguna di database. *Password* pengguna tidak dibandingkan dalam bentuk teks biasa, tetapi diperiksa menggunakan fungsi

check_password_hash. Hal ini bertujuan untuk meningkatkan keamanan sistem agar password tidak mudah diketahui oleh pihak lain.

Apabila *username* dan *password* benar, sistem akan menyimpan data pengguna ke dalam *session* dan mengarahkan pengguna menuju halaman *dashboard*. Namun, apabila *username* atau *password* salah, sistem akan menampilkan pesan bahwa *username* atau *password* tidak sesuai.



Gambar 4. 1

Halaman Login

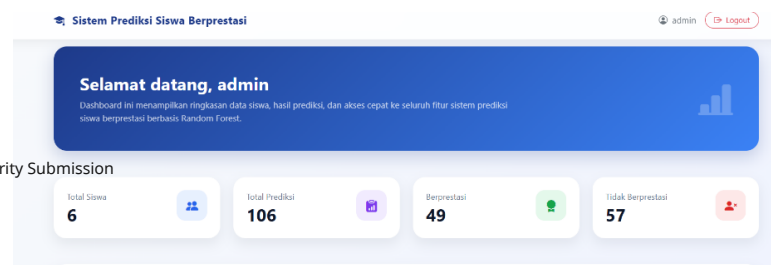
Gambar 4.1 menunjukkan halaman *login* yang digunakan oleh pengguna untuk masuk ke sistem. Halaman ini menjadi bagian penting karena seluruh fitur sistem hanya dapat diakses setelah pengguna berhasil *login*.

b) Implementasi Modul *Dashboard*

Modul *dashboard* merupakan halaman utama yang ditampilkan setelah pengguna berhasil *login*. Halaman ini berfungsi untuk menampilkan ringkasan informasi mengenai data yang terdapat di dalam sistem.

Informasi yang ditampilkan pada *dashboard* meliputi total data siswa, total hasil prediksi, jumlah siswa yang diprediksi berprestasi, dan jumlah siswa yang diprediksi tidak berprestasi. Data yang ditampilkan pada *dashboard* diperoleh dari tabel siswa dan tabel prediksi.

Dengan adanya *dashboard*, pengguna dapat memantau kondisi data dan hasil prediksi secara cepat tanpa harus membuka halaman lain satu per satu.



Gambar 4.2 Halaman Dashboard

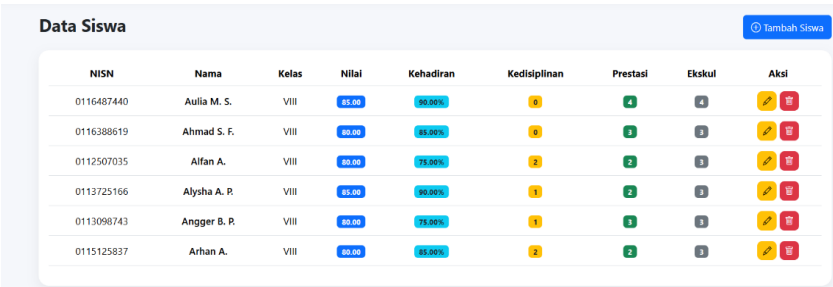
Gambar 4.2 menunjukkan tampilan *dashboard* sistem yang berisi ringkasan jumlah data siswa dan hasil prediksi. Halaman ini membantu pengguna mengetahui gambaran umum hasil klasifikasi siswa.

c) Implementasi Modul Data Siswa

Modul data siswa digunakan untuk menampilkan dan mengelola data siswa yang tersimpan di dalam database. Data siswa yang dikelola meliputi NISN, nama siswa, kelas, nilai rapor, kehadiran, kedisiplinan, prestasi lomba, dan ekstrakurikuler.

Modul ini menjadi bagian utama dalam sistem karena data siswa merupakan input yang digunakan untuk proses prediksi. Data yang tersimpan pada tabel siswa akan diambil oleh sistem, kemudian diproses melalui tahap konversi sebelum dimasukkan ke model *Random Forest*.

Pada modul ini, pengguna dapat melihat daftar siswa, menambahkan data siswa baru, mengubah data siswa, dan menghapus data siswa. Namun, data siswa yang sudah memiliki riwayat prediksi tidak dapat langsung dihapus agar hubungan data antara tabel siswa dan tabel prediksi tetap terjaga.



NISN	Nama	Kelas	Nilai	Kehadiran	Kedisiplinan	Prestasi	Ekskul	Aksi
0116487440	Aulia M. S.	VIII	85.00	90.90%	0	4	4	 
0116388619	Ahmad S. F.	VIII	80.00	85.90%	0	3	3	 
0112507035	Alfan A.	VIII	80.00	75.90%	2	2	3	 
0113725166	Alysha A. P.	VIII	85.00	90.90%	1	2	2	 
0113098743	Angger B. P.	VIII	80.00	75.90%	1	3	3	 
0115125937	Arhan A.	VIII	80.00	85.90%	2	2	3	 

Gambar 4.3 Halaman data siswa

Gambar 4.3 menunjukkan halaman data siswa yang berisi daftar siswa yang telah dimasukkan ke dalam sistem. Data pada halaman ini menjadi dasar dalam proses klasifikasi siswa berprestasi.

d) Implementasi Modul Input Data Siswa

Modul input data siswa digunakan untuk menambahkan data siswa baru ke dalam sistem. Pengguna mengisi form yang berisi data akademik dan non-akademik siswa.

Data yang dimasukkan pada form input meliputi NISN, nama, kelas, nilai rapor, kehadiran, kedisiplinan, prestasi lomba, dan ekstrakurikuler. Setelah pengguna menekan tombol simpan, sistem akan menyimpan data tersebut ke dalam tabel siswa. Modul input data siswa berperan penting karena kualitas data yang dimasukkan akan memengaruhi hasil prediksi. Apabila data yang dimasukkan tidak sesuai atau tidak lengkap, maka hasil prediksi juga dapat menjadi kurang akurat.

