

CEK TURNITIN

SKRIPSI KHUSNUL KHOTIMAH SI

 SATSET DIGITAL 3

Document Details

Submission ID

trn:oid:::3618:151077385

Submission Date

Sep 9, 2026, 11:07 AM GMT+7

Download Date

Sep 9, 2026, 11:10 AM GMT+7

File Name

SKRIPSI KHUSNUL KHOTIMAH SI.pdf

File Size

1.6 MB

113 Pages

20,422 Words

128,866 Characters

42% Overall Similarity

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

Filtered from the Report

- Bibliography

Top Sources

- 30%  Internet sources
- 24%  Publications
- 35%  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

- 30% Internet sources
- 24% Publications
- 35% 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.unisi.ac.id	2%
2	Internet	media.neliti.com	1%
3	Internet	ejournal.sisfokomtek.org	<1%
4	Internet	ejournal.uin-suka.ac.id	<1%
5	Student papers	Universitas Pamulang on 2026-08-05	<1%
6	Internet	jurnal.itscience.org	<1%
7	Publication	Fajar Dedi Pratama, Henny Dwi Bhakti. "IMPLEMENTASI APLIKASI PREDIKSI KETEP...	<1%
8	Internet	repository.uin-suska.ac.id	<1%
9	Publication	Ahmad Rizki Saputra, Rina - Wati, Tri - Susilowati. "KLASIFIKASI STATUS KELAYAKA...	<1%
10	Internet	proceeding.unpkediri.ac.id	<1%
11	Internet	ojs.selodangmayang.com	<1%

12	Student papers	Universitas Sangga Buana YPKP on 2026-07-24	<1%
13	Internet	repository.umsu.ac.id	<1%
14	Internet	cdn.juris.id	<1%
15	Internet	123dok.com	<1%
16	Publication	Sri Damayanti, Husna Sari, Bobby Ardiansyah, Dicky Apdillah. "Analisis Sentimen ...	<1%
17	Internet	docplayer.info	<1%
18	Internet	jurnal.pustakagalerimandiri.co.id	<1%
19	Internet	repository.ar-raniry.ac.id	<1%
20	Internet	repo.palcomtech.ac.id	<1%
21	Student papers	Konsorsium PTS Batch 5 on 2026-06-26	<1%
22	Student papers	Universitas Putera Batam on 2025-07-29	<1%
23	Internet	finkom.repository.unbin.ac.id	<1%
24	Internet	docs.google.com	<1%
25	Student papers	Universitas Islam Sumatera Utara on 2026-04-20	<1%

26	Internet	repositori.unibos.ac.id	<1%
27	Student papers	Konsorsium PTS Batch 5 on 2026-06-24	<1%
28	Internet	repository.uinsaizu.ac.id	<1%
29	Internet	download.garuda.kemdikbud.go.id	<1%
30	Internet	lab.tik.pnj.ac.id	<1%
31	Publication	Kamia Pertiwi, Usman Usman, Ilyas Ilyas. "Prediksi Risiko Ketergantungan Smart...	<1%
32	Student papers	UIN Ar-Raniry on 2025-04-25	<1%
33	Student papers	Universitas Pamulang on 2026-07-03	<1%
34	Internet	journal.appisi.or.id	<1%
35	Internet	etheses.uin-malang.ac.id	<1%
36	Internet	jtiik.ub.ac.id	<1%
37	Internet	ejurnal.seminar-id.com	<1%
38	Internet	repository.ulb.ac.id	<1%
39	Internet	journalcenter.org	<1%

40	Student papers	Konsorsium PTS Batch 5 on 2026-06-26	<1%
41	Student papers	Universitas Gunadarma on 2026-07-17	<1%
42	Internet	eskripsi.usm.ac.id	<1%
43	Student papers	Sriwijaya University on 2026-04-28	<1%
44	Student papers	Universitas Putera Batam on 2026-02-02	<1%
45	Student papers	Universitas Sanata Dharma on 2026-06-09	<1%
46	Student papers	Universitas Putera Batam on 2026-07-27	<1%
47	Student papers	Ajou University Graduate School on 2024-07-11	<1%
48	Student papers	Universitas Pamulang on 2026-08-03	<1%
49	Student papers	Universitas Islam Riau on 2026-07-08	<1%
50	Student papers	Universitas Khairun on 2026-07-13	<1%
51	Student papers	LL DIKTI IX Turnitin Consortium Part III on 2026-07-21	<1%
52	Publication	Cantriya Anastasya Simbolon, Maria Angelina Lubis, Sardo Pardingotan Sipayun...	<1%
53	Student papers	Dongguk University on 2024-10-05	<1%

54	Student papers	UPN Veteran Yogyakarta on 2026-04-02	<1%
55	Student papers	Universitas Pancasila on 2026-02-21	<1%
56	Student papers	UIN Ar-Raniry on 2026-06-29	<1%
57	Student papers	Universitas Pamulang on 2026-05-22	<1%
58	Student papers	Universitas Bengkulu on 2025-10-17	<1%
59	Internet	jurnal.polibatam.ac.id	<1%
60	Student papers	Universitas Tarumanagara on 2025-12-19	<1%
61	Internet	aksiologi.org	<1%
62	Internet	ejurnal.stmik-budidarma.ac.id	<1%
63	Internet	ojs.sains.ac.id	<1%
64	Internet	repository.unmuhjember.ac.id	<1%
65	Student papers	Universitas Darussalam Gontor on 2026-01-25	<1%
66	Student papers	Universitas Islam Negeri Sumatera Utara on 2026-06-18	<1%
67	Student papers	Universitas Pamulang on 2026-07-21	<1%

68	Student papers	Universitas Sangga Buana YPKP on 2026-07-20	<1%
69	Internet	dspace.uui.ac.id	<1%
70	Internet	repository.upnjatim.ac.id	<1%
71	Publication	Ahmad Dani, Muhammad Khairul Fahri, Harry Dwiki Pramadan, Fadli Anwar Sin...	<1%
72	Publication	Fidalina Nirigi, Mochammad Triyanto, Mohammad Rezza Pahlevi, Athia Saelan, Fa...	<1%
73	Student papers	UIN KH. Achmad Siddiq Jember on 2026-08-10	<1%
74	Student papers	UPN Veteran Yogyakarta on 2026-06-29	<1%
75	Student papers	Universitas Islam Riau on 2025-08-04	<1%
76	Internet	repository.telkomuniversity.ac.id	<1%
77	Internet	repository.unissula.ac.id	<1%
78	Student papers	UPN Veteran Yogyakarta on 2026-07-21	<1%
79	Student papers	Universitas Muhammadiyah Sukabumi on 2025-07-18	<1%
80	Internet	repository.ubpkarawang.ac.id	<1%
81	Student papers	UPN Veteran Yogyakarta on 2026-07-03	<1%

82	Student papers	Universitas Dinamika on 2026-06-11	<1%
83	Student papers	Universitas Muhammadiyah Makassar on 2026-06-25	<1%
84	Student papers	Universitas Pendidikan Ganesha on 2025-01-01	<1%
85	Student papers	Universitas Sangga Buana YPKP on 2026-07-31	<1%
86	Internet	ejournal.itn.ac.id	<1%
87	Publication	Arisman Arisman, Nurhayati Nurhayati, Tanti Tanti. "Analisis Klasifikasi Tingkat H...	<1%
88	Student papers	Fakultas Teknologi Informasi on 2025-11-03	<1%
89	Student papers	Rektorat (ADM) on 2026-08-27	<1%
90	Student papers	Universitas Khairun on 2026-05-25	<1%
91	Student papers	Universitas Putera Batam on 2026-07-24	<1%
92	Student papers	Universitas Putera Batam on 2026-07-28	<1%
93	Student papers	UNIVERSITAS BUDI LUHUR on 2026-07-14	<1%
94	Student papers	Universitas Khairun on 2026-08-14	<1%
95	Student papers	Universitas Pamulang on 2026-07-22	<1%

96	Internet	digilibadmin.unismuh.ac.id	<1%
97	Publication	M. Rizki Mahrifan, Fitri Yunita. "Analisis Sentimen pada Aplikasi MyXL Mengguna...	<1%
98	Publication	Fikriman Fikriman, Febri Ari Budiman, Evo Afrianto. "FAKTOR SOSIAL EKONOMI Y...	<1%
99	Student papers	Forum Perpustakaan Perguruan Tinggi Indonesia Jawa Timur II on 2025-08-20	<1%
100	Student papers	Konsorsium Perguruan Tinggi Swasta Indonesia II on 2021-08-23	<1%
101	Publication	Salman Salman. "Penerapan Metode SAW untuk Menentukan Penerima Bantuan ...	<1%
102	Student papers	UIN Sultan Syarif Kasim Riau on 2020-08-27	<1%
103	Student papers	Universitas Islam Riau on 2026-07-27	<1%
104	Student papers	Universitas Muhammadiyah Sukabumi on 2026-07-01	<1%
105	Student papers	Universitas Muria Kudus on 2016-06-09	<1%
106	Student papers	Universitas Tarumanagara on 2026-01-04	<1%
107	Internet	eprints.undip.ac.id	<1%
108	Internet	jurnal.buddhidharma.ac.id	<1%
109	Internet	lib.geo.ugm.ac.id	<1%

110	Internet	repository.ub.ac.id	<1%
111	Publication	Dwi Haryono Wiratno, Rahmawati Hanny Yustrianthe. "ANALISIS PERBANDINGA...	<1%
112	Student papers	Universitas Bengkulu on 2025-09-23	<1%
113	Internet	eprints.amikom.ac.id	<1%
114	Internet	journal.stmikjayakarta.ac.id	<1%
115	Internet	repository.ibs.ac.id	<1%
116	Publication	Diah Puspita Sari, Mario Ekoriano, Aditya Rahmadhony. "THE RELATIONSHIP BET...	<1%
117	Publication	Hidayat Hidayat, Musliadi KH, Indar Kusmanto, Munawirah Kadir, Kristian Kristia...	<1%
118	Student papers	Konsorsium PTS Batch 5 on 2026-06-26	<1%
119	Publication	Shofi Camila Laisya Amanda, Fitri Yunita, Samsudin Samsudin. "Optimasi Metode ...	<1%
120	Student papers	Universitas Mercu Buana Yogyakarta on 2026-07-23	<1%
121	Student papers	Universitas Pamulang on 2026-07-11	<1%
122	Internet	ejurnal.lkpkaryaprima.id	<1%
123	Internet	eprints.polbeng.ac.id	<1%

124	Internet	objectstorage.ap-dcc-gazipur-1.oraclecloud15.com	<1%
125	Internet	www.lembagakita.org	<1%
126	Publication	Adinda Purnamasari, Rini Astuti, Khaerul Anam, Gifthera Dwilestari, Mulyawan. ...	<1%
127	Publication	Nur Sakina, Farid Wajidi, Muh. Rafli Rasyid. "Evaluasi Algoritma KNN dan Naive B...	<1%
128	Student papers	Saddleback High School on 2025-09-30	<1%
129	Student papers	Sekolah Teknik Elektro & Informatika on 2026-08-03	<1%
130	Student papers	Telkom University on 2026-08-07	<1%
131	Student papers	Universitas Brawijaya on 2017-01-18	<1%
132	Internet	jurnal.kopertipindonesia.or.id	<1%
133	Internet	repository.upiypk.ac.id	<1%
134	Student papers	Universitas Andalas on 2026-08-20	<1%
135	Student papers	Universitas Putera Batam on 2026-02-03	<1%
136	Publication	M. Asyari Syahab, Aripin Rambe, Baginda Harahap, Dinur Syahputra. "Penerapan ...	<1%
137	Publication	Mochammad Dayke Xavier Lumankun, Ghofar Taufik. "Algoritma Naïve Bayes un...	<1%

138	Student papers	UNIVERSITAS BUDI LUHUR on 2026-08-06	<1%
139	Internet	ejournal.unama.ac.id	<1%
140	Internet	ejurnal.provisi.ac.id	<1%
141	Student papers	itera on 2025-01-22	<1%
142	Student papers	Universitas Brawijaya on 2017-07-27	<1%
143	Student papers	Universitas Katolik Musi Charitas on 2026-08-12	<1%
144	Student papers	Universitas Putera Batam on 2026-02-02	<1%
145	Internet	journal.fkpt.org	<1%
146	Student papers	Academic Library Consortium on 2024-12-27	<1%
147	Publication	Arief Budi Pratomo. "Penerapan Metode SAW dalam Menentukan Prioritas Penerimaan..."	<1%
148	Student papers	Forum Perpustakaan Perguruan Tinggi Indonesia Jawa Timur II on 2025-08-20	<1%
149	Student papers	Indian Institute of Technology Mandi on 2026-08-30	<1%
150	Student papers	Konsorsium PTS Batch 5 on 2026-06-26	<1%
151	Publication	Siti Noraisyah, Fitri Yunita. "Analisis Sentimen Penggunaan ChatGPT di Kalangan..."	<1%

152	Student papers	Universitas Brawijaya on 2017-12-10	<1%
153	Student papers	Universitas Raharja on 2024-08-04	<1%
154	Internet	ejournal.stmik-time.ac.id	<1%
155	Internet	www.jurnal.soratekno.co.id	<1%
156	Publication	Ahmad Baidowi, Sutisna. "Implementasi Data Mining Klasifikasi Fuel Surcharge ...	<1%
157	Student papers	Sriwijaya University on 2026-05-04	<1%
158	Internet	journal.artei.or.id	<1%
159	Internet	jurnal.globalhealthsciencegroup.com	<1%
160	Student papers	Fakultas Teknik on 2025-07-24	<1%
161	Student papers	Konsorsium Perguruan Tinggi Swasta I 2023 on 2026-07-24	<1%
162	Student papers	LL Dikti IX Turnitin Consortium on 2026-09-07	<1%
163	Student papers	Perpustakaan on 2025-08-08	<1%
164	Student papers	Universitas Jenderal Soedirman on 2022-05-08	<1%
165	Student papers	Universitas Muhammadiyah Palembang on 2026-04-26	<1%

166	Internet	dokumen.pub	<1%
167	Internet	lib.ibs.ac.id	<1%
168	Publication	Dinda Aulia, Munadya Salsa Billa Lubis, Steven R.J. Lumbansiantar, Vande Mart Si...	<1%
169	Student papers	UIN Sunan Ampel Surabaya on 2023-05-31	<1%
170	Student papers	UPN Veteran Yogyakarta on 2026-05-05	<1%
171	Student papers	UPN Veteran Yogyakarta on 2026-06-24	<1%
172	Student papers	Universitas Diponegoro on 2017-05-08	<1%
173	Student papers	Universitas Islam Riau on 2025-11-17	<1%
174	Internet	digilib.ulm.ac.id	<1%
175	Internet	eprints.polsri.ac.id	<1%
176	Internet	juris.id	<1%
177	Internet	repository.radenintan.ac.id	<1%
178	Internet	repository.staima-alhikam.ac.id	<1%
179	Publication	Dadang Iskandar Mulyana, Ali Muhammad Fadhil. "Penerapan Data Mining untu...	<1%

180	Student papers	Forum Perpustakaan Perguruan Tinggi Indonesia Jawa Timur on 2025-10-01	<1%
181	Student papers	Institut Teknologi Sumatera on 2026-09-08	<1%
182	Publication	Nadya Andhika, Randi Rian Putra. "SISTEM PENDUKUNG KEPUTUSAN MENENTUK...	<1%
183	Student papers	STT PLN on 2023-08-27	<1%
184	Publication	Salman Alfaris, Kusnawi. "Komparasi Metode KNN dan Naive Bayes Terhadap An...	<1%
185	Student papers	Universitas 17 Agustus 1945 Semarang on 2026-01-13	<1%
186	Student papers	Universitas Islam Negeri Raden Fatah on 2024-10-25	<1%
187	Student papers	Universitas Jenderal Achmad Yani on 2025-08-07	<1%
188	Student papers	Universitas Muria Kudus on 2016-08-04	<1%
189	Student papers	Universitas Pamulang on 2026-05-04	<1%
190	Student papers	Universitas Pamulang on 2026-08-05	<1%
191	Student papers	Universitas Pendidikan Ganesha on 2025-04-17	<1%
192	Internet	proceeding.unuha.ac.id	<1%
193	Internet	repository.unisbablitar.ac.id	<1%