Gambar 4. 4 Form Input Data Siswa

Gambar 4.4 menunjukkan form input data siswa. Form ini digunakan untuk memasukkan data akademik dan non-akademik siswa sebelum dilakukan proses prediksi.

e) Implementasi Import Data CSV ke Database

Selain input data secara manual melalui form, sistem juga dapat menggunakan data dari file CSV sebagai sumber data awal. Proses import data

CSV dilakukan untuk memasukkan dataset siswa ke dalam database agar dapat digunakan sebagai data latih maupun data input sistem prediksi.

Pada proses ini, file CSV dibaca menggunakan *library pandas*. Data yang telah dibaca kemudian dimasukkan ke *database MySQL* melalui koneksi database. Data tersebut disimpan ke tabel yang sesuai, misalnya tabel *data_latih* untuk proses pelatihan model atau tabel *siswa* untuk data yang akan diprediksi. Proses *import data CSV* membantu mempercepat pengisian data dalam jumlah banyak. Dengan demikian, pengguna tidak perlu memasukkan data satu per satu melalui form apabila dataset telah tersedia dalam bentuk file CSV

```
import_data.py > konversi_prestasi_lomba
1  import pandas as pd
2  import mysql.connector
3
4  db = mysql.connector.connect(
5      host="localhost",
6      user="root",
7      password="",
8      database="db_siswa_rf"
9  )
10
11 df = pd.read_csv("data_sekolah_skripsi.csv")
12 print("Kolom CSV:", df.columns.tolist())
13
14 # Rapikan nama kolom
15 df.columns = df.columns.str.strip()
```

Gambar 4. 5 Pseudocode Import Data CSV ke Database

Gambar 4.5 menunjukkan potongan kode program yang digunakan untuk membaca dataset berformat CSV menggunakan *library pandas*, kemudian menyimpan data tersebut ke dalam *database MySQL*. Kode lengkap import data CSV dapat diletakkan pada bagian lampiran.

f) Implementasi *Preprocessing* dan Konversi Data

Sebelum data siswa diproses oleh algoritma *Random Forest*, sistem terlebih dahulu melakukan proses *preprocessing*. Salah satu tahap *preprocessing* yang dilakukan adalah mengonversi data mentah menjadi skor numerik.

Konversi data dilakukan karena model *machine learning* membutuhkan input berupa angka. Data nilai rapor, kehadiran, kedisiplinan, prestasi lomba, dan

35

ekstrakurikuler harus diubah menjadi skor agar dapat diproses oleh model *Random Forest*.

113

Pada sistem ini, skor yang digunakan berada pada rentang 1 sampai 4. Skor 4 menunjukkan kategori sangat baik, sedangkan skor 1 menunjukkan kategori kurang. Rubrik konversi setiap variabel ditampilkan pada Tabel 4.1 sampai Tabel 4.5.

121

Tabel 4. 1 Rubrik Konversi Nilai Rapor

Nilai Rapor	Skor
≥ 85	4
75–84	3
65–74	2
< 65	1

Tabel 4. 2 Rubrik Konversi Kehadiran

Kehadiran	Skor
$\geq 95\%$	4
90%–94%	3
80%–89%	2
$< 80\%$	1

Tabel 4. 3 Rubrik Konversi Kedisiplinan

Jumlah Pelanggaran	Skor
0 pelanggaran	4
1–2 pelanggaran	3
3–4 pelanggaran	2
> 4 pelanggaran	1

Tabel 4. 4 Rubrik Konversi Prestasi Lomba

Prestasi Lomba	Skor
Nasional	4
Provinsi	3
Kota	2
Tidak ada	1

Tabel 4. 5 Rubrik Konversi Ekstrakurikuler

Ekstrakurikuler	Skor
Sangat aktif	4
Aktif	3
Cukup aktif	2
Tidak aktif	1

Setelah seluruh data dikonversi ke dalam skor numerik, sistem membentuk data input dengan fitur nilai rapor, kehadiran, kedisiplinan, prestasi lomba, dan ekstrakurikuler. Data tersebut kemudian digunakan sebagai *input model Random Forest* untuk menghasilkan hasil prediksi.

g) Implementasi Modul Prediksi Semua Siswa

Modul prediksi semua siswa digunakan untuk melakukan prediksi terhadap seluruh data siswa yang tersimpan di *database*. Modul ini membantu pengguna agar tidak perlu melakukan prediksi satu per satu.

Tahapan proses prediksi semua siswa adalah sebagai berikut:

1. Sistem mengambil seluruh data siswa dari tabel siswa.
2. Sistem melakukan perulangan terhadap setiap data siswa.
3. Setiap data siswa dikonversi ke bentuk skor numerik.
4. Data dimasukkan ke model *Random Forest*.

5. Model menghasilkan prediksi untuk setiap siswa.
6. Hasil prediksi disimpan ke tabel prediksi.
7. Sistem menampilkan jumlah siswa berprestasi dan tidak berprestasi.

Modul ini dapat di gunakan untuk proses klasifikasi terhadap banyak siswa sekaligus. Dengan adanya fitur ini, proses identifikasi siswa berprestasi menjadi lebih cepat dan efisien.

Tabel Hasil Prediksi Semua Siswa

Nama	Kelas	Nilai	Kehadiran	Kedisiplinan	Prestasi	Ekskul	Hasil	Confidence
Ahmad S. F.	VIII	80.0	85.0%	0	3	3	Berprestasi	72.55%
Alfan A.	VIII	80.0	75.0%	2	2	3	Tidak Berprestasi	96.74%
Alysha A. R.	VIII	85.0	90.0%	1	2	3	Berprestasi	79.07%
Angger B. P.	VIII	80.0	75.0%	1	3	3	Tidak Berprestasi	50.17%
Arhan A.	VIII	80.0	85.0%	2	2	3	Tidak Berprestasi	94.13%
Aulia M. S.	VIII	85.0	90.0%	0	4	4	Berprestasi	100.0%

Gambar 4. 6 Halaman Semua Prediksi Siswa

Gambar 4.6 menunjukkan hasil prediksi semua siswa yang telah diproses oleh sistem. Hasil prediksi ditampilkan dalam bentuk daftar sehingga mudah dibaca oleh pengguna.

h) Implementasi Modul Riwayat Prediksi

Modul riwayat prediksi digunakan untuk menampilkan seluruh hasil prediksi yang pernah dilakukan oleh sistem. Data riwayat diperoleh dari tabel prediksi yang dihubungkan dengan tabel siswa.

Informasi yang ditampilkan pada halaman riwayat meliputi nama siswa, kelas, hasil prediksi, *confidence score*, dan waktu prediksi. Modul riwayat prediksi berfungsi sebagai dokumentasi hasil klasifikasi. Dengan adanya riwayat, pihak sekolah dapat melihat kembali hasil prediksi yang pernah dilakukan sebelumnya.

Riwayat Prediksi Cetak Riwayat

Nama	Kelas	Hasil	Confidence	Waktu	Aksi
Ahmad S. F.	VIII	Berprestasi	72.5500%	2026-05-19 02:23:33	
Arhan A.	VIII	Tidak Berprestasi	94.1300%	2026-05-19 02:23:07	
Angger B. P.	VIII	Tidak Berprestasi	50.1700%	2026-05-19 02:23:07	
Alysha A. P.	VIII	Berprestasi	79.0700%	2026-05-19 02:23:07	
Alfan A.	VIII	Tidak Berprestasi	96.7400%	2026-05-19 02:23:07	

Gambar 4. 7 Halaman Riwayat Prediksi

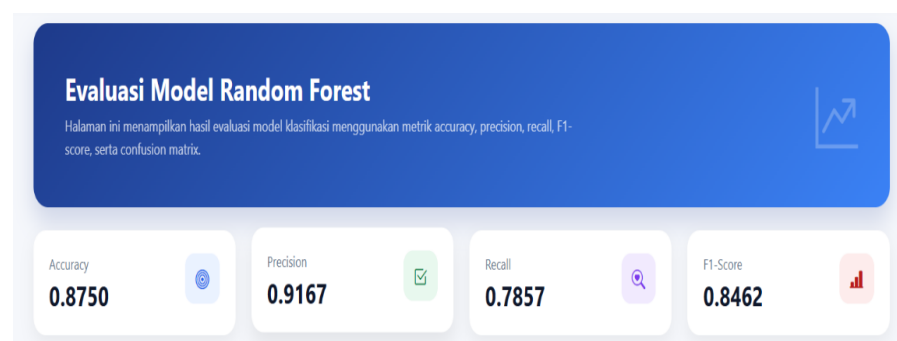
Gambar 4.7 menunjukkan halaman riwayat prediksi. Halaman ini menyimpan hasil prediksi siswa sehingga dapat digunakan sebagai bahan evaluasi dan dokumentasi.

i) Implementasi Modul Evaluasi Model

Modul evaluasi model digunakan untuk menampilkan hasil evaluasi performa algoritma *Random Forest*. Data evaluasi diperoleh dari file *model_info.json* yang dibuat setelah proses pelatihan model selesai.

Informasi yang ditampilkan pada halaman evaluasi model meliputi jumlah seluruh data, jumlah data training, jumlah data testing, *accuracy*, *precision*, *recall*, *F1-score*, *confusion matrix*, *feature importance*, jumlah pohon keputusan atau *n_estimators*, kedalaman maksimum pohon atau *max_depth*, serta daftar fitur yang digunakan.

Modul evaluasi model penting karena dapat menunjukkan tingkat keberhasilan algoritma *Random Forest* dalam melakukan klasifikasi siswa berprestasi dan tidak berprestasi.



Gambar 4. 8 Halaman Evaluasi Model

Gambar 4.8 menunjukkan halaman evaluasi model yang berisi hasil pengujian model *Random Forest*. Halaman ini digunakan untuk melihat nilai *accuracy*, *precision*, *recall*, *F1-score*, *confusion matrix*, dan *feature importance*.

Tabel 4. 6 Implementasi Modul Sistem

No	Modul	Fungsi	Hasil Implementasi
1	<i>Login</i>	Membatasi akses pengguna	Sistem dapat memvalidasi <i>username</i> dan <i>password</i>
2	<i>Dashboard</i>	Menampilkan ringkasan data	Sistem menampilkan total siswa, total prediksi, berprestasi, dan tidak berprestasi
3	Data Siswa	Mengelola data siswa	Sistem dapat menampilkan, menambah, mengedit, dan menghapus data siswa
4	<i>Input Data Siswa</i>	Memasukkan data akademik dan non-akademik siswa	Data siswa berhasil disimpan ke <i>database</i>
5	<i>Import CSV</i>	Memasukkan <i>dataset</i> dari <i>file CSV</i> ke <i>database</i>	<i>Data CSV</i> berhasil diproses dan disimpan ke <i>database</i>
6	<i>Preprocessing dan Konversi Data</i>	Mengubah data mentah menjadi skor numerik	Data berhasil dikonversi ke skor 1 sampai 4
7	Prediksi Siswa	Melakukan prediksi terhadap satu siswa	Sistem menghasilkan prediksi Berprestasi atau Tidak Berprestasi
8	Prediksi Semua Siswa	Melakukan prediksi seluruh data siswa	Sistem menghasilkan prediksi untuk semua siswa
9	Riwayat Prediksi	Menyimpan dan menampilkan hasil prediksi	Sistem menampilkan riwayat hasil prediksi
10	Evaluasi Model	Menampilkan performa model	Sistem menampilkan <i>accuracy, precision, recall, F1-score, confusion matrix, dan feature importance</i>

Berdasarkan Tabel 4.6, seluruh modul utama yang dibutuhkan dalam sistem telah berhasil diimplementasikan. Setiap modul memiliki fungsi yang saling mendukung dalam proses prediksi siswa berprestasi.