194	Internet	www.ejournal.itn.ac.id	<1%
195	Student papers	Konsorsium Perguruan Tinggi Swasta Indonesia II on 2025-08-26	<1%
196	Student papers	Konsorsium Perguruan Tinggi Swasta Indonesia II on 2025-09-07	<1%
197	Student papers	SDM Universitas Gadjah Mada on 2021-10-08	<1%
198	Student papers	Sriwijaya University on 2023-08-29	<1%
199	Publication	Tia Sutianah, Petrus Sokibi, Viar Dwi Kartika. "Sistem Klasifikasi Penentuan Pener..."	<1%
200	Student papers	Universitas Brawijaya on 2017-05-19	<1%
201	Student papers	Universitas Brawijaya on 2021-05-24	<1%
202	Student papers	Universitas Maritim Raja Ali Haji on 2023-06-23	<1%
203	Student papers	Universitas Muhammadiyah Palembang on 2026-05-02	<1%
204	Student papers	Universitas Muhammadiyah Purwokerto on 2024-11-25	<1%
205	Student papers	Universitas Putera Batam on 2024-07-20	<1%
206	Student papers	Universitas Putera Batam on 2026-02-03	<1%
207	Student papers	Universitas Sanata Dharma on 2026-06-15	<1%

208	Internet	eprints.unisla.ac.id	<1%
209	Internet	journal.irpi.or.id	<1%
210	Internet	jurnal.unived.ac.id	<1%
211	Internet	repository.upstegal.ac.id	<1%
212	Publication	Afifah Nurul Izzati. "Analisis Sentimen Hasil Putusan MK Terkait Sengketa Pilpres ...	<1%
213	Student papers	Fakultas Ekonomi Universitas Indonesia on 2015-03-25	<1%
214	Student papers	Fakultas Ekonomi Universitas Indonesia on 2015-07-07	<1%
215	Publication	Maulidya Dwi Nurmalasari, Kusri Kusri, Sudarmawan Sudarmawan. "Kompar...	<1%
216	Student papers	Politeknik Manufaktur Negeri Bangka Belitung on 2026-07-27	<1%
217	Publication	Siti Mutia Dunggio, Rauf A. Hatu, Sainudin Latore. "Dampak Bantuan Langsung T...	<1%
218	Student papers	UIN Sjech M.Djamil Djambek Bukittinggi on 2026-07-08	<1%
219	Student papers	UIN Sultan Syarif Kasim Riau on 2020-06-01	<1%
220	Student papers	UNIVERSITAS BUDI LUHUR on 2026-07-02	<1%
221	Student papers	Universitas Brawijaya on 2021-12-21	<1%

222	Student papers	Universitas Brawijaya on 2023-08-02	<1%
223	Student papers	Universitas Budi Luhur on 2026-01-27	<1%
224	Student papers	Universitas Islam Riau on 2026-06-15	<1%
225	Student papers	Universitas Jambi on 2021-10-21	<1%
226	Student papers	Universitas Muhammadiyah Surakarta on 2013-04-10	<1%
227	Student papers	Universitas Pembangunan Panca Budi on 2026-08-02	<1%
228	Student papers	Universitas Putera Batam on 2026-01-30	<1%
229	Student papers	Universitas Sangga Buana YPKP on 2026-08-12	<1%
230	Internet	core.ac.uk	<1%
231	Internet	repository.mercubuana.ac.id	<1%
232	Internet	repository.usd.ac.id	<1%
233	Internet	www.passeidireto.com	<1%
234	Student papers	Erciyes Üniversitesi on 2026-07-08	<1%
235	Publication	Gema Otheliansyah, Raynal Yasni. "Pengaruh Penyaluran Dana Desa pada Indika...	<1%

236	Publication	Ichsani Mursidah, Remi Sanjaya, Bambang Yulianto, Dhian Sweetania, Puji Sularsi...	<1%
237	Publication	Jupriyanto -, Jamaludin Apandi, Anderias Eko Wijaya, Rian Hermawan et al. "Perb...	<1%
238	Student papers	Politeknik Negeri Cilacap on 2025-12-18	<1%
239	Publication	Putri Pratiwi, Dima Dwifa, Anita Desiani, Ali Amran, Bambang Suprihatin. "Klasifik...	<1%
240	Student papers	Sekolah Teknik Elektro & Informatika on 2026-07-20	<1%
241	Student papers	Sriwijaya University on 2020-04-13	<1%
242	Student papers	Tarumanagara University on 2024-12-10	<1%
243	Student papers	UIN Sultan Syarif Kasim Riau on 2021-02-02	<1%
244	Student papers	UPN Veteran Yogyakarta on 2026-03-31	<1%
245	Student papers	UPN Veteran Yogyakarta on 2026-07-02	<1%
246	Publication	Ummu Habibah Nur'Aini, Ratih HafSarah Maharrani, Andesita Prihantara, Riyadi ...	<1%
247	Student papers	Universitas Brawijaya on 2016-08-01	<1%
248	Student papers	Universitas Brawijaya on 2017-04-30	<1%
249	Student papers	Universitas Indonesia on 2026-07-03	<1%

250	Student papers	Universitas Islam Riau on 2026-05-10	<1%
251	Student papers	Universitas Negeri Semarang on 2022-07-25	<1%
252	Student papers	Universitas Negeri Semarang on 2025-07-16	<1%
253	Student papers	Universitas Pamulang on 2026-07-22	<1%
254	Student papers	Universitas Putera Batam on 2026-01-30	<1%
255	Internet	bip.wum.edu.pl	<1%
256	Internet	e-journal.hamzanwadi.ac.id	<1%
257	Internet	ejournal.unuja.ac.id	<1%
258	Internet	journal.ajirapublisher.id	<1%
259	Internet	mafiadoc.com	<1%
260	Internet	repositori.uma.ac.id	<1%
261	Internet	share.pdfonline.com	<1%
262	Internet	www.skamax.com	<1%
263	Publication	A. Reza Baehaga Jamroni Jamroni, Wahyu Hadikristanto, Muhamad Fatchan. "An...	<1%

264	Student papers	Erciyes Üniversitesi on 2026-07-09	<1%
265	Publication	Gagan Suganda, Marsani Asfi, Ridho Taufiq Subagio, Ricky Perdana Kusuma. "PE...	<1%
266	Publication	Maulana Fansyuri, Devi Yunita. "Analisis Kinerja Algoritma K-Nearest Neighbor d...	<1%
267	Publication	Muhamad Panji Riyansyah. "Penerapan Algoritma Apriori Dalam Data Mining Unt...	<1%
268	Publication	Muhammad Irvan Maulana, Savana Putra Aditama, Harun Al Rosyid. "Komparasi...	<1%
269	Publication	Mulia, Fina Adinda. "Peran Notaris Dalam Perjanjian Yang Dibuat Oleh Perseroan ...	<1%
270	Publication	Oktaviani Oktaviani, Endah Kurniasari, Rosdiana Rosdiana, Faturachman Kurniaw...	<1%
271	Student papers	Politeknik Manufaktur Negeri Bangka Belitung on 2024-10-31	<1%
272	Publication	Putri Tasya Agustina Tampubolon, T.Sofia Chairani, Debi Yandra Niska. "Sistem P...	<1%
273	Student papers	STT PLN on 2024-08-06	<1%
274	Publication	Sepriyan Gunawan, Rini Astuti, Willy Prihartono, Ryan Hamonangan. "PREDIKSI D...	<1%
275	Publication	Suci Mulyani, Afril Efan Pajri, Muhammad Fikram. "Klasifikasi Tingkat Kemiskina...	<1%
276	Student papers	Tarumanagara University on 2025-04-24	<1%
277	Student papers	UPN Veteran Yogyakarta on 2026-05-04	<1%

278	Student papers	UPN Veteran Yogyakarta on 2026-06-24	<1%
279	Student papers	Universitas Airlangga on 2021-03-01	<1%
280	Student papers	Universitas Brawijaya on 2017-07-13	<1%
281	Student papers	Universitas Brawijaya on 2021-04-20	<1%
282	Student papers	Universitas Gunadarma on 2025-12-08	<1%
283	Student papers	Universitas Kristen Duta Wacana on 2026-05-07	<1%
284	Student papers	Universitas Muhammadiyah Surakarta on 2015-03-10	<1%
285	Student papers	Universitas Negeri Manado on 2025-05-04	<1%
286	Student papers	Universitas Pendidikan Ganesha on 2024-07-12	<1%
287	Student papers	Universitas Putera Batam on 2025-07-25	<1%
288	Student papers	Universitas Putera Batam on 2026-02-02	<1%
289	Student papers	Universitas Putera Batam on 2026-02-03	<1%
290	Student papers	Universitas Sangga Buana YPKP on 2026-07-28	<1%
291	Student papers	Universitas Sangga Buana YPKP on 2026-08-11	<1%

292	Internet	beritamerdeka.net	<1%
293	Internet	ejurnal.methodist.ac.id	<1%
294	Internet	eprints.binadarma.ac.id	<1%
295	Internet	eprints3.upgris.ac.id	<1%
296	Internet	es.scribd.com	<1%
297	Internet	geograf.id	<1%
298	Student papers	iGroup on 2013-01-04	<1%
299	Internet	journal.ppkia.ac.id	<1%
300	Internet	journal.sinergi.or.id	<1%
301	Internet	ojs.stmik-banjarbaru.ac.id	<1%
302	Internet	repository.ibrahimy.ac.id	<1%
303	Internet	repository.pnj.ac.id	<1%
304	Internet	www.batamnews.co.id	<1%
305	Internet	www.science.gov	<1%

306	Publication	Anindya Khrisna Wardhani, Rano Indradi Sudra, Ega Nugraha, Astrid Novita Putri....	<1%
307	Publication	Arifin, Joko. "Pengaruh Program Kewirausahaan Digital dan Motivasi Berwirausa...	<1%
308	Publication	Dela Puspita Helmi, Jumadil Nangi, Asa Hari Wibowo. "Sistem Informasi Analisis K...	<1%
309	Student papers	FAKULTAS EKONOMI DAN BISNIS on 2026-05-04	<1%
310	Publication	M. Saleh, Dwi Yuli Prasetyo. "Analisis Sentimen Konsumen Terhadap Usaha Food ...	<1%
311	Publication	Tri Achya Ngasuko. "Faktor Yang Mempengaruhi Rumah Tangga Untuk Mengaks...	<1%
312	Student papers	UIN Raden Intan Lampung on 2026-05-13	<1%
313	Student papers	UNIVERSITAS BUDI LUHUR on 2025-09-09	<1%
314	Student papers	Universitas Brawijaya on 2019-12-06	<1%
315	Student papers	Universitas Islam Negeri Mataram on 2026-03-05	<1%
316	Student papers	Universitas Islam Riau on 2025-12-11	<1%
317	Student papers	Universitas Islam Riau on 2026-04-13	<1%
318	Student papers	Universitas Katolik Musi Charitas on 2026-08-11	<1%
319	Student papers	Universitas Muhammadiyah Palembang on 2025-07-06	<1%

320	Student papers	Universitas Muhammadiyah Surakarta on 2014-05-30	<1%
321	Student papers	Universitas Negeri Makassar on 2013-08-13	<1%
322	Student papers	Universitas Putera Batam on 2026-07-24	<1%
323	Publication	Zulfa Alviandri Mukharyahya, Yani Parti Astuti, Okta Nur Cahyani. "Perbandingan...	<1%
324	Publication	Achmad Hakim Qoirul Haq, Harminto Mulyo, Adi Sucipto. "Optimasi Kinerja Algor...	<1%
325	Publication	Charladinna Arvianne Reyvalda. "Analisis Kinerja YOLOv11 untuk Deteksi Makana...	<1%
326	Publication	Indra Dwi Rizqi, Retno Wardhani, Moh. Rosidi Zamroni. "Implementasi Algoritma ...	<1%
327	Student papers	LL Dikti IX Turnitin Consortium on 2026-05-22	<1%
328	Publication	Mhd Zahir Az Zikri, Muhammad Yasin Simargolang. "Klasifikasi Tingkat Tindak Pi...	<1%
329	Publication	Putri Aisha, M. Fathurahman, Surya Prangga. "IMPLEMENTASI METODE NEIGHBO...	<1%
330	Publication	Rinda Fitriana Azmi, Galih Hendro Martono, Neny Suliastianingsih. "Komparasi Al...	<1%
331	Publication	Siti Hajar, Nurdin Nurdin. "PERBANDINGAN KLASIFIKASI ALGORITMA DECISION T...	<1%
332	Student papers	Sriwijaya University on 2021-02-22	<1%
333	Student papers	Sriwijaya University on 2021-06-25	<1%

334	Student papers	State Islamic University of Alauddin Makassar on 2025-06-12	<1%
335	Student papers	UIN Maulana Malik Ibrahim Malang on 2020-07-06	<1%
336	Student papers	UIN Sultan Syarif Kasim Riau on 2021-08-12	<1%
337	Student papers	UNIVERSITAS BUDI LUHUR on 2026-01-22	<1%
338	Student papers	Universitas Atma Jaya Yogyakarta on 2017-09-07	<1%
339	Student papers	Universitas Diponegoro on 2023-07-04	<1%
340	Student papers	Universitas Islam Riau on 2021-12-01	<1%
341	Student papers	Universitas Islam Riau on 2026-06-19	<1%
342	Student papers	Universitas Jambi on 2026-05-12	<1%
343	Student papers	Universitas Negeri Manado on 2021-12-04	<1%
344	Student papers	Universitas Putera Batam on 2024-07-19	<1%
345	Student papers	Universitas Wijaya Kusuma Surabaya on 2026-07-21	<1%
346	Publication	Wan Dinulaqli, Nurdin Nurdin. "PERBANDINGAN PERFORMA NAÏVE BAYES DAN M...	<1%

10

**IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK
MENENTUKAN KELAYAKAN PENERIMA BANTUAN
LANGSUNG TUNAI DI DESA SUNGAI AMBAT**

SKRIPSI



Disusun Oleh :

KHUSNUL KHOTIMAH

403221010080

1

**PROGRAM STUDI INFORMASI
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS ISLAM INDRAGIRI
2026**

**IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK
MENENTUKAN KELAYAKAN PENERIMA BANTUAN
LANGSUNG TUNAI DI DESA SUNGAI AMBAT**

**IMPLEMENTATION OF THE NAIVE BAYES ALGORITHM
FOR DETERMINING THE ELIGIBILITY OF DIRECT CASH
ASSISTANCE (BLT) RECIPIENTS IN SUNGAI AMBAT
VILLAGE**

SKRIPSI

*Diajukan Sebagai Salah Satu Syarat Untuk Mendapatkan Gelar Sarjana Sistem
Informasi Strata (S1) Pada Fakultas Teknik Dan Ilmu Komputer
Universitas Islam Indragiri*



Disusun Oleh :

KHUSNUL KHOTIMAH

403221010080

**PROGRAM STUDI INFORMASI
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS ISLAM INDRAGIRI**

2026

PERSETUJUAN

IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK MENENTUKAN KELAYAKAN PENERIMA BANTUAN LANGSUNG TUNAI DI DESA SUNGAI AMBAT

yang dipersiapkan dan disusun oleh

KHUSNUL KHOTIMAH

403221010080

telah disetujui oleh Dosen Pembimbing Tugas Akhir

pada tanggal 24 Juli 2026

Pembimbing Utama

Pembimbing Pendamping

Fitri Yunita, S.SI., M.Kom

NIDN. 1012089001

Usman, ST., M.Kom

NIDN. 1017078301

Dekan Fakultas Teknik dan Ilmu Komputer

Dr. Bayu Rianto, S.SI., M.Kom

NIPY. 169109321

PENGESAHAN

IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK MENENTUKAN KELAYAKAN PENERIMA BANTUAN LANGSUNG TUNAI DI DESA SUNGAI AMBAT

Dipersiapkan dan disusun oleh

KHUSNUL KHOTIMAH

403221010080

telah diuji dan dipertahankan didepan Dewan Penguji

pada tanggal 24 Juli 2026

Susunan Dewan Penguji

Pembimbing Utama

Dewan Penguji I

Fitri Yunita, S.SI., M.Kom

NIDN. 1012089001

Prof.Dr.H. Abdullah, S. Si., M.Kom

NIDN. 1014039101

Pembimbing Pendamping

Dewan Penguji II

Usman, ST., M.Kom

NIDN. 1017078301

Muh. Rasyid Ridha, S.SI.,M.Kom

NIDN. 1013089001

Dewan Penguji III

Samsudin, S. Kom., M. Kom

NIDN.1009098501

Tugas Akhir ini telah diterima sebagai salah satu persyaratan
untuk menyelesaikan Mata Kuliah Tugas Akhir

Dekan Fakultas Teknik dan Ilmu Komputer

Ka. Prodi Sistem Informasi

Dr.Bayu Rianto, S.SI., M.Kom

NIPY. 169109321

Fitri Yunita,S.SI., M.Kom

NIPY.1590 10 293

1

PERNYATAAN

Saya yang bertanda tangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya (ASLI), dan ini dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis disuatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacuan dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu terkait naskah dan karya yang telah dibuat menjadi tanggung jawab saya pribadi.