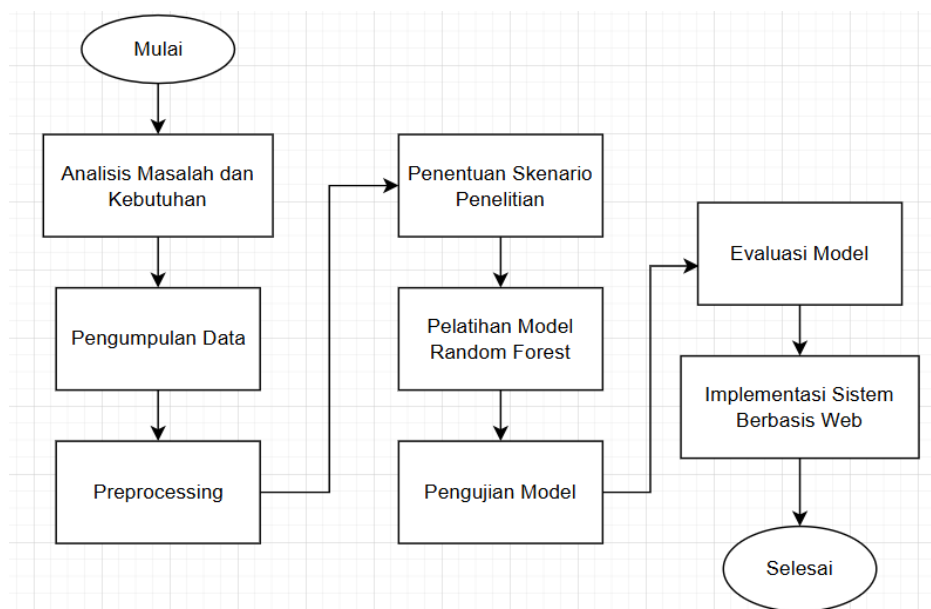
2. Keterkaitan Antar Modul Sistem

Setiap modul dalam sistem saling berhubungan dan membentuk alur kerja sistem prediksi siswa berprestasi. Proses dimulai dari *login* pengguna. Setelah berhasil *login*, pengguna masuk ke halaman *dashboard* untuk melihat ringkasan data.

Dari *dashboard*, pengguna dapat mengakses halaman data siswa. Pada halaman data siswa, pengguna dapat menambahkan, mengubah, atau melihat data siswa. Data siswa yang telah tersimpan kemudian dapat digunakan sebagai input pada proses prediksi.

Sebelum dilakukan prediksi, data siswa melalui tahap *preprocessing* atau konversi data. Pada tahap ini, nilai rapor, kehadiran, kedisiplinan, prestasi lomba, dan ekstrakurikuler diubah menjadi skor numerik. Setelah itu, data dimasukkan ke model *Random Forest* untuk menghasilkan hasil prediksi.

Hasil prediksi kemudian disimpan ke dalam tabel prediksi dan dapat dilihat pada halaman riwayat prediksi. Sementara itu, performa model dapat dilihat pada halaman evaluasi model yang membaca data dari *file model_info.json*.



Gambar 4. 9 Alur Keterkaitan Antar Modul Sistem

Dengan adanya alur keterkaitan antar modul tersebut, dapat dijelaskan bahwa setiap bagian sistem memiliki peran masing-masing. Modul *login* berfungsi untuk keamanan, *dashboard* untuk ringkasan data, data siswa sebagai sumber input, *preprocessing* untuk mengubah data menjadi skor numerik, model *Random Forest* untuk melakukan klasifikasi, riwayat prediksi untuk menyimpan hasil, dan evaluasi model untuk menampilkan kinerja algoritma.

3. Implementasi Pelatihan Model *Random Forest*

Implementasi pelatihan model *Random Forest* dilakukan untuk membangun model klasifikasi yang dapat memprediksi siswa ke dalam kategori Berprestasi dan Tidak Berprestasi. Data yang digunakan dalam proses pelatihan berasal dari tabel *data_latih*.

Tahapan pelatihan model adalah sebagai berikut:

- Sistem melakukan koneksi ke *database MySQL*.
- Sistem mengambil data dari tabel *data_latih*.
- Data yang digunakan terdiri dari nilai rapor, kehadiran, kedisiplinan, prestasi lomba, ekstrakurikuler, dan label.
- Nilai rapor, kehadiran, dan kedisiplinan dikonversi ke bentuk skor numerik.
- Data fitur dipisahkan dari data label.
- Dataset dibagi menjadi data *training* dan data *testing*.
- Model *Random Forest* dilatih menggunakan data *training*.
- Model diuji menggunakan data *testing*.
- Hasil evaluasi model dihitung menggunakan *accuracy, precision, recall, F1-score, dan confusion matrix*.
- Model disimpan ke dalam file *model_rf.pkl*.
- Informasi evaluasi disimpan ke dalam file *model_info.json*.

```
X = df[["nilai_rapor", "kehadiran", "kedisiplinan", "prestasi_lomba", "ekstrakurikuler"]]
y = df["label"]

X_train, X_test, y_train, y_test = train_test_split(
    X, y,
    test_size=0.3, # testing lebih banyak
    random_state=42,
    stratify=y
)

model = RandomForestClassifier(
    n_estimators=50, # lebih sedikit pohon
    max_depth=4, # lebih dangkal
    random_state=42
)
```

Gambar 4.10 Pelatihan Model Random Forest

Gambar 4.10 menunjukkan potongan code pelatihan model *Random Forest*. Pada proses ini, dataset dibagi menjadi *data training* dan *data testing*, kemudian model dilatih menggunakan algoritma *Random Forest*.

Tabel 4. 7 Parameter Model Random Forest

Parameter	Nilai	Keterangan
<i>n_estimators</i>	50	Jumlah pohon keputusan yang digunakan dalam <i>Random Forest</i>
<i>max_depth</i>	4	Batas maksimum kedalaman setiap pohon keputusan
<i>random_state</i>	42	Digunakan agar hasil pembagian data dan pelatihan tetap konsisten
<i>test_size</i>	0.3	Persentase data testing sebesar 30%
<i>stratify</i>	y	Menjaga proporsi label pada data training dan testing

Berdasarkan Tabel 4.7, model *Random Forest* menggunakan 50 pohon keputusan dengan kedalaman maksimum 4. Pembagian data dilakukan dengan *data testing* sebesar 30% dan *data training* sebesar 70%. Parameter *stratify* = y digunakan agar proporsi label pada *data training* dan testing tetap seimbang.

1. Pengujian Fungsional

Pengujian fungsional dilakukan untuk memastikan bahwa setiap fitur pada sistem berjalan sesuai dengan kebutuhan yang telah dirancang. Pengujian dilakukan menggunakan metode *Black Box Testing*, yaitu pengujian berdasarkan *input* dan *output* sistem tanpa melihat *code program* secara langsung.