Tembilahan , Juli 2026

Khusnul Khotimah

NIM. 403221010080

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas segala limpah rahmat, taufik, serta hidayah Nya sehingga penulis dapat menyelesaikan skripsi ini yang berjudul “Implementasi Algoritma Naive Bayes Untuk Menentukan Kelayakan Penerima Bantuan Langsung Tunai Di Desa Sungai Ambat” tepat pada waktunya. Shalawat serta salam semoga selalu tercurahkan kepada Nabi Muhammad SAW beserta keluarga dan para sahabatnya.

Penyusunan skripsi ini bertujuan untuk memenuhi salah satu persyaratan akademik pada program studi sistem informasi, fakultas teknik dan ilmu komputer, universitas islam indragiri. Selain itu, penelitian ini untuk memberikan kontribusi dalam upaya meningkatkan akurasi dan objektivitas proses penentuan kelayakan penerima bantuan langsung tunai (BLT) melalui penerapan metode data mining, khususnya algoritma Naive Bayes dan K-Nearest Neighbor(KNN).

Dalam proses penyusunan skripsi ini tidaklah mudah. Setiap tahapnya diwarnai dengan tantangan, pembelajaran, dan pengalaman baru yang memperkaya penulis, baik secara akademik maupun pribadi. Namun, berkat bimbingan, dukungan, dan doa dari berbagai pihak, skripsi ini akhirnya dapat terselesaikan dengan baik :

1. Bapak Dr. H. Najamuddin, Lc., M.A, selaku Rektor Universitas Islam Indragiri, yang telah memfasilitasi, dukungan, serta kesempatan berharga kepada penulis, sehingga proses penyusunan skripsi ini dapat berjalan dengan lancar hingga selesai.

2. Bapak Dr. Bayu Rianto, S.SI, M.Kom, selaku Dekan Fakultas Teknik dan Ilmu Komputer Universitas Islam Indragiri, atas dukungan, arahan, dan bimbingan yang telah membantu penulis selama menjalani perkuliahan.
3. Ibu Fitri Yunita, S.SI., M.Kom, selaku Ketua Program Studi Sistem Informasi sekaligus Pembimbing utama, yang telah memberikan bimbingan, perhatian, dukungan, serta masukan yang sangat berharga kepada penulis selama proses perkuliahan hingga penyusunan skripsi ini dapat terselesaikan dengan baik.
4. Bapak Usman, ST, M.Kom, selaku pembimbing pendamping, yang dengan penuh kesabaran dan ketulusan hati telah memberikan bimbingan, arahan, serta masukan yang sangat berarti bagi penulis hingga skripsi ini dapat terselesaikan dengan baik.
5. Bapak Samsudin, S.Kom, M.Kom, selaku Dosen Pembimbing Akademik, atas kesabaran, nasihat, serta pendampingannya yang tiada henti selama tujuh semester penulis menempuh studi hingga skripsi penyusunan proposal ini.
6. Seluruh dosen, staf, dan civitas akademika Program Studi Sistem Informasi Universitas Islam Indragiri, yang telah memberikan ilmu, bimbingan, serta pengalaman berharga selama masa perkuliahan, sehingga penulis dapat berkembang dan menyelesaikan studi ini dengan baik.
7. Aparatur Desa Sungai Ambat yang telah memberikan izin serta data pendukung yang diperlukan dalam penelitian ini.
8. Kedua orang tua dan keluarga tercinta, atas cinta, doa, dan pengorbanan yang tak ternilai, yang senantiasa menjadi penopang semangat dan kekuatan bagi penulis dalam melalui setiap langkah awal perjalanan Studi hingga tahap akhir penyusunan skripsi ini.

9. Teman seperjuangan, Elvi Suciati, Buana Mustikawati, Kamia Pertiwi, Ainur Rahmadina, dan Marliyah. Terima kasih sudah menemaniku selama masa perkuliahan. Terima kasih sudah menjadi teman, sahabat, dan saudara yang selalu menguatkan. Mengambil banyak peran penting dibalik layar, kebersamai dalam perjuangan dan tidak pernah mengeluh ketika direpotkan. Semoga sama-sama dilancarkan sampai akhir perjuangan. See u on top guys!

Penulis berharap proposal ini dapat memberikan manfaat, baik dalam pengembangan ilmu pengetahuan khususnya pada bidang data mining maupun dalam penerapan teknologi untuk mendukung penyaluran bantuan sosial yang lebih tepat sasaran. Penulis juga menyadari bahwa skripsi ini masih jauh dari sempurna, sehingga kritik dan saran yang membangun sangat diharapkan untuk kesempurnaan penelitian selanjutnya.

Tembilahan 17 Desember 2025
Penulis

Khusnul Khotimah
403221010080

1

DAFTAR ISI

HALAMAN JUDUL

PERSETUJUAN

PENGESAHAN

PERNYATAAN

KATA PENGANTARv

DAFTAR ISI viii

DAFTAR TABEL ix

DAFTAR GAMBARx

INTISARI xiii

ABSTRACT xiv

BAB I PENDAHULUAN

1.1 Latar Belakang1

1.2 Rumusan Masalah3

1.3 Batasan Masalah4

1.4 Tujuan Penelitian4

1.5 Manfaat Penelitian5

1.6 Sistematika Penulisan5

BAB II TINJAUAN PUSTAKA

2.1 Kemiskinan9

2.2 Bantuan Langsung Tunai (BLT)10

2.3 Data Mining11

2.4 Klasifikasi dalam Data Mining12

233

2.5 Algoritma Naive Bayes	14
2.6 Algoritma K-Nearest Neighbor (KNN)	15
2.7 Google Colab	17
2.8 Penelitian Terdahulu	18
2.9 Variabel Penelitian	27
2.10 Rangkuman	30
BAB III METODE PENELITIAN	
3.1 Kerangka Penelitian	32
3.2 Alat dan Perangkat Penelitian	43
BAB IV HASIL DAN PEMBAHASAN	
4.1 Data Awal	47
4.2 Preprocessing Data	50
4.3 Implementasi Algoritma.....	64
4.4 Perbandingan Performa Algoritma	87
4.5 Uji Statistik Paired T-Test.....	88
BAB V PENUTUP	
5.1 Kesimpulan	92
5.3 Saran	93
DAFTAR PUSTAKA	
LAMPIRAN	

DAFTAR TABEL

Tabel 2.1 Riview Jurnal Terdahulu	18
Tabel 3.2 Variabel Masukan	34
Tabel 3.3 Variabel Output	35
Tabel 4.1 Data Awal Penelitian	49
Tabel 4.2 Data Yang Telah Di Encoding	53
Tabel 4.3 Data Telah Di Normalisasi.....	56
Tabel 4.4 Distribusi Kelas Keseluruhan.....	61
Tabel 4.5 Pembagian Data Training dan Data Testing	61
Tabel 4.6 Distribusi Data Training.....	62
Tabel 4.7 Distribusi Data Testing	62
Tabel 4.8 Hasil Likelihood	67
Tabel 4.9 Hasil Prdikasi	70
Tabel 4.10 Classification Report Naive Bayes	76
Tabel 4.11 Classification Report KNN	85
Tabel 4.12 Perhitungan Model	87
Tabel 4.13 Uji Statistik Paired T-Test.....	88

174

DAFTAR GAMBAR

100

113

50

Gambar 3.1 Alur Kerangka Penelitian	33
Gambar 3.2 Alur Kerja Naive Bayes	38
Gambar 4.1 Diagram Data Training dan Data Testing	62
Gambar 4.2 Heatmap korelasi Antar Fitur	63
Gambar 4.3 Diagram Jumlah Status Kelayakan	71
Gambar 4.4 Confusion Matrix Naive Bayes	72
Gambar 4.5 Grafik Akurasi Berdasarkan Nilai k.....	80
Gambar 4.6 Confusion Matrix KNN	80
Gambar 4.7 Boxplot Perbandingan Akurasi	89
Gambar 4.8 Rata Rata Akurasi Metode	90

INTISARI

Penentuan kelayakan penerima Bantuan Langsung Tunai (BLT) memerlukan proses yang objektif agar penyaluran bantuan tepat sasaran. Penelitian ini bertujuan mengimplementasikan algoritma Naive Bayes untuk menentukan kelayakan penerima BLT di Desa Sungai Ambat serta membandingkan kinerjanya dengan algoritma K-Nearest Neighbor (KNN). Data penelitian diperoleh dari Kantor Desa Sungai Ambat sebanyak 524 data warga yang terdiri atas 13 variabel independen dan satu variabel dependen. Penelitian menggunakan pendekatan Knowledge Discovery in Databases (KDD) melalui tahapan data cleaning, transformasi data, seleksi fitur, pembagian data latih dan data uji dengan rasio 80:20, pemodelan, serta evaluasi menggunakan confusion matrix, accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa algoritma Naive Bayes memberikan performa lebih baik dengan accuracy 96%, precision 92%, recall 92%, dan F1-score 92%, sedangkan algoritma KNN memperoleh accuracy 92%, precision 96%, recall 69%, dan F1-score 76%. Hasil paired t-test menunjukkan adanya perbedaan performa yang signifikan antara kedua algoritma. Dengan demikian, algoritma Naive Bayes lebih efektif dalam menentukan kelayakan penerima BLT sehingga dapat mendukung pengambilan keputusan yang lebih objektif, akurat, dan tepat sasaran.

Kata Kunci: Bantuan Langsung Tunai, Data Mining, Naive Bayes, K-Nearest Neighbor, Klasifikasi.

ABSTRACT

Determining the eligibility of Direct Cash Assistance (Bantuan Langsung Tunai/BLT) recipients requires an objective process to ensure that social assistance is distributed accurately and fairly. This study aims to implement the Naive Bayes algorithm to determine the eligibility of BLT recipients in Sungai Ambat Village and to compare its performance with the K-Nearest Neighbor (KNN) algorithm. The research data were obtained from the Sungai Ambat Village Office, consisting of 524 citizen records with 13 independent variables and one dependent variable. The study employed the Knowledge Discovery in Databases (KDD) approach, which includes data cleaning, data transformation, feature selection, an 80:20 training-testing data split, model development, and performance evaluation using a confusion matrix, accuracy, precision, recall, and F1-score. The results indicate that the Naive Bayes algorithm outperformed the K-Nearest Neighbor algorithm, achieving an accuracy of 96%, precision of 92%, recall of 92%, and an F1-score of 92%, while the KNN algorithm achieved an accuracy of 92%, precision of 96%, recall of 69%, and an F1-score of 76%. Furthermore, the paired t-test demonstrated a statistically significant difference in performance between the two algorithms. Therefore, the Naive Bayes algorithm is more effective in determining the eligibility of BLT recipients and can support more objective, accurate, and targeted decision-making.

Keywords: Direct Cash Assistance (BLT), Data Mining, Naive Bayes, K-Nearest Neighbor, Classification.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kemiskinan merupakan sebuah masalah yang amat serius yang masih dihadapi oleh beberapa negara di dunia, termasuk indonesia. Masalah kemiskinan tidak hanya berkaitan dengan rendahnya pendapatan masyarakat, tetapi juga melibatkan berbagai aspek kehidupan seperti pendidikan, kesehatan, lapangan pekerjaan, dan akses terhadap sumber daya ekonomi. Tingginya tingkat kemiskinan dapat menghambat pembangunan nasional dan menurunkan kualitas hidup masyarakat, sehingga diperlukan berbagai upaya strategis dan kelanjutan untuk mengatasinya. Oleh karena itu, maka pemerintah membuat kebijakan ataupun program untuk menanggulangi masalah tersebut. Salah satu dari sekian banyak program tersebut adalah Bantuan Langsung Tunai (BLT). Bantuan Langsung Tunai (BLT) dapat dipahami sebagai pemberian sejumlah uang (dana tunai) kepada masyarakat miskin setelah pemerintah menaikkan harga BBM dengan mengurangi subsidi BBM yang kemudian selisih subsidi tersebut diberikan kepada masyarakat miskin[1].

Pemerataan kesejahteraan sosial merupakan salah satu prioritas penting dalam pembangunan nasional, terutama dalam rangka membantu masyarakat kurang mampu agar dapat terpenuhi kebutuhan dasarnya[2]. Bantuan langsung tunai (BLT) merupakan salah satu program pemerintah yang bertujuan untuk mengurangi kemiskinan dan meningkatkan kesejahteraan masyarakat dengan memberikan bantuan secara langsung kepada keluarga yang berhak[3]. Dari kondisi tersebut, maka pemerintah membuat kebijakan ataupun program untuk menanggulangi masalah yang terjadi. Salah satunya yaitu program Bantuan Langsung Tunai (*cash*

transfers) atau sering disingkat BLT adalah program bantuan pemerintah berjenis pemberian uang tunai atau berupa bantuan lainnya baik bersyarat (*conditional cash transfer*) maupun tidak bersyarat (*unconditional cash transfer*) untuk masyarakat miskin supaya dapat meningkatkan hidup yang lebih baik[4].

Namun, penentuan kelayakan penerima BLT sering kali menghadapi tantangan seperti data yang tidak terorganisir dengan baik, subyektivitas dalam proses verifikasi, dan terbatasnya sumber daya manusia untuk melakukan seleksi secara efisien. Hal ini dapat menyebabkan kurang tepat sasaran, dimana bantuan bisa saja diterima oleh pihak yang kurang berhak atau sebaliknya, masyarakat yang sangat membutuhkan justru tidak mendapatkan bantuan tersebut.

Dalam era digital dan big data saat ini, penerapan metode-metode kecerdasan buatan, khususnya algoritma klasifikasi seperti Naive Bayes, dapat menjadi solusi efektif untuk membantu menentukan kelayakan penerima BLT secara otomatis dan lebih akurat. Algoritma Naive Bayes dikenal memiliki keunggulan dalam menangani data kategorikal dan bekerja dengan baik meskipun data pelatihan terbatas, sehingga cocok digunakan dalam konteks penentuan status kelayakan berdasarkan sejumlah variabel sosial ekonomi[5].

Di Desa Sungai Ambat, pengimplementasian algoritma Naive Bayes dan K-Nearest Neighbor (KNN) diharapkan dapat membantu aparat desa dalam melakukan seleksi penerima BLT. Melalui analisis data sosial ekonomi masyarakat, algoritma ini dapat membantu dengan memperhatikan data-data kondisi keluarga seperti pendapatan, jumlah anggota keluarga, status pekerjaan, dan faktor relevan lainnya. Dengan begitu, proses pendataan penerima bantuan menjadi lebih objektif, transparan, dan dapat dipertanggungjawabkan.