Tabel 4. 8 Pengujian Fungsional Sistem

No	Fitur yang Diuji	Skenario Pengujian	Hasil yang Diharapkan	Hasil Pengujian	Keterangan
1	Login	Pengguna memasukkan <i>username</i> dan <i>password</i> benar	Sistem masuk ke halaman <i>dashboard</i>	Sesuai	Berhasil
2	Login	Pengguna memasukkan <i>username</i> atau <i>password</i> salah	Sistem menampilkan pesan <i>error</i>	Sesuai	Berhasil
3	Dashboard	Pengguna membuka halaman <i>dashboard</i>	Sistem menampilkan ringkasan data siswa dan prediksi	Sesuai	Berhasil
4	Tambah Data Siswa	Pengguna mengisi form data siswa lalu menyimpan	Data siswa tersimpan ke database	Sesuai	Berhasil
5	Edit Data Siswa	Pengguna mengubah data siswa	Data siswa berhasil diperbarui	Sesuai	Berhasil
6	Hapus Data Siswa	Pengguna menghapus data siswa	Data siswa berhasil dihapus jika belum memiliki riwayat prediksi	Sesuai	Berhasil

No	Fitur yang Diuji	Skenario Pengujian	Hasil yang Diharapkan	Hasil Pengujian	Keterangan
7	Import CSV	Pengguna menjalankan proses <i>import data CSV</i>	Data dari <i>CSV</i> tersimpan ke <i>database</i>	Sesuai	Berhasil
8	Prediksi Siswa	Pengguna memilih satu siswa lalu menekan tombol prediksi	Sistem menampilkan hasil prediksi siswa	Sesuai	Berhasil
9	Prediksi Semua Siswa	Pengguna menekan tombol prediksi semua siswa	Sistem menampilkan hasil prediksi seluruh siswa	Sesuai	Berhasil
10	Riwayat Prediksi	Pengguna membuka halaman riwayat prediksi	Sistem menampilkan data hasil prediksi	Sesuai	Berhasil
11	Evaluasi Model	Pengguna membuka halaman evaluasi model	Sistem menampilkan hasil evaluasi model	Sesuai	Berhasil
12	Logout	Pengguna menekan tombol <i>logout</i>	<i>Session</i> pengguna dihapus	Sesuai	Berhasil

Berdasarkan hasil pengujian pada Tabel 4.8, seluruh fitur utama sistem dapat berjalan sesuai dengan rancangan. Sistem mampu melakukan autentikasi pengguna, mengelola data siswa, melakukan *import data CSV*, melakukan prediksi, menyimpan riwayat prediksi, menampilkan evaluasi model, dan mengakhiri *session* pengguna melalui fitur *logout*.

2. Pengujian Non Fungsional

Pengujian non-fungsional dilakukan untuk mengetahui kualitas sistem dari beberapa aspek, yaitu performa model, keamanan, kemudahan penggunaan, dan keandalan sistem.

a. Pengujian Performa Model

Pengujian performa model dilakukan untuk mengetahui kemampuan algoritma *Random Forest* dalam mengklasifikasikan siswa ke dalam kategori Berprestasi dan Tidak Berprestasi.

Model dilatih menggunakan data dari tabel data_latih. Data dibagi menjadi data *training* sebanyak 73 siswa (70%) dan data *testing* sebanyak 32 siswa (30%). Pembagian data dilakukan menggunakan fungsi *train_test_split* dengan parameter *test_size=0,3*, *random_state=42*, dan *stratify=y* untuk menjaga proporsi label pada data *training* dan *testing* tetap seimbang.

Setelah model dilatih, dilakukan prediksi terhadap data *testing*. Hasil prediksi kemudian dibandingkan dengan label sebenarnya untuk memperoleh *confusion matrix* dan metrik evaluasi.

b. Confusion Matrix

Confusion matrix digunakan untuk melihat jumlah data yang berhasil diprediksi dengan benar dan data yang salah diprediksi. Hasil *confusion matrix* model *Random Forest* ditampilkan pada Tabel 4.9.

Tabel 4.9 Confusion Matrix

Aktual / Prediksi	Prediksi Berprestasi	Prediksi Tidak Berprestasi
Aktual Berprestasi	TP = 11	FN = 3
Aktual Tidak Berprestasi	FP = 1	TN = 17

Keterangan:

86

1. TP (*True Positive*) = 11: jumlah siswa berprestasi yang diprediksi dengan benar sebagai Berprestasi.

71

2. FN (*False Negative*) = 3: jumlah siswa berprestasi yang salah diprediksi sebagai Tidak Berprestasi.

179

3. FP (*False Positive*) = 1: jumlah siswa tidak berprestasi yang salah diprediksi sebagai Berprestasi.

4. TN (*True Negative*) = 17: jumlah siswa tidak berprestasi yang diprediksi dengan benar sebagai Tidak Berprestasi. Total data *testing* adalah:

$$Total = TP + FN + FP + TN$$

$$Total = 11 + 3 + 1 + 17 = 32$$

63

Dari total 32 data *testing*, sebanyak 28 data berhasil diprediksi dengan benar dan 4 data mengalami kesalahan prediksi.

119

c. Perhitungan *Accuracy*

Accuracy digunakan untuk mengukur tingkat ketepatan model dalam memprediksi seluruh data.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Accuracy = \frac{11+17}{11+17+1+3}$$

$$Accuracy = \frac{28}{32}$$

$$Accuracy = 0.8750$$

224

Berdasarkan hasil perhitungan tersebut, nilai *accuracy* yang diperoleh adalah 0,875 atau 87,50%. Nilai ini menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data *testing* dengan benar.