Untuk mengatasi permasalahan tersebut, diperlukan suatu metode yang dapat membantu dalam pengambilan keputusan secara objektif dan efisien. Algoritma Naive Bayes merupakan salah satu teknik dalam machine learning yang dapat digunakan untuk klasifikasi data berdasarkan probabilitas[6]. Dengan menggunakan algoritma ini, diharapkan dapat membantu pemerintah desa dalam menentukan kelayakan penerima BLT secara lebih akurat dan cepat berdasarkan data yang tersedia.

Oleh karena itu, penelitian ini bertujuan untuk mengimplementasikan serta membandingkan kinerja algoritma Naive Bayes dan K-Nearest Neighbor (KNN) dalam menentukan kelayakan penerima Bantuan Langsung Tunai di Desa Sungai Ambat. Perbandingan ini dilakukan untuk mengetahui algoritma, sehingga proses penyaluran bantuan dapat berjalan lebih tepat sasaran dan memberikan manfaat yang optimal bagi masyarakat yang membutuhkan.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah di atas maka masalah penelitian ini dapat dikemukakan sebagai berikut:

1. Bagaimana penerapan algoritma Naive Bayes dan K- Nearest Neighbor (KNN) dalam klasifikasi kelayakan penerima bantuan langsung tunai (BLT)?
2. Bagaimana hasil klasifikasi kelayakan penerima bantuan langsung tunai menggunakan algoritma Naive Bayes dan K- Nearest Neighbor (KNN) berdasarkan data sosial ekonomi warga?
3. Bagaimana perbandingan efektivitas algoritma Naive Bayes dan K-Nearest Neighbor (KNN) dalam menentukan kelayakan penerima BLT?

1.3 Batasan Penelitian

Agar penelitian ini lebih terpadu serta mudah dalam pembahasan, maka penulis memberikan pembatasan masalah yaitu sebagai berikut:

1. Data yang digunakan terbatas pada data sosial ekonomi warga, seperti usia kepala keluarga, jenis kelamin kepala keluarga, jumlah anggota rumah tangga, pendidikan kepala keluarga, pendapatan bulanan, status pekerjaan kepala keluarga, penerima bantuan sosial lain, kepemilikan aset dasar, kelayakan kondisi rumah, akses air bersih, akses listrik, keberadaan lansia, keberadaan penyandang disabilitas.
2. Algoritma yang digunakan untuk klasifikasi adalah Naive Bayes, dan hanya membandingkan dengan satu algoritma K-Nearest Neighbor (KNN).
3. Penentuan kelayakan difokuskan pada proses seleksi penerima BLT di tingkat desa, tanpa memperhitungkan kebijakan pemerintah pusat atau faktor eksternal lain.
4. Penelitian tidak mencakup pengembangan sistem atau aplikasi berbasis web maupun mobile, melainkan hanya sampai pada tahap analisis dan penerapan algoritma Naive Bayes dan K-Nearest Neighbor (KNN).

1.4 Tujuan Penelitian

Tujuan dari penelitian ini untuk memperoleh pengetahuan dan wawasan agar dapat memecahkan masalah sesuai rumusan masalah di atas:

1. Mengimplementasikan algoritma Naive Bayes dan K-Nearest Neighbor(KNN) dalam proses klasifikasi kelayakan penerima bantuan langsung tunai (BLT).

2. Mengklasifikasi data sosial ekonomi warga menggunakan algoritma Naive Bayes agar proses penentuan penerima BLT menjadi lebih objektif dan efisien.
3. Melakukan uji efektivitas dengan membandingkan hasil algoritma Naive Bayes dengan metode K-Nearest Neighbor menggunakan uji statistik paired t-test untuk mengetahui metode yang paling optimal dan tepat sasaran.

1.5 Manfaat Penelitian

Penelitian yang dilakukan penulis akan memberikan beberapa manfaat sebagai berikut:

1. Memberikan metode yang efisien dan akurat dalam menentukan kelayakan penerima Bantuan Langsung Tunai sehingga dapat mempercepat proses seleksi dan penyaluran bantuan.
2. Membantu aparat desa dalam membuat keputusan yang lebih objektif dan transparan terkait penerima bantuan sosial berdasarkan data yang valid dan terstruktur.
3. Mengurangi konflik sosial atau kecemburuan antarwarga, melalui mekanisme penilaian kelayakan yang objektif dan terukur.
4. Mengurangi potensi kesalahan penyaluran bantuan sehingga penerima yang benar-benar membutuhkan dapat tersentuh dan mengurangi potensi penyalahgunaan bantuan.

1.6 Sistematika Penulisan

Agar pembahasan dalam skripsi ini lebih terstruktur dan mudah di pahami, penulisan disusun ke dalam beberapa bab dengan sistematika sebagai berikut:

BAB I PENDAHULUAN

Berisi latar belakang penelitian, rumusan masalah, batas masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan skripsi.

BAB II TINJAUAN PUSTAKA

Memuat landasan teori yang mendukung penelitian, kajian literatur terkait dengan penerima bantuan langsung tunai, data mining, serta penjelasan algoritma Naive Bayes. Selain itu, bab ini juga mencakup penelitian-penelitian terdahulu yang relevan sebagai dasar dalam merumuskan kerangka pemikiran penelitian.

BAB III METODOLOGI PENELITIAN

Menjelaskan pendekatan penelitian yang digunakan, sumber data, teknik pengumpulan data melalui observasi dan wawancara di Kantor Desa Sungai Ambat, variabel penelitian, tahap pengolahan data, serta metode menganalisis menggunakan algoritma Naive Bayes dan K-Nearest Neighbor(KNN).

BAB IV HASIL PENELITIAN

Bab ini menyajikan hasil pelaksanaan penelitian mulai dari proses pengolahan data, preprocessing, pembentukan model menggunakan algoritma Naive Bayes dan K-Nearest Neighbor (KNN), hasil pengujian model menggunakan metrik evaluasi seperti Accuracy, Precision, Recall, F1-Score, dan Confusion Matrix, serta hasil perbandingan performa kedua algoritma.

73

BAB V PENUTUP

139

Bab ini berisi kesimpulan yang diperoleh berdasarkan hasil penelitian dan pembahasan yang telah dilakukan. Selain itu, bab ini juga memuat saran yang diharapkan dapat menjadi masukan bagi pemerintah desa maupun peneliti selanjutnya dalam mengembangkan penelitian mengenai klasifikasi penerima bantuan tunai (BLT) dengan metode data mining.

BAB II

TINJAUAN PUSTAKA

Kajian pustaka memiliki peranan yang sangat penting dalam suatu penelitian ilmiah karena berfungsi sebagai landasan teoretis yang memperkuat dasar konseptual maupun metodologis dari penelitian yang dilakukan. Melalui kajian pustaka, peneliti dapat memperoleh pemahaman mendalam mengenai sejauh mana topik penelitian telah dikaji oleh penelitian-penelitian sebelumnya, mengidentifikasi kesenjangan penelitian (research gap), serta menentukan posisi penelitian yang sedang dilakukan dalam konteks keilmuan yang lebih luas. Dengan demikian, penyusunan landasan teori menjadi aspek yang esensial untuk memastikan bahwa penelitian memiliki arah yang terarah, relevan, serta memberikan kontribusi nyata terhadap pengembangan ilmu pengetahuan.

Dalam konteks penyaluran Bantuan Langsung Tunai (BLT), salah satu tantangan signifikan adalah penentuan penerima bantuan yang layak secara objektif dan akurat. Proses penentuan ini sering kali masih dilakukan secara manual dan subjektif, sehingga berpotensi menimbulkan kesalahan sasaran serta ketidakadilan dalam distribusi bantuan. Oleh karena itu, diperlukan pendekatan berbasis data yang dapat mendukung klasifikasi atau prediksi kelayakan penerima BLT.

Salah satu metode dalam bidang data mining yang dapat diterapkan untuk proses klasifikasi adalah algoritma Naive Bayes. Algoritma ini beroperasi berdasarkan prinsip probabilitas yang mengacu pada Teorema Bayes, serta memiliki kemampuan untuk memproses data berukuran kecil hingga menengah dengan cepat dan efisien. Algoritma Naive Bayes juga dikenal sebagai metode yang mudah diimplementasikan dan memberikan hasil yang cukup akurat, khususnya

dalam kasus-kasus yang melibatkan data kategorikal seperti data sosial dan ekonomi masyarakat desa.

Bab ini akan mengelaborasi teori-teori yang mendukung penelitian ini, mulai dari konsep dasar Bantuan Langsung Tunai, pengantar data mining dan proses klasifikasi, penjelasan mendalam mengenai algoritma Naive Bayes, serta metode evaluasi kinerja model klasifikasi. Selain itu, bab ini juga mengkaji beberapa penelitian terdahulu yang relevan sebagai bahan perbandingan dan landasan dalam pengembangan penelitian ini. Dengan tinjauan pustaka ini, diharapkan penelitian yang dilakukan memiliki arah yang jelas serta dapat memberikan kontribusi signifikan dalam upaya meningkatkan ketepatan sasaran penyaluran bantuan sosial di Desa Sungai Ambat.

2.1 Kemiskinan

Kemiskinan merupakan kondisi dimana seseorang atau sekelompok orang tidak mampu memenuhi kebutuhan dasar lain nya secara layak, seperti kebutuhan pangan, sandang, papan , pendidikan, dan kesehatan[7]. Menurut Badan Pusat Statistik (BPS), Kemiskinan diartikan sebagai ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan dasar yang diukur dari sisi pengeluaran. Kemiskinan bukan hanya disebabkan oleh faktor rendahnya pendapatan, tetapi juga oleh rendahnya akses terhadap pendidikan, pekerjaan, serta sarana dan prasarana ekonomi. Oleh karena itu, penanganan kemiskinan memerlukan pendekatan yang menyeluruh, baik dari sisi kebijakan ekonomi, sosial, maupun teknologi.

Pemerintah Indonesia telah melakukan berbagai upaya untuk menanggulangi kemiskinan, salah satunya dengan memberikan bantuan sosial kepada masyarakat miskin dalam bentuk bantuan langsung tunai (BLT). Program ini bertujuan

agar masyarakat miskin tetap dapat memenuhi kebutuhan hidup sehari-hari serta menjaga stabilitas ekonomi nasional ditengah kondisi kritis atau inflasi harga barang kebutuhan pokok.

2.2 Bantuan Langsung Tunai (BLT)

Bantuan Langsung Tunai (BLT) merupakan salah satu bentuk program kebijakan sosial yang dirancang oleh pemerintah untuk memberikan bantuan berupa uang tunai secara langsung kepada masyarakat berpenghasilan rendah atau kelompok rentan[8]. Program ini bertujuan untuk meningkatkan daya beli masyarakat serta mengurangi beban ekonomi yang timbul akibat berbagai kondisi yang memengaruhi kestabilan ekonomi, seperti krisis ekonomi, kenaikan harga bahan pokok, maupun bencana alam. Melalui penyaluran bantuan secara langsung, BLT diharapkan dapat memberikan dampak nyata terhadap kesejahteraan masyarakat dengan cara memperkuat daya tahan ekonomi rumah tangga miskin dan menjaga stabilitas sosial di tengah tekanan ekonomi nasional.

Bantuan Langsung Tunai (BLT) merupakan program bantuan sosial yang diselenggarakan oleh pemerintah melalui transfer langsung dana tunai kepada individu atau keluarga yang memenuhi kriteria spesifik[9]. Program ini dirancang untuk menyediakan dukungan finansial yang cepat dan adaptif bagi kelompok masyarakat yang terpengaruh oleh krisis ekonomi, termasuk inflasi, pandemi, atau peningkatan harga bahan pokok. Di Indonesia, BLT umumnya dihubungkan dengan inisiatif seperti BLT BBM (Bahan Bakar Minyak) atau BLT Desa, yang dikelola oleh Kementerian Sosial atau entitas terkait[10].

Dalam konteks jaminan sosial, BLT berfungsi sebagai bentuk perlindungan sosial yang memungkinkan pemerintah memenuhi tanggung jawabnya dalam

mendukung masyarakat yang membutuhkan. BLT tidak hanya memberikan bantuan finansial sementara, tetapi juga berperan sebagai instrumen dalam mengurangi kemiskinan dan meningkatkan kesejahteraan masyarakat, terutama pada masa krisis. Dengan demikian, program ini secara langsung berkontribusi pada upaya pencapaian Tujuan meningkatkan kesejahteraan masyarakat miskin. Mengurangi beban ekonomi keluarga penghasilan rendah, memastikan pemerataan bantuan sosial agar tepat sasaran.

2.3 Data Mining

Data mining adalah proses menemukan pola atau informasi menarik dari data dalam jumlah besar yang disimpan dalam database, data warehouse, atau tempat penyimpanan lainnya[11]. Data mining menggabungkan teknik dari berbagai bidang seperti statistik, pembelajaran mesin, kecerdasan buatan, dan basis data untuk mengekstraksi informasi yang bermanfaat dan relevan secara otomatis atau semiotomatis. Tujuannya adalah untuk menemukan pola tersembunyi yang dapat mendukung pengambilan keputusan strategis dan memberikan nilai tambah ekonomi. Proses ini merupakan inti dari *Knowledge Discovery in Databases (KDD)*, yang melibatkan tahapan seleksi data, pre-processing, transformasi, data mining, dan interpretasi hasil[12].

Data mining merupakan proses penemuan pola atau informasi yang berguna dari kumpulan data yang besar. Data mining adalah proses mengekstraksi pengetahuan yang sebelumnya tidak diketahui, namun berpotensi berguna, dari data yang tersimpan dalam basis data. Tujuan utama data mining adalah untuk mengidentifikasi pola tersembunyi dan hubungan antara variabel dalam suatu dataset. Beberapa tahapan dalam proses data mining meliputi:

1. Pembersihan data (data cleaning) : menghilangkan data yang tidak relevan atau mengandung kesalahan.
2. Integrasi data (data integration) : menggabungkan data dari berbagai sumber.
3. Seleksi data (data selection) : memilih data yang relevan untuk analisis.
4. Transformasi data (data transformation) : mengubah data menjadi format yang sesuai untuk pemrosesan.
5. Proses mining (data mining process) : penerapan algoritma untuk menemukan pola.
6. Evaluasi dan interpretasi (pattern evaluation) : menilai pola yang ditemukan agar bermanfaat bagi pengambilan keputusan.

Salah satu cabang dari data mining adalah klasifikasi, yaitu proses memprediksi katagori atau kelas dari suatu data berdasarkan atribut yang ada. Dalam kontes penelitian ini, klasifikasi digunakan untuk menentukan apakah seseorang layak atau tidak layak menerima BLT berdasarkan data sosial ekonominya.

2.4 Klasifikasi dalam Data Mining

Klasifikasi merupakan salah satu teknik utama dalam data mining yang digunakan untuk mengelompokkan atau memprediksi suatu data ke dalam kelas atau kategori tertentu berdasarkan pola yang telah dipelajari dari data sebelumnya[13]. Proses klasifikasi dilakukan dengan cara membangun model prediktif yang diperoleh dari data latih (training data), di mana setiap data memiliki atribut-atribut tertentu (fitur) dan label kelas yang telah diketahui. Model yang dihasilkan kemudian digunakan untuk menentukan kelas dari data baru yang belum diketahui kategorinya.

Klasifikasi adalah proses menemukan model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep, dengan tujuan untuk dapat memprediksi label kelas dari objek yang belum diketahui kelasnya. Model klasifikasi ini dapat direpresentasikan dalam berbagai bentuk, seperti aturan klasifikasi (classification rules), pohon keputusan (decision tree), rumus matematis, atau jaringan probabilistik seperti Naive Bayes.

Dalam konteks data mining, proses klasifikasi terdiri atas dua tahap utama, yaitu:

1. Tahap pembelajaran (learning phase) : sistem mempelajari hubungan antara atribut-atribut dalam data dengan label kelas yang telah diketahui, untuk membangun model klasifikasi.
2. Tahap pengujian (testing phase) : model yang telah dibangun diuji menggunakan data uji (testing data) untuk mengukur akurasi dan kemampuan model dalam memprediksi kelas data baru.

Klasifikasi banyak digunakan dalam berbagai bidang, seperti prediksi kelayakan kredit, diagnosis medis, pengenalan pola, dan termasuk dalam penentuan kelayakan penerima Bantuan Langsung Tunai (BLT). Melalui klasifikasi, sistem dapat menganalisis faktor-faktor seperti pendapatan, pekerjaan, jumlah tanggungan, dan kondisi rumah, kemudian mengkategorikan setiap warga ke dalam kelompok layak atau tidak layak menerima bantuan sosial secara objektif dan terukur.

Dengan demikian, klasifikasi dalam data mining berperan penting dalam mendukung proses pengambilan keputusan berbasis data (data-driven decision making), karena mampu mengekstraksi pola tersembunyi dari data historis dan

menggunakannya untuk memprediksi keputusan di masa depan dengan tingkat akurasi yang dapat diukur.

2.5 Algoritma Naive Bayes

Algoritma Naive Bayes merupakan salah satu metode klasifikasi dalam data mining yang berlandaskan pada teori probabilitas Bayes. Metode ini digunakan untuk memprediksi kelas atau kategori dari suatu data berdasarkan nilai probabilitas tertinggi yang dihitung dari data pelatihan (training data). Disebut naive karena algoritma ini mengasumsikan bahwa setiap fitur atau atribut dalam data bersifat independen secara kondisional terhadap fitur lainnya, meskipun dalam kenyataannya hal tersebut tidak selalu benar[14]. Algoritma ini menggunakan teorema Bayes dengan asumsi bahwa setiap fitur (atribut) bersifat independen terhadap fitur lainnya.

Meskipun demikian, model ini tetap terbukti mampu memberikan performa klasifikasi yang baik dalam berbagai aplikasi nyata. Naive Bayes bekerja dengan menghitung probabilitas prior dari setiap kelas serta probabilitas likelihood dari masing-masing atribut terhadap kelas tersebut. Selanjutnya, probabilitas posterior dihitung menggunakan Teorema Bayes, dan kelas dengan nilai probabilitas terbesar akan dipilih sebagai hasil prediksi.

Pendekatan probabilistik ini menjadikan Naive Bayes sebagai algoritma yang sederhana namun sangat efisien, karena tidak memerlukan proses pelatihan yang kompleks. Selain itu, Naive Bayes mampu bekerja optimal pada data berukuran besar maupun data dengan fitur yang beragam, terutama ketika fitur-fitur tersebut berupa kategori. Kecepatan komputasi, ketahanan terhadap noise, dan kemampuan dalam menangani data berdimensi tinggi menjadikan algoritma ini

banyak digunakan dalam berbagai bidang seperti deteksi spam, text mining, analisis sentimen, serta klasifikasi data sosial ekonomi. Dengan karakteristik tersebut, Naive Bayes menjadi salah satu metode klasifikasi yang sering dipilih dalam penelitian maupun implementasi sistem pendukung keputusan yang memerlukan proses klasifikasi cepat dan hasil yang konsisten.

Kelebihan Naive Bayes:

1. Cepat dan efisien untuk dataset besar.
2. Dapat bekerja dengan baik pada data kategori.
3. Tidak memerlukan parameter kompleks.

Kelemahan Naive Bayes:

1. Asumsi independensi antar atribut sering tidak sesuai dengan kenyataan.
2. Tidak cocok untuk data dengan hubungan antar variabel yang kuat.

Algoritma Naive Bayes banyak digunakan karena memiliki kecepatan tinggi dalam proses pelatihan dan prediksi, serta kinerja yang baik pada data berukuran besar. Selain itu, metode ini efektif pada data dengan variabel kategori, seperti data sosial ekonomi untuk penentuan penerima bantuan sosial[15]. Meskipun demikian, algoritma ini memiliki kelemahan utama pada asumsi independensi antar atribut, yang dapat mengurangi akurasi jika antar variabel terdapat hubungan yang kuat. Oleh karena itu, dalam beberapa penelitian, Naive Bayes sering dikombinasikan dengan metode lain untuk meningkatkan performa, seperti feature selection atau ensemble learning[16].

2.6 Algoritma K-Nearest Neighbor(KNN)

K-Nearest Neighbor (K-NN) merupakan salah satu algoritma dalam bidang machine learning yang termasuk ke dalam metode supervised learning, yaitu

metode pembelajaran yang menggunakan data berlabel sebagai dasar untuk melakukan proses prediksi atau klasifikasi terhadap data baru[17]. Algoritma ini bekerja dengan prinsip bahwa suatu objek atau data yang memiliki karakteristik serupa cenderung berada pada kelompok atau kelas yang sama. Oleh karena itu, proses penentuan kelas suatu data dilakukan berdasarkan kedekatannya dengan sejumlah data tetangga terdekat yang terdapat pada data pelatihan (*training data*).

Dalam implementasinya, pengukuran kedekatan antar data umumnya menggunakan beberapa metode perhitungan jarak, seperti Euclidean Distance, Manhattan Distance, dan Minkowski Distance. Di antara metode tersebut, Euclidean Distance merupakan metode yang paling sering digunakan karena mampu mengukur jarak lurus antara dua titik dalam ruang multidimensi[18]. Semakin kecil nilai jarak yang dihasilkan, maka semakin tinggi tingkat kemiripan antara data yang dibandingkan. K-Nearest Neighbor termasuk algoritma yang sederhana namun efektif karena tidak memerlukan proses pembentukan model yang kompleks. Seluruh data pelatihan disimpan dan digunakan secara langsung pada saat proses prediksi dilakukan. Oleh karena itu, algoritma ini sering digunakan dalam berbagai bidang seperti klasifikasi data, pengenalan pola, sistem rekomendasi, data mining, dan analisis prediktif[19].

Keunggulan utama K-NN terletak pada kemudahan implementasi dan kemampuannya menghasilkan tingkat akurasi yang baik pada dataset yang sesuai. Namun, performa algoritma ini sangat dipengaruhi oleh pemilihan nilai k , ukuran dataset, serta metode pengukuran jarak yang digunakan. Algoritma K-NN dapat didefinisikan sebagai metode klasifikasi non-parametrik yang menentukan kategori suatu objek berdasarkan mayoritas kelas dari sejumlah tetangga terdekatnya dalam

ruang fitur (*feature space*). Karena kesederhanaan dan efektivitasnya, algoritma ini masih menjadi salah satu metode yang banyak digunakan sebagai dasar perbandingan dalam penelitian-penelitian di bidang *data mining*, *machine learning*, dan *artificial intelligence*.

2.7 Google Colab

Google Colab (Google Colaboratory) adalah sebuah layanan berbasis komputasi awan (*cloud computing*) yang dikembangkan oleh Google untuk menyediakan lingkungan pemrograman interaktif berbasis *Jupyter Notebook*[20]. Platform ini memungkinkan pengguna untuk menulis, menjalankan, dan mendokumentasikan kode program, khususnya bahasa Python, secara langsung melalui peramban (*browser*) tanpa memerlukan instalasi perangkat lunak tambahan pada komputer lokal[21]. Dengan memanfaatkan infrastruktur komputasi berbasis awan milik google, pengguna dapat menjalankan berbagai proses komputasi dengan lebih efisien serta mengakses proyek yang sedang dikerjakan dari berbagai perangkat yang memiliki koneksi internet.

Google Colab berfungsi sebagai sarana yang mendukung kegiatan pembelajaran, penelitian, analisis data, serta pengembangan Perangkat lunak, implementasi berbagai metode dalam bidang *machine learning*, *deep learning*, *data mining* dan *artificial intelligence* (AI)[22]. Platform ini menyediakan berbagai pustaka (*library*) Python yang umum digunakan dalam bidang ilmu data, seperti NumPy, Pandas, Matplotlib, TensorFlow, dan Scikit-learn, sehingga memudahkan pengguna dalam melakukan pengolahan data dan eksperimen komputasi. Salah satu keunggulan utama Google Colab adalah kemampuannya menyediakan akses ke sumber daya komputasi berperforma tinggi, seperti Graphics Processing Unit

(GPU) dan Tensor Processing Unit (TPU), yang dapat digunakan untuk mempercepat proses komputasi tanpa memerlukan perangkat keras khusus[23]. Selain itu, Google Colab terintegrasi dengan Google Drive sehingga memudahkan pengguna dalam menyimpan, mengelola, dan berbagi dokumen notebook dan dataset secara daring. Dengan berbagai fitur yang dimilikinya, Google Colab menjadi platform yang efektif untuk mendukung kegiatan akademik maupun profesional, terutama dalam bidang pemrograman, data science, data mining, dan kecerdasan buatan. Kemudahan akses, fleksibilitas penggunaan, serta dukungan impor data dari berbagai sumber, seperti file lokal, Google Drive, GitHub, maupun repositori data lainnya. Google Colab sebagai salah satu alat yang banyak digunakan dalam lingkungan pendidikan dan penelitian modern[24].

2.8 Penelitian Terdahulu

Berdasarkan hasil review jurnal terdahulu, maka dapat perbandingan sebagaimana ditampilkan pada Tabel 2.1 di bawah ini:

Tabel 2.1 Riview Jurnal Terdahulu

No	Judul, Penulis, Dan Tahun	Hasil Penelitian
1	Sistem Klasifikasi Penyakit Diabetes Mellitus Menggunakan Metode K-Nearest Neighbor (K-NN). Fitri Yunita (2016)	Hasil dari penelitian ini yang telah dilakukan, algoritma K-Nearest Neighbor (K-NN) mampu digunakan untuk mengklasifikasikan data dengan baik berdasarkan beberapa variabel seperti usia, berat badan, tinggi badan, riwayat keluarga, gangguan destruksi sel beta, dan pola makan. Hasil pengujian menunjukkan bahwa tingkat akurasi yang diperoleh cukup tinggi, yaitu mencapai 96%, sehingga metode ini tergolong efektif dalam proses klasifikasi. Selain itu, jumlah data latih (training data) sangat berpengaruh terhadap tingkat akurasi yang dihasilkan. Semakin banyak data latih yang digunakan, maka semakin tinggi pula tingkat akurasi klasifikasi.

Tabel 2.1 Riview Jurnal Terdahulu Lanjutan

No	Judul, Penulis, Dan Tahun	Hasil Penelitian
		Dengan demikian, dapat disimpulkan bahwa metode K-Nearest Neighbor (K-NN) merupakan metode yang cukup baik dan layak digunakan dalam menyelesaikan permasalahan klasifikasi data.
2	Penentuan Kelayakan Masyarakat Miskin Penerima Bantuan Menggunakan Metode Naive Bayes (Studi Kasus: Kabupaten Penajam Paser Utara) Nur Madia, Anindita Septiarini, Heliza Rahmania Hatta, Hamdani Hamdani, Masna Wati (2023)	Implementasi dan hasil pengujian pada penelitian penentuan kelayakan masyarakat miskin penerima bantuan dapat diambil beberapa kesimpulan yaitu metode Naïve Bayes dapat diterapkan pada sistem penentuan kelayakan masyarakat miskin penerima bantuan karena mendapatkan accuracy yang cukup baik, dan berdasarkan hasil pengujian menunjukkan bahwa metode Naïve Bayes dapat menghasilkan accuracy tertinggi pada scenario ketiga yaitu 60% atau 328 data latih dan 40% atau 218 data uji dengan nilai accuracy sebanyak 77,98%. Pada penelitian lebih lanjut dapat dikembangkan dengan menggunakan Pada penelitian lebih lanjut dapat dikembangkan dengan menggunakan metode lain seperti Artificial Neural Network, Support Vector Machine, atau K-Nearest Neighbor untuk meningkatkan nilai akurasi. Kelebihan dari penelitian ini adalah pendekatan metodologis yang sistematis, penggunaan data resmi dan relevan, evaluasi performan yang lengkap, jumlah variabel yang komprehensif. Kekurangan dari penelitian ini adalah nilai akurasi relatif rendah, tidak dilakukan perbandingan dengan metode lain. minim visualisasi dan analisis hasil lebih lanjut, tidak ada uji signifikansi atau validasi silang (cross-validation).
3	Implementasi Algoritma Naive Bayes untuk Sistem Pendukung	Hasil dari penelitian ini yaitu berhasil mengimplementasikan algoritma Naive Bayes sebagai sistem pendukung keputusan untuk menentukan kelayakan penerima bantuan pendidikan. Algoritma Naive Bayes mampu mengklasifikasikan siswa berdasarkan kriteria.

Tabel 2.1 Riwiew Jurnal Terdahulu Lanjutan

No	Judul, Penulis, Dan Tahun	Hasil Penelitian
	Penerima Bantuan Pendidikan Chandra Wisnu Nugroho, Sigrid Tamriesfatno, dan Sulkifli(2025).	Algoritma ini dapat membantu pihak terkait dalam membuat keputusan yang lebih objektif dan terukur. Kelebihan dari penelitian ini adalah Tujuan penelitian jelas dan kontekstual, Pemanfaatan algoritma Naive Bayes yang sesuai, Data dan variabel penelitian relevan, Hasil penelitian mudah diinterpretasikan. Kekurangan dari penelitian ini adalah Jumlah data terbatas (hanya 60 data latih dan 10 data uji), Tidak dilakukan pengujian akurasi kuantitatif, Tidak ada perbandingan metode lain, Validasi model kurang komprehensif.
4	Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako Amat Damuri, Umbar Riyanto, Hengki Rusdianto,Mohammad Aminudin (2021).	Berdasarkan penelitian yang dilakukan oleh Amat Damuri, dkk (2021), diterapkan algoritma Naive Bayes untuk melakukan klasifikasi kelayakan penerima bantuan sembako. Penelitian ini menggunakan total 175 data, yang terdiri dari 135 data training dan 40 data testing. Hasil evaluasi menggunakan confusion matrix didapatkan akurasi yang dihasilkan untuk 135 data training dengan 40 data testing dan tujuh atribut yang digunakan menghasilkan akurasi sebesar 86%, recall 85%, dan presisi 88%. Akurasi dapat dipengaruhi oleh beberapa faktor, diantaranya: jumlah data training, data testing dan atribut yang digunakan. Kelebihan dari penelitian ini adalah Metode penelitian jelas dan sistematis, Menggunakan data riil dan Hasil evaluasi model cukup tinggi. Kekurangan dari penelitian ini adalah Ukuran dataset terbatas, Tidak ada perbandingan algoritma, Minim analisis faktor Menggunakan data riil dan Hasil evaluasi model cukup tinggi. Kekurangan dari penelitian ini adalah Ukuran dataset terbatas, Tidak ada perbandingan algoritma, Minim analisis faktor dominan, Kurangnya visualisasi hasil pengujian, Validasi model sederhana.