106

d. Perhitungan *Precision*

Precision digunakan untuk mengukur ketepatan model ketika memprediksi siswa sebagai kategori Berprestasi.

$$Precision = \frac{TP}{TP+FP}$$

$$Precision = \frac{11}{11+1}$$

$$Precision = \frac{11}{12}$$

$$Precision = 0.9167$$

Nilai *precision* yang diperoleh adalah 0,9167 atau 91,67%, menunjukkan bahwa sebagian besar siswa yang diprediksi Berprestasi memang benar termasuk kategori Berprestasi.

e. Perhitungan *Recall*

Recall digunakan untuk mengukur kemampuan model dalam menemukan seluruh siswa yang benar-benar berprestasi.

$$Recall = \frac{TP}{TP+FN}$$

$$Recall = \frac{11}{11+3}$$

$$Recall = \frac{11}{14}$$

$$Recall = 0.7857$$

Nilai *recall* sebesar 0,7857 atau 78,57%, menunjukkan bahwa beberapa siswa berprestasi tidak terdeteksi oleh model (*False Negative*).

f. Perhitungan *F1-Score*

F1-score digunakan untuk melihat keseimbangan antara nilai *precision* dan *recall*.

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

$$F1 - Score = \frac{2 \times 0.9167 \times 0.7857}{0.9167 + 0.7857}$$

$$F1 - Score = 0.8462$$

Nilai *F1-score* sebesar 0,8462 atau 84,62%, menunjukkan bahwa model memiliki keseimbangan yang baik antara ketepatan prediksi dan kemampuan mengenali siswa berprestasi.

g. Rekapitulasi Hasil Evaluasi Model

Hasil evaluasi model *Random Forest* berdasarkan *accuracy*, *precision*, *recall*, dan *F1-score* ditampilkan pada Tabel 4.10

Tabel 4. 10 Hasil Evaluasi Model *Random Forest*

Metrik Evaluasi	Presentase
<i>Accuracy</i>	87,50%
<i>Precision</i>	91,67%,
<i>Recall</i>	78,57%,
<i>F1-Score</i>	84,62%,

Berdasarkan Tabel 4.10, model *Random Forest* memperoleh *accuracy* sebesar 87,50%, *precision* sebesar 91,67%, *recall* sebesar 78,57%, dan *F1-score* sebesar 84,62%. Hasil tersebut menunjukkan bahwa model memiliki performa yang baik dalam mengklasifikasikan siswa berprestasi dan tidak berprestasi, meskipun beberapa siswa berprestasi masih tidak terdeteksi (*False Negative*).

h. *Feature Importance*

Feature importance digunakan untuk mengetahui pengaruh masing-masing fitur terhadap hasil prediksi model *Random Forest*. Fitur yang digunakan dalam penelitian ini adalah nilai rapor, kehadiran, kedisiplinan, prestasi lomba, dan ekstrakurikuler.

Nilai *feature importance* diperoleh dari model *Random Forest* setelah proses training selesai. Semakin besar nilai *importance* suatu fitur, maka semakin besar pengaruh fitur tersebut terhadap hasil klasifikasi.

Tabel 4. 11 Feature Importance

Fitur	Nilai
Nilai rapor	0.1230
Kehadiran	0.2764
Kedisiplinan	0.0562
Ekastrakurikuler	0.2601
Prestasi lomba	0.2843

Berdasarkan Tabel 4.11, fitur dengan nilai importance tertinggi merupakan fitur yang paling berpengaruh

terhadap hasil prediksi siswa berprestasi. Informasi ini dapat digunakan oleh pihak sekolah untuk memahami faktor mana yang paling dominan dalam proses klasifikasi.

i. Pengujian Keamanan

Pengujian keamanan dilakukan untuk memastikan bahwa sistem tidak dapat diakses oleh pengguna yang tidak memiliki hak akses. Sistem menggunakan halaman login sebagai pengaman awal.

Pada proses *login*, pengguna harus memasukkan *username* dan *password*. Jika data yang dimasukkan benar, maka pengguna dapat masuk ke sistem. Jika data salah, maka sistem menampilkan pesan *error* dan pengguna tidak dapat mengakses halaman utama.

Selain itu, sistem juga menggunakan fungsi *login_required* untuk melindungi halaman tertentu. Apabila pengguna belum *login* dan mencoba mengakses halaman sistem, maka pengguna akan diarahkan kembali ke halaman *login*. Berdasarkan hasil pengujian, sistem telah memiliki mekanisme keamanan dasar yang cukup baik untuk membatasi akses pengguna

j. Pengujian Kemudahan Penggunaan

Pengujian kemudahan penggunaan dilakukan untuk mengetahui apakah sistem mudah digunakan oleh pengguna. Sistem memiliki menu yang jelas, seperti *dashboard*, data siswa, prediksi, riwayat, evaluasi, dan *logout*.

Form input data siswa dibuat sesuai dengan variabel penelitian, sehingga pengguna dapat memasukkan data dengan mudah. Hasil prediksi juga ditampilkan secara langsung dalam bentuk kategori Berprestasi atau Tidak Berprestasi.

Selain itu, sistem juga menampilkan penjelasan proses prediksi, seperti konversi nilai dan voting mayoritas dari *decision tree*. Hal ini membantu pengguna memahami bagaimana sistem menghasilkan keputusan prediksi. Berdasarkan pengujian, sistem dapat digunakan dengan mudah oleh pengguna karena tampilan dan alur sistem disusun secara sederhana.

k. Pengujian Keandalan Sistem

Pengujian keandalan dilakukan untuk memastikan bahwa sistem dapat menjalankan fungsi utama dengan baik. Sistem mampu melakukan proses login, menampilkan *dashboard*, mengelola data siswa, melakukan konversi data, memuat model *Random Forest*, melakukan prediksi, menyimpan hasil prediksi, dan menampilkan evaluasi model.

Sistem juga memiliki mekanisme penanganan kesalahan. Apabila file *model_rf.pkl* belum tersedia, maka sistem akan menampilkan pesan bahwa model belum dilatih. Dengan demikian, pengguna mengetahui bahwa proses *training* model harus dilakukan terlebih dahulu sebelum menggunakan fitur prediksi. Berdasarkan hasil pengujian, sistem dapat berjalan dengan baik dan dapat diandalkan untuk melakukan prediksi siswa berprestasi.

B. Pembahasan

Penelitian ini menghasilkan sistem prediksi siswa berprestasi menggunakan algoritma *Random Forest*. Sistem ini dibangun untuk menjawab permasalahan yang telah dijelaskan pada Bab I, yaitu proses penentuan siswa berprestasi yang masih dilakukan secara manual dan berpotensi kurang objektif.