Tabel 2.1 Review Jurnal Terdahulu Lanjutan

No	Judul, Penulis, Dan Tahun	Hasil Penelitian
5	Optimalisasi Pengalokasian Bantuan Sosial di Kota Banjarmasin Menggunakan Pendekatan Naive Bayes Muhammad Iqbal Adha, Finki Dona Marleny, Ayu Ahadi Ningrum (2025)	Hasil penelitian menunjukkan bahwa model Naive Bayes mencapai akurasi 95% pada data pengujian dan seluruh data, dengan distribusi prioritas sebagai berikut: Prioritas Tinggi (46,2%), Prioritas Sedang (29,3%), dan Prioritas Rendah (24,5%). Kecamatan Banjarmasin Selatan, Banjarmasin Barat, dan Banjarmasin Utara diidentifikasi sebagai wilayah dengan jumlah prioritas tinggi terbanyak. Implikasi praktis dari penelitian ini adalah rekomendasi alokasi bantuan yang lebih terarah, strategi distribusi cepat, dan program pendukung seperti pelatihan keterampilan untuk kelompok prioritas sedang. Kesimpulan penelitian menegaskan bahwa pendekatan ini efektif untuk alokasi bantuan yang efisien, namun evaluasi berkala diperlukan untuk memastikan relevansi model di masa depan. Kelebihan dari penelitian ini adalah Pendekatan Metodologis yang Jelas dan Sistematis, Akurasi Model Sangat Tinggi (95%), Proses Preprocessing Data yang Lengkap, Visualisasi Data Informatif, Relevansi Sosial yang Kuat. Kekurangan dari penelitian ini adalah Ketergantungan pada Data Statis Tidak Menggunakan Penyeimbangan Data (SMOTE), Variabel Terbatas.
6	Perbandingan Algoritma Naïve Bayes, C4.5, dan K-Nearest Neighbor untuk Klasifikasi Kelayakan Program Keluarga Harapan Putri Ramadani1, Riszki Fadillah, Quratih Adawiyah, Suerni, Baginda	Hasil penelitian menunjukkan bahwa algoritma C4.5 memberikan kinerja terbaik dengan akurasi 91,8%, presisi 90,7%, recall 92,3%, dan AUC 0,944. Naive Bayes mencatat akurasi 87,2% dan recall 88,9%, sedangkan K-NN menghasilkan akurasi 89,9% dan recall 91,1%, namun memerlukan komputasi lebih tinggi. Temuan ini menunjukkan bahwa C4.5 lebih efektif dalam mengklasifikasikan kelayakan penerima PKH secara akurat dan efisien. Penelitian ini menegaskan potensi algoritma machine learning dalam mendukung pengambilan keputusan pada

Tabel 2.1 Riwiew Jurnal Terdahulu Lanjutan

No	Judul, Penulis, Dan Tahun	Hasil Penelitian
	RestuAl Ghazali(2024)	program bantuan sosial. Studi lanjutan disarankan untuk memperluas cakupan data dan mengeksplorasi metode klasifikasi lainnya guna optimalisasi distribusi bantuan. Kelebihan dari penelitian ini adalah Menggunakan Data Lokal dan Aktual, Membandingkan Tiga Algoritma Populer, Evaluasi Model Lengkap, Visualisasi dan Implementasi Praktis, Metodologi Tersusun Baik. Kekurangan dari penelitian ini adalah Jumlah Data Terbatas (109 entri), Tidak Ada Seleksi Fitur Otomatis / Parameter Tuning, Konteks Wilayah Terbatas.
7	Implementasi Algoritma K-Nearest Neighbor Dalam Menentukan Penerima Bantuan Langsung Tunai. Anwar Pauji, Siti Aisyah, Agus Surip, Riko Saputra4, Irfan Ali(2020)	Penelitian yang dilakukan oleh Anwar Pauji dan rekan-rekan ini berfokus pada implementasi algoritma K-Nearest Neighbor (K-NN) dalam proses klasifikasi penerima Bantuan Langsung Tunai (BLT) di Desa Ciperna, Kabupaten Cirebon. Penelitian ini dilatarbelakangi oleh permasalahan yang sering muncul dalam penyaluran bantuan sosial, di mana banyak penerima yang tidak tepat sasaran. Hal tersebut terjadi akibat kurang optimalnya proses penetapan status keluarga miskin sebagai penerima manfaat. Oleh karena itu, peneliti berupaya mengembangkan model klasifikasi berbasis data mining dengan algoritma K-NN agar proses distribusi bantuan dapat lebih objektif, transparan, dan akurat. Kelebihan dari penelitian ini adalah Pendekatan sederhana dan mudah diimplementasikan, Dapat membantu pemerintah desa dalam pengambilan keputusan, Penggunaan atribut yang relevan. Kekurangan dari penelitian ini adalah Tingkat akurasi masih rendah (68,82%), Ketidakseimbangan data antar kelas, Keterbatasan variabel penelitian.

Tabel 2.1 Riview Jurnal Terdahulu Lanjutan

No	Judul, Penulis, Dan Tahun	Hasil Penelitian
8	Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Warga Penerima Bantuan Sosial Pajar Pahrudin, Kusno Harianto(2022)	Hasil dari penelitian ini adalah algoritma K-Nearest Neighbor (K-NN) terbukti efektif dalam mengklasifikasikan warga yang layak menerima bantuan sosial. Dengan memanfaatkan metode data mining, penelitian ini mengolah data warga berdasarkan atribut pendapatan, kondisi rumah, jumlah tanggungan, dan pekerjaan. Melalui perhitungan jarak menggunakan Euclidean Distance dan parameter $K = 5$, hasil menunjukkan bahwa data baru diklasifikasikan sebagai layak menerima bantuan sosial. Secara keseluruhan, metode K-NN dinilai sederhana namun akurat, serta berpotensi menjadi solusi untuk meningkatkan ketepatan penyaluran bantuan sosial di masyarakat. Kelebihan dari penelitian ini adalah Metode mudah diimplementasikan dan dipahami, Hasil klasifikasi bersifat objektif dan berbasis data, Fleksibel terhadap berbagai jenis atribut sosial ekonomi. Kekurangan dari penelitian ini adalah Belum dilakukan pengujian terhadap akurasi model secara kuantitatif, Ukuran dataset sangat terbatas, Sensitif terhadap nilai parameter K, Belum mempertimbangkan normalisasi dan pembobotan atribut, Tidak ada perbandingan dengan algoritma lain.
9	Klasifikasi Program Bantuan Sosial Menggunakan Algoritma K-Nearest Neighbor (K-NN) (Studi Kasus Kecamatan Malangbong Kabupaten Garut) Hamzah Nurriq Fakhri Fikrillah(2025)	Hasil dari penelitian yang dilakukan oleh Fikrillah (2025), penerapan algoritma K-Nearest Neighbor (K-NN) terbukti efektif dalam mengklasifikasikan keluarga penerima Program Bantuan Sosial (BANSOS) di Kecamatan Malangbong, Kabupaten Garut. Dengan menggunakan pendekatan CRISP-DM dan perangkat lunak RapidMiner, penelitian ini berhasil mengolah 19.943 data keluarga yang telah Melalui perhitungan jarak menggunakan Euclidean Distance dan parameter $K = 5$, hasil menunjukkan bahwa data baru diklasifikasikan sebagai layak menerima bantuan sosial. Secara

Tabel 2.1 Review Jurnal Terdahulu Lanjutan

No	Judul, Penulis, Dan Tahun	Hasil Penelitian
		keseluruhan, metode K-NN dinilai sederhana namun akurat, serta berpotensi menjadi solusi untuk meningkatkan ketepatan penyaluran bantuan sosial di masyarakat. Kelebihan dari penelitian ini adalah Metode mudah diimplementasikan dan dipahami, Hasil klasifikasi bersifat objektif dan berbasis data, Fleksibel terhadap berbagai jenis atribut sosial ekonomi. Kekurangan dari Nilai recall sangat rendah, Tidak dijelaskan metode normalisasi atau standarisasi, Tidak ada perbandingan dengan algoritma lain, Tidak membahas kompleksitas komputasi, Ketidakseimbangan kelas tidak dijelaskan.
10	Sistem Pendukung Keputusan Penerima Bantuan Langsung Tunai Dana Desa Menggunakan Metode Algoritma K-Means Clustering Sarmila Sari, Joy Nashar Utamajaya (2022),	Berdasarkan hasil penelitian yang telah dilakukan secara manual dengan metode K-means clustering dan evaluasi hasil keseluruhan data menggunakan Software Rapidminer versi 5 dihasilkan Nama-Nama masyarakat yang layak menerima bantuan dengan hasil yang sama yaitu terdapat 3 cluster yang dihasilkan dalam mengelompokkan data penduduk kurang mampu penerima bantuan langsung tunai dana desa, dengan cluster 1 adalah kelompok dengan status penerima bantuan sosial tepat sasaran sebanyak 87 data Penduduk, cluster 2 dengan status penerima bantuan sosial dipertimbangkan sebanyak 193 Penduduk, dan cluster 3 dengan status penerima bantuan sosial tidak tepat sasaran sebanyak 56 Penduduk, dimana Pengolahan data yang dilakukan menghasilkan kelompok penerima bantuan yang telah dipertimbangkan menggunakan 5 variabel penelitian yaitu Nama, Pekerjaan, Tanggungan, Kondisi Rumah, dan penghasilan sehingga pihak desa bisa lebih mudah dalam menentukan penerima bantuan Langsung tunai dana desa dengan teknik Data Mining. Dari hasil analisis

Tabel 2.1 Riview Jurnal Terdahulu Lanjutan

No	Judul, Penulis, Dan Tahun	Hasil Penelitian
10		data dapat ditarik kesimpulan yaitu masyarakat yang menerima bantuan berada pada cluster 0 (cluster 1) yaitu masyarakat yang benar- benar membutuhkan bantuan dari pemerintah, dimana penerima bantuan bekerja sebagai buruh,Kondisi rumah terbuat dari papan, dan memiliki penghasilan di bawah Rp 500.000. Kelebihan dari penelitian ini adalah Pendekatan berbasis data yang objektif, Mempermudah pengambilan keputusan pemerintah desa, Pemanfaatan variabel sosial ekonomi yang relevan, Hasil konsisten antara perhitungan manual dan sistem (RapidMiner), Meningkatkan efisiensi proses penyaluran bantuan. Kekurangan dari penelitian ini adalah Jumlah variabel terbatas, Tidak dilakukan evaluasi model dengan metrik kuantitatif, Sensitif terhadap skala data dan outlier, Sensitif terhadap skala data dan outlier, Tidak ada perbandingan dengan algoritma lain.

Tabel 2.1 Berdasarkan tinjauan terhadap sepuluh jurnal terdahulu yang telah dianalisis, dapat disimpulkan bahwa penerapan algoritma machine learning dalam sistem pendukung keputusan penyaluran bantuan sosial telah menunjukkan efektivitas yang signifikan. Algoritma Naive Bayes menjadi metode yang paling banyak diteliti dengan tingkat akurasi yang bervariasi antara 77,98% hingga 95%, dimana penelitian Adha et al. (2025) mencapai akurasi tertinggi sebesar 95% dalam optimalisasi pengalokasian bantuan sosial di Kota Banjarmasin. Sementara itu, algoritma K-Nearest Neighbor (K-NN) juga menunjukkan performa yang memadai dengan rentang akurasi 68,82% hingga 89,9%, meskipun cenderung memerlukan komputasi yang lebih tinggi dan sensitif terhadap pemilihan parameter K.

Studi komparatif yang dilakukan oleh Ramadani et al. (2024) mengungkapkan bahwa algoritma C4.5 memberikan kinerja superior dengan akurasi 91,8%, presisi 90,7%, recall 92,3%, dan nilai AUC 0,944 dalam mengklasifikasikan kelayakan penerima Program Keluarga Harapan, mengungguli Naive Bayes (87,2%) dan K-NN (89,9%). Temuan ini mengindikasikan bahwa pemilihan algoritma yang tepat sangat bergantung pada karakteristik dataset dan konteks permasalahan yang dihadapi. Selain itu, pendekatan clustering menggunakan metode K-Means yang diterapkan oleh Sari dan Utamajaya (2022) mampu mengelompokkan penerima bantuan menjadi tiga kategori prioritas yaitu tepat sasaran, dipertimbangkan, dan tidak tepat sasaran, yang memfasilitasi pengambilan keputusan yang lebih terstruktur.

Meskipun demikian, terdapat beberapa keterbatasan metodologis yang konsisten ditemukan dalam sebagian besar penelitian, antara lain: pertama, ukuran dataset yang relatif terbatas dengan rentang 60 hingga 19.943 data yang berpotensi mengurangi generalisasi model; kedua, minimnya validasi model secara komprehensif seperti cross-validation atau uji signifikansi statistik; ketiga, kurangnya perbandingan sistematis antar algoritma untuk menentukan metode optimal; keempat, tidak adanya penanganan ketidakseimbangan kelas (class imbalance) yang dapat mempengaruhi performa klasifikasi; dan kelima, keterbatasan variabel penelitian yang belum mencakup faktor ekonomi dan sosial yang lebih kompleks. Oleh karena itu, penelitian lanjutan disarankan untuk mengintegrasikan multiple algorithms comparison dengan teknik ensemble learning, menerapkan *feature selection* dan *hyperparameter tuning* untuk

optimalisasi model, serta melakukan evaluasi berkala terhadap relevansi model dalam konteks dinamika sosial ekonomi masyarakat yang terus berkembang.

2.9 Variabel Penelitian

Variabel penelitian adalah sesuatu yang menjadi fokus pengamatan dalam suatu penelitian dan dapat diukur, dihitung, atau dikategorikan. Variabel digunakan untuk menjelaskan hubungan, pengaruh, atau perbedaan antara satu konsep dengan konsep lainnya. Dalam penelitian kuantitatif, variabel biasanya dibagi menjadi variabel yang memengaruhi (variabel independen) dan variabel yang dipengaruhi (variabel dependen). Dengan kata lain, variabel penelitian adalah unsur yang dijadikan acuan untuk mengumpulkan data sehingga peneliti dapat menjawab rumusan masalah.

2.9.1 Variabel Independen

Variabel independen (variabel bebas) merupakan variabel yang memengaruhi atau menjadi faktor penyebab perubahan pada variabel dependen, yaitu kelayakan penerima Bantuan Langsung Tunai (BLT). Dalam penelitian berjudul “Implementasi Algoritma Naive Bayes untuk Menentukan Kelayakan Penerima Bantuan Langsung Tunai di Desa Sungai Ambat”, variabel independen adalah faktor-faktor sosial ekonomi rumah tangga yang diduga berkontribusi terhadap penentuan status kelayakan penerima BLT. Variabel-variabel tersebut berkaitan langsung dengan kondisi kesejahteraan warga, sehingga menjadi masukan penting bagi model klasifikasi berbasis algoritma Naive Bayes. Adapun variabel independen yang digunakan dalam penelitian ini meliputi:

1. Usia Kepala Keluarga, yaitu umur kepala keluarga dalam satuan tahun;

2. Jumlah Anggota Rumah Tangga, yakni total anggota yang tinggal satu rumah dan menjadi tanggungan ekonomi;
3. Jenis kelamin kepala keluarga, yaitu Jenis kelamin kepala keluarga yang tercatat sebagai penanggung jawab rumah tangga;
4. Pendidikan Kepala Keluarga berdasarkan tingkat pendidikan terakhir yang diselesaikan;
5. Pendapatan Bulanan, yaitu total penghasilan keluarga dalam satu bulan;
6. Status Pekerjaan Kepala Keluarga sebagai sumber nafkah utama;
7. Status Penerima Bantuan Sosial Lain, yaitu apakah keluarga menerima bantuan sosial lain dari pemerintah;
8. Kepemilikan Aset Dasar, seperti rumah, kendaraan, atau tanah;
9. Kelayakan Kondisi Rumah, yaitu kualitas fisik bangunan rumah;
10. Akses Air Bersih berdasarkan sumber air yang digunakan;
11. Akses Listrik sebagai ketersediaan fasilitas listrik di rumah tangga;
12. Keberadaan Lansia, yaitu apakah terdapat anggota keluarga berusia di atas 60 tahun; dan
13. Keberadaan Penyandang Disabilitas di dalam rumah tangga.

Hubungan variabel independen dengan penelitian ini sangat penting karena seluruh variabel tersebut menjadi input (fitur) bagi algoritma Naive Bayes untuk melakukan proses analisis dan klasifikasi. Melalui data yang dikumpulkan dari variabel-variabel sosial ekonomi tersebut, algoritma Naive Bayes akan mencari pola dan hubungan probabilistik yang dapat digunakan untuk memprediksi nilai variabel dependen, yaitu status kelayakan penerima BLT. Dengan demikian, kedua belas variabel independen tersebut berfungsi sebagai dasar bagi model untuk

menghasilkan prediksi yang objektif dan terukur, sekaligus mencerminkan kondisi sosial ekonomi masyarakat Desa Sungai Ambat sebagai dasar pengambilan keputusan oleh pemerintah desa.

2.9.2 Variabel Dependen

Dalam penelitian ini, variabel dependen yang digunakan adalah kelayakan penerima Bantuan Langsung Tunai (BLT) di Desa Sungai Ambat. Variabel ini menjadi fokus utama yang ingin diprediksi menggunakan algoritma Naive Bayes. Dengan kata lain, variabel dependen merupakan hasil atau keluaran dari proses analisis yang menunjukkan apakah seorang warga layak atau tidak layak menerima BLT berdasarkan kriteria yang berlaku dalam pendataan sosial ekonomi masyarakat desa.

Hubungan variabel dependen dengan penelitian ini sangat erat, karena tujuan utama dari penelitian adalah untuk memprediksi status kelayakan tersebut berdasarkan berbagai faktor yang memengaruhinya. Faktor-faktor tersebut, yang disebut sebagai variabel independen, meliputi antara lain jumlah pendapatan keluarga per bulan, jumlah tanggungan, jenis pekerjaan, kondisi fisik rumah, tingkat pendidikan kepala keluarga, akses terhadap fasilitas dasar, serta kepemilikan aset tertentu.

Melalui penerapan algoritma Naive Bayes, penelitian ini berupaya mengidentifikasi pola hubungan antara variabel independen dengan variabel dependen sehingga sistem mampu menghasilkan prediksi yang akurat mengenai status kelayakan penerima BLT. Dengan demikian, variabel dependen berperan penting sebagai tujuan akhir dari proses analisis dan prediksi dalam penelitian ini, sekaligus menjadi dasar bagi pemerintah desa dalam menentukan rekomendasi

pemberian BLT secara objektif dan tepat sasaran. Variabel kelayakan ini dikategorikan ke dalam dua kelas utama, yaitu:

1. Layak: warga memenuhi kriteria sosial ekonomi yang ditetapkan dan dianggap perlu menerima bantuan.
2. Tidak Layak: warga tidak memenuhi kriteria atau berada pada kondisi ekonomi yang lebih baik dibanding penerima prioritas.

Data mengenai kelayakan diperoleh melalui pendataan sosial ekonomi desa yang mencakup indikator kesejahteraan rumah tangga serta hasil klasifikasi model machine learning menggunakan algoritma Naive Bayes.

2.10 Rangkuman

Berdasarkan hasil kajian terhadap sepuluh penelitian terdahulu, dapat disimpulkan bahwa penerapan teknik machine learning dan data mining memiliki peran yang signifikan dalam mendukung proses pengambilan keputusan, khususnya pada penentuan kelayakan penerima bantuan sosial. Berbagai algoritma klasifikasi, seperti Naive Bayes, K-Nearest Neighbor (K-NN), C4.5, serta metode clustering K-Means, telah banyak diterapkan untuk mengolah data sosial ekonomi masyarakat sehingga menghasilkan keputusan yang lebih objektif, cepat, dan terukur. Secara umum, penelitian-penelitian tersebut menunjukkan bahwa pemanfaatan algoritma klasifikasi mampu meningkatkan ketepatan sasaran penyaluran bantuan dibandingkan proses seleksi yang dilakukan secara manual.

Hasil kajian menunjukkan bahwa algoritma Naive Bayes merupakan metode yang paling banyak digunakan dalam penelitian terkait klasifikasi penerima bantuan sosial. Hal ini disebabkan karena Naive Bayes memiliki proses komputasi yang sederhana, efisien, serta mampu menghasilkan tingkat akurasi yang cukup

tinggi pada data kategorikal. Berdasarkan penelitian yang dianalisis, tingkat akurasi Naive Bayes berada pada rentang 77,98% hingga 95%, bergantung pada karakteristik data, jumlah data latih, serta tahapan preprocessing yang diterapkan. Di sisi lain, algoritma K-Nearest Neighbor (K-NN) juga menunjukkan performa yang baik dengan tingkat akurasi yang kompetitif, namun kinerjanya sangat dipengaruhi oleh pemilihan nilai parameter K, metode pengukuran jarak, serta distribusi data yang digunakan. Selain kedua algoritma tersebut, beberapa penelitian juga membandingkan Naive Bayes dengan algoritma lain, seperti C4.5, yang pada beberapa kasus menghasilkan performa lebih tinggi. Temuan tersebut menunjukkan bahwa tidak terdapat satu algoritma yang selalu memberikan hasil terbaik pada seluruh dataset, melainkan performa model sangat dipengaruhi oleh karakteristik data, jumlah atribut, kualitas data, serta teknik pengolahan data yang diterapkan sebelum proses klasifikasi dilakukan. Oleh karena itu, proses evaluasi dan perbandingan beberapa algoritma menjadi langkah penting dalam menentukan model yang paling sesuai untuk suatu permasalahan.

Meskipun hasilnya menjanjikan, banyak penelitian memiliki keterbatasan, termasuk ukuran dataset yang kecil, minimnya validasi model, kurangnya perbandingan antar algoritma, tidak adanya penanganan ketidakseimbangan data, dan variabel penelitian yang belum komprehensif. Oleh sebab itu, penelitian selanjutnya perlu mengoptimalkan model dengan perbandingan multi-algoritma, teknik ensemble, feature selection, hyperparameter tuning, serta evaluasi berkala agar hasil lebih akurat dan relevan terhadap dinamika sosial ekonomi masyarakat.

BAB III

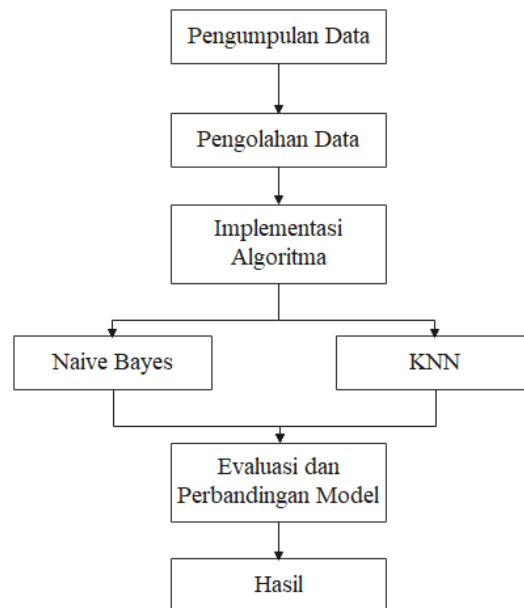
METODE PENELITIAN

Dalam bab metode penelitian ini dijelaskan langkah-langkah yang digunakan dalam proses menentukan kelayakan penerima Bantuan Langsung Tunai (BLT) di Desa Sungai Ambat. Metode penelitian disusun mulai dari tahap identifikasi masalah, pengumpulan data, pengolahan data, hingga proses klasifikasi dan evaluasi hasil. Penelitian ini menggunakan pendekatan kuantitatif, karena seluruh analisis dilakukan berdasarkan data sosial ekonomi masyarakat yang diolah secara sistematis dan terukur. Model yang digunakan dalam penelitian ini adalah algoritma Naive Bayes, yang berfungsi sebagai alat untuk melakukan menentukan kelayakan penerima BLT berdasarkan variabel independen yang telah ditentukan. Melalui penerapan algoritma tersebut, penelitian bertujuan menghasilkan kelayakan penerima bantuan yang objektif, akurat, dan dapat membantu pemerintah desa dalam menyalurkan bantuan secara tepat sasaran.

3.1 Kerangka Penelitian

Kerangka penelitian menggambarkan alur berpikir dan tahap kegiatan yang dilakukan penelitian dalam mencapai tujuan penelitian. Setiap langkah disusun secara sistematis agar proses pengumpulan, pengolahan, dan analisis data dapat berjalan dengan terarah dan menghasilkan kelayakan penerima Bantuan Langsung Tunai (BLT) yang valid. Proses dan langkah penelitian ini meliputi pengumpulan data warga Desa Sungai Ambat, pengolahan data pra-pemrosesan (preprocessing), implementasi algoritma Naive Bayes Sebagai Model utama, serta model *K-Nearest Neighbor* (KNN) sebagai pembanding. Evaluasi performa dilakukan menggunakan

metrik seperti *accuracy*, *precision*, *recall*, *F1-score* dan *confusional matrix* untuk menilai kualitas model. Selanjutnya, hasil penentuan dianalisis untuk menentukan efektifitas masing-masing algoritma serta mengidentifikasi variabel-variabel utama yang berpengaruh terhadap keputusan kelayakan penerima BLT di Desa Sungai Ambat.



Gambar 3.1 Alur Kerangka Penelitian

3.1.1 Pengumpulan data

Pengumpulan data yaitu mengumpulkan data warga Desa Sungai Ambat yang menjadi calon penerima BLT. Untuk memperoleh data yang akurat dan sesuai dengan kebutuhan penelitian, beberapa teknik pengumpulan data digunakan, yaitu:

1. Observasi

Teknik observasi digunakan untuk mengamati langsung kondisi masyarakat serta mengetahui proses administrasi penyaluran BLT di kantor

desa. Observasi membantu peneliti memahami situasi dan kondisi nyata yang relevan dengan penelitian.

2. Wawancara

Wawancara dilakukan secara terstruktur dengan perangkat desa guna mendapatkan penjelasan mengenai prosedur penentuan penerima BLT, kriteria-kriteria yang digunakan, dan permasalahan yang sering terjadi dalam proses seleksi. Wawancara ini memberikan informasi tambahan sebagai pendukung analisis data sekunder.

3. Dokumentasi

Dokumentasi dilakukan dengan mengumpulkan data yang sudah tersedia dalam arsip desa. Teknik ini digunakan untuk memperoleh dataset yang berisi atribut-atribut sosial ekonomi warga sebagai bahan utama dalam proses pemodelan Naive Bayes. Data yang diperoleh melalui dokumentasi merupakan data aktual dan telah digunakan dalam proses administrasi desa.

Data yang diperoleh terdiri dari tiga belas variabel input dan target sebanyak satu kelas. Adapun variabel input dalam penelitian ini dapat dilihat pada tabel 3.2 dibawah ini :

Tabel 3.2 Variabel Masukan

No	Variabel Independen	Definisi Operasional	Indikator	Jenis Data	Kategori / Skala
1	Usia Kepala Keluarga	Umur kepala keluarga yang dihitung dalam satuan tahun.	Usia berdasarkan tanggal lahir	Numerik	Tahun
2	Jumlah Anggota Rumah Tangga	Total anggota yang tinggal dalam satu rumah dan menjadi	Jumlah anggota keluarga	Numerik	Orang

Tabel 3.2 Variabel Masukan Lanjutan

No	Variabel Independen	Definisi Operasional	Indikator	Jenis Data	Kategori / Skala
		tanggungan ekonomi.			
3	Jenis kelamin kepala keluarga	Jenis kelamin kepala keluarga yang tercatat sebagai penanggung jawab rumah tangga	Laki-laki/ Perempuan	kategorikal	Biner
4	Pendidikan Kepala Keluarga	Tingkat pendidikan terakhir yang diselesaikan oleh kepala keluarga.	SD, SMP, SMA, >SMA	Kategorikal	Ordinal
5	Pendapatan Bulanan	Total penghasilan keluarga dalam satu bulan.	Upah, hasil usaha, hasil pertanian	Numerik / Kategori	Rupiah / Rendah–Sedang–Tinggi
6	Status Pekerjaan Kepala Keluarga	Jenis pekerjaan utama yang menjadi sumber nafkah keluarga.	Buruh, Petani, Wiraswasta, Tidak bekerja	Kategorikal	Nominal
7	Penerima Bantuan Sosial Lain	Status apakah rumah tangga menerima program bantuan lain.	Ya / Tidak	Kategorikal	Biner
8	Kepemilikan Aset Dasar	Kondisi kepemilikan aset seperti rumah, tanah, atau kendaraan.	Memiliki rumah, kendaraan, tanah	Kategorikal	Biner / Nominal

Tabel 3.2 Variabel Masuk

No	Variabel Independen	Definisi Operasional	Indikator	Jenis Data	Kategori / Skala
9	Kelayakan Kondisi Rumah	Kualitas bangunan rumah berdasarkan kondisi fisik.	Layak, Kurang Layak, Tidak Layak	Kategorikal	Ordinal
10	Akses Air Bersih	Jenis sumber air yang digunakan rumah tangga.	PDAM, Sumur, Sungai	Kategorikal	Nominal
11	Akses Listrik	Ketersediaan listrik di rumah tangga.	Ada / Tidak ada	Kategorikal	Biner
12	Keberadaan Lansia	Adanya anggota rumah tangga berusia >60 tahun.	Ya / Tidak	Kategorikal	Biner
13	Keberadaan Penyandang Disabilitas	Adanya anggota rumah tangga yang memiliki keterbatasan fisik atau mental.	Ya / Tidak	Kategorikal	Biner

Sedangkan sebagai unit output atau target yang akan dihasilkan dapat dilihat pada tabel 3.3 dibawah ini :

Tabel 3.3 Variabel Output

No	Variabel	Keterangan
1	0	Tidak Layak
2	1	Layak

3.1.2 Pengolahan data

Pengolahan data yaitu meliputi pembersihan data, penanganan nilai hilang, transformasi atribut, dan normalisasi data agar siap digunakan dalam

model. Pada tahap ini juga dilakukan pembagian dataset menjadi data latih dan data uji. Pengolahan data dilakukan agar data siap diproses ke dalam algoritma Naive Bayes dan KNN. Tahapan pengolahan data meliputi:

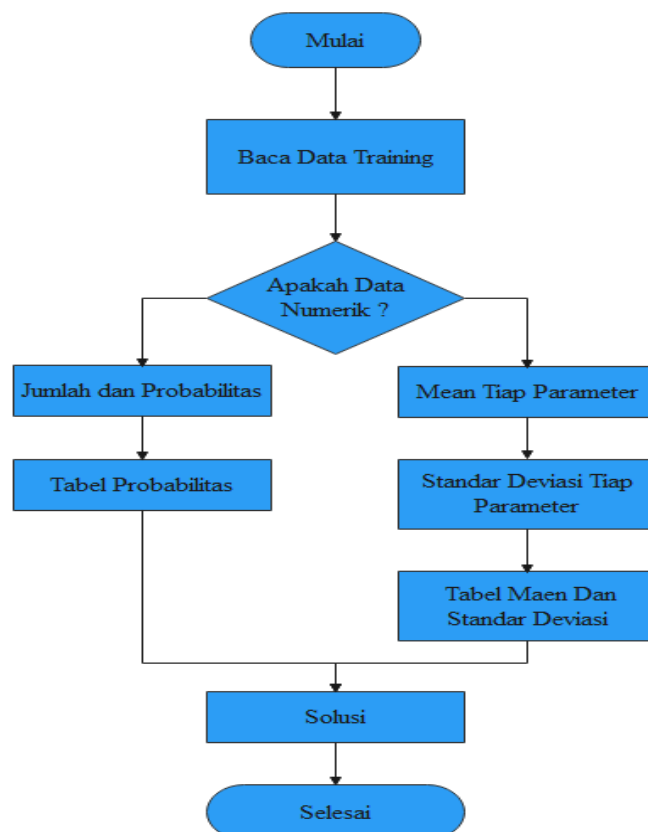
1. **Pembersihan Data (Data Cleaning)** Proses pembersihan data dilakukan untuk memastikan data yang digunakan benar-benar siap diproses oleh algoritma. menghapus data duplikat, menangani nilai hilang, dan memperbaiki kesalahan input
2. **Transformasi Atribut (Data Transformation)** Data yang telah dibersihkan selanjutnya ditransformasikan agar sesuai dengan kebutuhan model. Tahap tersebut yaitu : Diskritisasi atribut numerik, Pengkodean data kategorikal (label encoding atau one-hot encoding), Normalisasi data, Transformasi format, seperti mengubah tipe data (string → numerik).
3. **Pemilihan Fitur (Feature Selection)** Tahap ini bertujuan memilih atribut-atribut yang paling relevan dalam menentukan kelayakan penerima BLT. Pemilihan fitur dilakukan agar model lebih akurat dan tidak dipengaruhi oleh atribut yang tidak penting.
4. **Pembagian Dataset**, Data yang telah melewati tahap pembersihan dan transformasi selanjutnya dibagi menjadi dua yaitu Data Latih (Training Set) – 80% Digunakan untuk melatih model Naive Bayes dan KNN dengan mempelajari pola dari atribut-atribut yang ada. Data Uji (Testing Set) – 20% Digunakan untuk menguji kinerja model secara objektif dan mengukur kemampuan model dalam melakukan prediksi pada data baru.