Dengan adanya sistem ini, proses identifikasi siswa berprestasi dapat dilakukan secara lebih cepat, terukur, dan berbasis data. Sistem tidak hanya menggunakan data akademik berupa nilai rapor, tetapi juga menggunakan data non-akademik seperti kehadiran, kedisiplinan, prestasi lomba, dan ekstrakurikuler.

Penggunaan kombinasi data akademik dan non-akademik bertujuan agar hasil prediksi lebih menyeluruh. Siswa berprestasi tidak hanya dilihat dari nilai akademik saja, tetapi juga dari sikap, kehadiran, prestasi, dan keaktifan dalam kegiatan sekolah.

1. Implementasi Sistem

Berdasarkan hasil implementasi, sistem telah berhasil dibangun sesuai dengan rancangan pada Bab III. Sistem memiliki modul *login*, *dashboard*, data siswa, input data siswa, *import CSV*, prediksi, riwayat prediksi, dan evaluasi model.

Modul login berfungsi untuk membatasi akses pengguna. Modul dashboard membantu pengguna melihat ringkasan data secara cepat. Modul data siswa digunakan untuk mengelola data siswa. Modul *import CSV* membantu memasukkan dataset ke database. Modul prediksi digunakan untuk melakukan klasifikasi siswa berprestasi. Modul riwayat digunakan untuk menyimpan hasil prediksi, sedangkan modul evaluasi digunakan untuk melihat performa model.

Dengan adanya modul-modul tersebut, sistem dapat membantu pihak sekolah dalam melakukan proses identifikasi siswa berprestasi secara lebih sistematis.

2. Preprocessing Data

Tahap preprocessing pada sistem dilakukan dengan cara mengonversi data mentah menjadi skor numerik. Proses ini sangat penting karena algoritma *Random Forest* memerlukan data input dalam bentuk angka.

Nilai rapor dikonversi berdasarkan rentang nilai. Kehadiran dikonversi berdasarkan persentase. Kedisiplinan dikonversi berdasarkan jumlah pelanggaran. Prestasi lomba dikonversi berdasarkan tingkat lomba. Ekstrakurikuler dikonversi berdasarkan tingkat keaktifan siswa.

Dengan adanya proses konversi ini, data yang awalnya memiliki bentuk berbeda-beda dapat diseragamkan menjadi skor 1 sampai 4. Data yang sudah seragam kemudian dapat diproses oleh model *Random Forest*.

3. Penerapan Random Forest

Algoritma *Random Forest* bekerja dengan membentuk banyak *decision tree*. Setiap *decision tree* memberikan hasil prediksi masing-masing. Hasil akhir kemudian ditentukan berdasarkan voting mayoritas.

Pada sistem ini, model Random Forest menggunakan parameter *n_estimators* = 50, yang berarti model membentuk 50 pohon keputusan. Parameter *max_depth* = 4 digunakan untuk membatasi kedalaman pohon agar model tidak terlalu kompleks dan menghasilkan prediksi yang lebih realistis.

Penerapan *Random Forest* pada sistem ini sesuai dengan kebutuhan penelitian karena algoritma tersebut mampu menangani beberapa fitur sekaligus dan dapat menghasilkan klasifikasi yang stabil.

4. Hasil Evaluasi Model

Berdasarkan hasil evaluasi, model memperoleh *accuracy* sebesar 87,50%, *precision* sebesar 91,67%, *recall* sebesar 78,57%, dan *F1-score* sebesar 84,62%.

- a) Nilai *accuracy* sebesar 87,50% menunjukkan bahwa model mampu memprediksi sebagian besar data *testing* dengan benar. Dari total 32 data *testing*, sebanyak 28 data berhasil diprediksi dengan benar.
- b) Nilai *precision* sebesar 91,67% menunjukkan bahwa model cukup tepat dalam memprediksi siswa sebagai Berprestasi. Dari seluruh siswa yang diprediksi Berprestasi, sebagian besar benar-benar termasuk kategori Berprestasi..
- c) Nilai *recall* sebesar 78,57% menunjukkan bahwa model mampu mengenali sebagian besar siswa yang benar-benar berprestasi. Masih terdapat 3 siswa berprestasi yang salah diklasifikasikan sebagai Tidak Berprestasi (*False Negative*).
- d) Nilai *F1-score* sebesar 84,62% menunjukkan bahwa model memiliki keseimbangan yang baik antara *precision* dan *recall*. Dengan demikian, model *Random Forest* dapat dikatakan memiliki performa yang baik dalam melakukan klasifikasi siswa berprestasi secara realistis.

5. Kesalahan Prediksi

Berdasarkan *confusion matrix*, terdapat satu kesalahan prediksi, yaitu siswa yang sebenarnya tidak berprestasi tetapi diprediksi sebagai berprestasi. Kesalahan ini termasuk dalam kategori *False Positive*.

Kesalahan prediksi dapat terjadi karena pola data siswa tersebut memiliki kemiripan dengan siswa berprestasi. Misalnya, siswa memiliki nilai cukup baik pada beberapa variabel, tetapi secara label sebenarnya masuk kategori tidak berprestasi.

Kesalahan seperti ini masih wajar terjadi dalam model *machine learning*. Untuk mengurangi kesalahan prediksi, beberapa upaya dapat dilakukan, antara lain menambah jumlah data latih, memperbaiki kualitas data, melakukan tuning parameter

Random Forest, menambahkan variabel lain yang relevan, dan memastikan proses pelabelan data dilakukan secara konsisten.

6. Perbandingan Sebelum dan Sesudah Sistem

Untuk melihat manfaat sistem yang dikembangkan, dilakukan perbandingan antara kondisi sebelum dan sesudah implementasi sistem. Perbandingan tersebut ditampilkan pada Tabel 4.12.

Tabel 4. 12 Perbandingan Sebelum dan Sesudah Implementasi Sistem

Aspek	Sebelum Sistem	Sesudah Sistem
Proses penentuan siswa berprestasi	Dilakukan secara manual	Dilakukan otomatis menggunakan sistem
Dasar penilaian	Berdasarkan perhitungan dan pertimbangan manual	Berdasarkan data akademik dan non-akademik
Kecepatan proses	Membutuhkan waktu lebih lama	Lebih cepat karena diproses oleh sistem
Objektivitas	Masih berpotensi subjektif	Lebih objektif karena menggunakan model <i>Random Forest</i>
Penyimpanan hasil	Tidak terdokumentasi secara sistematis	Tersimpan dalam riwayat prediksi
Evaluasi model	Tidak tersedia	Tersedia <i>accuracy, precision, recall, F1-score, dan confusion matrix</i>
Kemudahan monitoring	Sulit dilakukan karena data tersebar	Lebih mudah melalui <i>dashboard</i> dan riwayat prediksi

Berdasarkan Tabel 4.12, sistem yang dikembangkan memberikan beberapa manfaat dibandingkan proses manual. Proses prediksi menjadi lebih cepat karena dilakukan secara otomatis. Hasil prediksi juga lebih objektif karena berdasarkan data dan model *machine learning*. Selain itu, hasil prediksi dapat disimpan di dalam

database sehingga lebih mudah digunakan sebagai dokumentasi dan bahan evaluasi. *Dashboard* dan riwayat prediksi juga membantu pengguna dalam memantau hasil klasifikasi siswa.

7. Kesesuaian Hasil dengan Tujuan Penelitian

Tujuan penelitian ini adalah membuat model klasifikasi berbasis *machine learning* yang dapat memprediksi siswa berprestasi dan menilai performa algoritma *Random Forest* dalam melakukan prediksi.

Berdasarkan hasil implementasi, sistem telah berhasil dibangun dan dapat melakukan prediksi siswa berprestasi. Berdasarkan hasil evaluasi, algoritma *Random Forest* memperoleh *accuracy* sebesar 87,50%, *precision* sebesar 91,67%, *recall* sebesar 78,57%, dan *F1-score* sebesar 84,62%.

Dengan demikian, tujuan penelitian telah tercapai. Sistem yang dibuat mampu melakukan klasifikasi siswa berprestasi berdasarkan data akademik dan non-akademik, serta algoritma *Random Forest* menunjukkan performa yang baik.

8. Kelebihan dan Keterbatasan Sistem

a. Kelebihan Sistem

Kelebihan sistem yang dikembangkan adalah sebagai berikut:

- 1) Sistem dapat membantu proses prediksi siswa berprestasi secara otomatis.
- 2) Sistem menggunakan data akademik dan non-akademik sehingga penilaian lebih menyeluruh.
- 3) Sistem memiliki fitur *dashboard* untuk menampilkan ringkasan data.
- 4) Sistem menyimpan riwayat prediksi sehingga hasil dapat dilihat kembali.
- 5) Sistem menampilkan evaluasi model sehingga pengguna dapat mengetahui performa algoritma.
- 6) Sistem menggunakan *Random Forest* yang mampu menghasilkan klasifikasi cukup akurat.
- 7) Sistem mendukung data dari *CSV* sehingga proses pengisian dataset menjadi lebih cepat.

b. Keterbatasan Sistem

Keterbatasan sistem yang dikembangkan adalah sebagai berikut:

162

50

- 1) Sistem dapat menggunakan algoritma *Random Forest* dan belum dibandingkan dengan algoritma lain.
- 2) Data yang digunakan terbatas pada variabel nilai rapor, kehadiran, kedisiplinan, prestasi lomba, dan ekstrakurikuler.
- 3) Sistem belum memberikan rekomendasi tindak lanjut setelah hasil prediksi diperoleh.
- 4) Akurasi model masih bergantung pada jumlah data dan kualitas data latih.

13

BAB V

PENUTUP

A. Kesimpulan

Berdasarkan hasil penelitian mengenai penerapan algoritma *Random Forest* dalam sistem prediksi siswa berprestasi berdasarkan data akademik dan non-akademik di SMP Negeri 4 Kediri, maka dapat disimpulkan sebagai berikut:

1. Penelitian ini berhasil merancang dan membangun sistem prediksi siswa berprestasi berbasis *web* menggunakan algoritma *Random Forest*. Sistem memanfaatkan data akademik dan non-akademik siswa, seperti nilai rapor, kehadiran, kedisiplinan, prestasi lomba, dan keaktifan ekstrakurikuler. Sistem yang dikembangkan mampu membantu pihak sekolah dalam melakukan identifikasi siswa berprestasi secara lebih terstruktur dan objektif.
2. Algoritma *Random Forest* mampu digunakan sebagai metode klasifikasi dalam memprediksi siswa berprestasi dengan performa yang baik. Berdasarkan evaluasi terbaru, model memperoleh *accuracy* sebesar 87,50%, *precision* sebesar 91,67%, *recall* sebesar 78,57%, dan *F1-score* sebesar 84,62%, menunjukkan bahwa sebagian besar siswa berprestasi dapat dikenali dengan tepat, meskipun beberapa siswa masih tidak terdeteksi (*False Negative*). Sistem ini membantu mengurangi subjektivitas dalam penentuan siswa berprestasi sehingga hasil prediksi lebih konsisten dan realistis. Hasil pengujian menunjukkan bahwa sistem prediksi yang dibangun dapat membantu mempercepat proses penilaian siswa berprestasi. Selain itu, penggunaan sistem berbasis data membantu mengurangi unsur subjektivitas dalam proses penentuan siswa berprestasi sehingga hasil yang diperoleh menjadi lebih konsisten dan akurat.

B. Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang diharapkan dapat menjadi bahan pengembangan pada penelitian selanjutnya, yaitu sebagai berikut:

1. Penelitian selanjutnya dapat menambahkan variabel lain seperti, sikap sosial, hasil psikotes, atau data perilaku siswa, agar prediksi siswa berprestasi menjadi lebih lengkap dan akurat.

2. Penelitian berikutnya dapat membandingkan algoritma *Random Forest* dengan algoritma *machine learning* lain, seperti *Support Vector Machine (SVM)*, *Naive Bayes*, *Decision Tree*, atau *Neural Network*, untuk mengetahui metode terbaik dalam klasifikasi siswa berprestasi.