3.1.3 Implementasi Algoritma

Tahap implementasi algoritma merupakan proses memasukkan data yang telah melalui tahap pengolahan ke dalam model klasifikasi. Pada penelitian ini digunakan dua algoritma, yaitu Naive Bayes sebagai model utama dan K-Nearest Neighbor (KNN) sebagai model pembanding. Implementasi dilakukan menggunakan perangkat lunak Google Colab dengan langkah-langkah sebagai berikut:

1. Implementasi Algoritma Naive Bayes

Naive Bayes adalah algoritma klasifikasi berbasis probabilitas yang bekerja dengan menerapkan *Teorema Bayes* dan asumsi independensi antar atribut.



Gambar 3.2 Alur Kerja Naive Bayes

Teorema Bayes dirumuskan sebagai berikut :

$$P(H|X) = \frac{P(X | H) \times P(H)}{P(X)} \dots \dots \dots [1]$$

Adapun tahapan implementasinya dalam penelitian ini yaitu:

1. Memasukkan data latih ke model
2. Menghitung nilai prior dan likelihood untuk setiap kelas (Layak / Tidak Layak)
3. Menerapkan rumus probabilitas posterior Naive Bayes
4. Menguji model menggunakan data uji
5. Menghasilkan prediksi kelayakan penerima BLT

2. Implementasi Algoritma K-Nearest Neighbor(KNN)

KNN digunakan sebagai algoritma pembandingan untuk melihat perbedaan performa dengan model utama. KNN bekerja dengan mencari sejumlah tetangga terdekat berdasarkan jarak antar data. Adapun tahapan implementasinya dalam penelitian ini yaitu:

1. Melakukan normalisasi data karena KNN sensitif terhadap skala atribut
2. Menentukan nilai **K** terbaik (misalnya $K = 3, 5$, atau 7)
3. Menghitung jarak Euclidean antara data uji dan data latih

$$d(x, y) = \sqrt{\left(\sum (x_i - y_i)^2\right)} \dots \dots \dots [2]$$

4. Mengambil mayoritas kelas tetangga terdekat
5. Menghasilkan prediksi kelayakan penerima BLT

3.1. 4 Evaluasi Kinerja Model

Evaluasi dilakukan untuk mengetahui seberapa baik setiap algoritma melakukan klasifikasi. Ini dilakukan untuk melihat seberapa baik model yang digunakan dalam memberikan hasil. Proses evaluasi model ini adalah confusion matriks, akurasi, presisi, recall dan F1-score. Metrik yang digunakan adalah:

1. Accuracy (akurasi): mengukur persentase prediksi yang benar.

Rumus:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \dots \dots \dots [3]$$

2. Precision : Mengukur tingkat ketepatan ketika model memprediksi "layak"

Rumus :

$$Precision = \frac{TP}{TP + FP} \dots \dots \dots [4]$$

3. Recall : Kemampuan model mendeteksi warga yang benar-benar "layak"

Rumus:

$$Recall = \frac{TP}{TP + FN} \dots \dots \dots [5]$$

4. F1-Score : Gabungan harmonis antara precision dan recall

Rumus :

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \dots \dots \dots [6]$$

5. Confusion Matrix

Confusion matrix menggambarkan hubungan antara kelas aktual dan prediksi model.

Tabel 3.4 Confusion Matrix

Aktual / Prediksi	Layak	Tidak Layak
Layak	TP	FN
Tidak Layak	FP	TN

3.1.5 Paired t-test untuk Membandingkan Performa Algoritma

Paired t-test (uji t berpasangan) adalah uji statistik parametrik yang digunakan untuk menentukan apakah terdapat perbedaan yang signifikan antara dua kelompok nilai yang berpasangan atau menggunakan sampel yang sama. *Paired t-test* dilakukan melalui serangkaian tahapan untuk menentukan apakah terdapat perbedaan yang signifikan antara dua kelompok nilai berpasangan. Pada penelitian ini, nilai yang dibandingkan adalah akurasi algoritma Naive Bayes dan KNN pada setiap fold k-fold cross validation. Langkah-Langkah Paired t-test :

1. Mengumpulkan Pasangan Data (X_1 dan X_2)

Langkah pertama adalah mengumpulkan dua kelompok data yang saling berpasangan. Pada konteks penelitian, data tersebut adalah:

X_1 = akurasi Naive Bayes

X_2 = akurasi KNN

2. Menghitung Selisih Antar Pasangan (d_i)

Untuk setiap pasangan data, hitung selisihnya:

$$d_i = x_{(1,i)} - x_{(2,i)} \dots \dots \dots [7]$$

Hasilnya membentuk daftar selisih:

$$d_1, d_2, d_3, \dots, d_n$$

3. Menghitung Rata-rata Selisih ($\bar{d}$)

Rata-rata selisih dihitung menggunakan rumus:

$$\bar{d} = \frac{\sum_{i=1}^n d_i}{n} \dots \dots \dots [8]$$

Rata-rata ini menunjukkan kecenderungan selisih, apakah Naive Bayes cenderung lebih tinggi, lebih rendah, atau setara dengan KNN.

4. Menghitung Deviasi (Varians) Selisih (s_e^2)

Hitung varians dari selisih tersebut:

$$s_e^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-k} \dots \dots \dots [9]$$

Varians ini menggambarkan tingkat keragaman selisih antarpasangan.

5. Menghitung Standar Deviasi Selisih (s_e)

Standar deviasi diperoleh dari akar varians:

$$s_d = \sqrt{(s_d^2)} \dots \dots \dots [10]$$

Standar deviasi menunjukkan seberapa besar perbedaan data terhadap rata-rata selisih.

6. Menghitung Standard Error (SE)

Kesalahan standar dihitung dengan:

$$SE = \frac{sd}{\sqrt{n}} \dots \dots \dots [11]$$

Nilai ini digunakan untuk menormalkan rata-rata selisih dalam perhitungan uji t.

7. Menghitung Nilai t-hitung (t)

Nilai t-hitung diperoleh dengan rumus:

$$t = \frac{\bar{d}}{SE} = \frac{\bar{d}}{(\frac{s_d}{\sqrt{n}})} \dots \dots \dots [12]$$

Semakin besar nilai t-hitung, semakin besar kemungkinan terdapat perbedaan signifikan antara kedua algoritma.

8. Menentukan Derajat Kebebasan (df)

Derajat kebebasan untuk paired t-test adalah:

$$df = n - 1 \dots \dots \dots [13]$$

9. Menentukan Nilai p-value

Nilai p-value diperoleh dari tabel distribusi t atau perangkat lunak statistik

Google Colab berdasarkan perhitungan nilai t-hitung dan derajat kebebasan.

10. Menarik kesimpulan

Bandingkan p-value dengan tingkat signifikansi ($\alpha = 0.05$):

Jika p-value < 0.05 : Terdapat perbedaan signifikan antara performa Naive

Bayes dan KNN, Salah satu algoritma terbukti lebih unggul secara statistik.

Jika p-value ≥ 0.05 : Tidak terdapat perbedaan signifikan, Kedua algoritma

memiliki performa yang setara secara statistik.

3.2 Alat dan perangkat penelitian

Alat dan perangkat penelitian merupakan komponen penting yang mendukung proses pengolahan data, analisis, dan penyusunan laporan. Pemilihan

perangkat yang tepat bertujuan untuk memastikan bahwa seluruh tahapan penelitian

dapat dilakukan secara efektif, efisien, dan akurat. Pada penelitian ini, alat dan

perangkat yang digunakan meliputi perangkat keras (hardware) serta perangkat

lunak (software) yang dijelaskan sebagai berikut:

1. Perangkat Keras (Hardware)

Perangkat keras digunakan untuk menjalankan aplikasi, mengolah data, dan melakukan pemodelan algoritma data mining. Adapun spesifikasi perangkat keras

yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Laptop/Komputer dengan spesifikasi minimum:

2. Prosesor: Intel Core i7 atau setara
3. RAM: Minimal 8 GB
4. Penyimpanan: Minimal 256 GB SSD
5. Sistem Operasi: Windows 11

Spesifikasi tersebut dipilih untuk memastikan proses pengolahan data dan pemodelan algoritma dapat berjalan tanpa hambatan, terutama saat memproses dataset berukuran sedang hingga besar.

2. Perangkat Lunak (Software)

Perangkat lunak digunakan untuk memfasilitasi pengolahan data, analisis, visualisasi, dan penyusunan dokumen penelitian. Perangkat lunak yang digunakan dalam penelitian ini meliputi:

1. Microsoft Excel, digunakan untuk pengelolaan data awal, pembersihan data sederhana, serta visualisasi dasar.
2. Google Colab, digunakan untuk proses pemodelan data mining seperti algoritma Naive Bayes, pengujian model, serta evaluasi hasil.
3. Microsoft Word, digunakan dalam proses penyusunan laporan penelitian, dokumentasi hasil, serta pembuatan grafik dan interpretasi data.

3. Data Penelitian

Selain perangkat keras dan lunak, penelitian ini juga memanfaatkan data sebagai objek utama analisis. Data yang digunakan adalah data sosial ekonomi warga Desa Sungai Ambat yang terdiri dari 13 variabel input dan 1 variabel output yang kemudian diolah dan dianalisis menggunakan metode data mining. Data ini merupakan komponen utama dalam proses pemodelan dan menjadi dasar dalam perumusan hasil peneliti

BAB IV

HASIL DAN PEMBAHASAN

Pada bab ini dijelaskan hasil implementasi algoritma Naive Bayes dan K-Nearest Neighbor (KNN) dalam proses klasifikasi kelayakan penerima bantuan langsung tunai (BLT) di Desa Sungai Ambat. Penelitian dilakukan menggunakan data sosial ekonomi masyarakat yang telah melalui tahap preprocessing sehingga siap digunakan dalam proses klasifikasi. Penelitian ini menggunakan perangkat lunak Google Colab untuk melakukan pengolahan data, pelatihan model, pengujian model, serta evaluasi performa algoritma. Variabel yang digunakan terdiri dari 13 variabel independen dan 1 variabel dependen berupa status kelayakan penerima BLT. Penelitian ini dilakukan di Desa Sungai Ambat dengan tujuan untuk mengklasifikasikan kelayakan masyarakat penerima Bantuan Langsung Tunai (BLT) menggunakan pendekatan data mining berbasis machine learning. Dalam penelitian ini digunakan dua algoritma klasifikasi, yaitu Naive Bayes sebagai metode Utama dan K-Nearest Neighbor (KNN) sebagai metode perbandingan. Pemilihan kedua algoritma tersebut didasarkan pada kemampuan masing-masing algoritma dalam melakukan proses klasifikasi data berdasarkan pola dan karakteristik data yang dipelajari sebelumnya. Algoritma Naive Bayes dipilih karena memiliki kemampuan dalam menghitung probabilitas suatu kelas berdasarkan atribut yang dimiliki data, sehingga sangat efektif digunakan untuk proses pengambilan keputusan yang cepat dan sederhana. Sementara itu, algoritma K-Nearest Neighbor (KNN) digunakan karena mampu melakukan klasifikasi

berdasarkan kedekatan jarak antar data, sehingga dapat memberikan hasil klasifikasi berdasarkan kemiripan karakteristik data masyarakat.

Data yang digunakan dalam penelitian ini merupakan data masyarakat Desa Sungai Ambat yang diperoleh langsung dari kantor desa melalui proses wawancara dan dokumentasi. Proses wawancara dilakukan dengan perangkat desa untuk memperoleh informasi mengenai mekanisme penyaluran BLT, kriteria yang digunakan dalam menentukan penerima bantuan, serta berbagai permasalahan yang sering terjadi dalam proses seleksi penerima bantuan. Selain itu, dokumentasi dilakukan dengan mengumpulkan data administrasi desa yang berkaitan dengan kondisi sosial ekonomi masyarakat sebagai bahan utama dalam penelitian ini.

Jumlah data yang digunakan dalam penelitian ini sebanyak 524 data Masyarakat. Seluruh data tersebut kemudian melalui tahap pengolahan data (preprocessing) agar data yang digunakan memiliki kualitas yang baik dan siap di proses oleh algoritma klasifikasi. Tahap pengolahan data meliputi pembersihan data (data cleaning), penanganan data kosong atau tidak lengkap, transformasi data kategorikal menjadi numerik, serta normalisasi data khusus pada algoritma KNN agar perhitungan jarak antar data dapat dilakukan dengan lebih optimal.

Setelah proses pengolahan data selesai dilakukan, dataset kemudian dibagi menjadi dua bagian yaitu data latih (training data) dan data uji (testing data) dengan perbandingan 80% : 20%. Sebanyak 80% data digunakan sebagai data latih untuk membangun model klasifikasi dan mempelajari pola hubungan antar variabel, sedangkan 20% data digunakan sebagai data uji untuk mengukur kemampuan model dalam melakukan prediksi terhadap data baru. Pembagian dataset ini dilakukan agar proses evaluasi model dapat dilakukan secara objektif dan hasil

klasifikasi yang diperoleh dapat menggambarkan performa algoritma secara lebih akurat dalam menentukan kelayakan penerima BLT di Desa Sungai Ambat.

4.1 Data Awal

Data awal yang digunakan dalam penelitian ini merupakan data masyarakat Desa Sungai Ambat yang diperoleh dari kantor desa melalui proses dokumentasi dan wawancara dengan perangkat desa. Data tersebut digunakan sebagai dasar dalam proses klasifikasi kelayakan penerima Bantuan Langsung Tunai (BLT) menggunakan algoritma Naive Bayes dan K-Nearest Neighbor (KNN).

Dataset penelitian terdiri dari 524 data masyarakat dengan berbagai atribut sosial ekonomi yang berkaitan dengan kondisi kesejahteraan rumah tangga. Data yang diperoleh masih berupa data mentah (raw data) sehingga perlu dilakukan proses pengolahan terlebih dahulu sebelum digunakan pada tahap implementasi algoritma. Data awal ini mencakup beberapa variabel penting yang digunakan sebagai atribut masukan dalam proses klasifikasi, yaitu usia kepala keluarga, jumlah anggota rumah tangga, jenis kelamin kepala keluarga, pendidikan kepala keluarga, pendapatan bulanan, status pekerjaan kepala keluarga, status penerima bantuan sosial lain, kepemilikan aset dasar, kondisi rumah, akses air bersih, akses listrik, keberadaan lansia, dan keberadaan penyandang disabilitas.

Pada tahap awal, data yang diperoleh dari kantor desa masih memiliki beberapa serta beberapa data kategorikal yang belum sesuai untuk diproses oleh algoritma machine learning. Oleh karena itu, dilakukan tahap preprocessing untuk membersihkan dan menyesuaikan data agar siap digunakan dalam proses klasifikasi. Berikut merupakan contoh data awal masyarakat sebelum dilakukan proses pengolahan data:

Tabel 4.1 Data Awal Penelitian

No	Nama Kepala Keluarga	Usia	Jumlah Tanggungan	L/P	Pendidikan	Pendapatan	Status Pekerjaan	Akses Air Bersih	Listrik	Penerima Bantuan Lain	Kepemilikan Aset	Kondisi Rumah	Keberadaan Lansia	Keberadaan Penyandang Disabilitas	Status Kelayakan
1	Ahmad Subhan	39	3	L	SLTA	Rp. 1.000.000	Wiraswasta	Sumur	Ada	Tidak Ada	Ada	Layak	Tidak Ada	Tidak Ada	Tidak Layak
2	Ahmad Junaidi	49	3	L	SLTA	Rp. 1.000.000	Petani	Sumur	Ada	Tidak Ada	Ada	Layak	Tidak Ada	Tidak Ada	Tidak Layak
3	Abdullah Purnomo	27	2	L	SLTA	Rp. 1.000.000	Wiraswasta	Sumur	Ada	Tidak Ada	Ada	Layak	Tidak Ada	Tidak Ada	Tidak Layak
4	Kholilur Rahman	64	2	L	SLTA	Rp. 500.000	Petani	Sumur	Ada	Tidak Ada	Tidak Ada	Layak	Tidak Ada	Tidak Ada	Layak
5	Imam Syafi'i	47	4	L	SLTA	Rp. 1.000.000	Petani	Sumur	Ada	Tidak Ada	Ada	Layak	Ada	Tidak Ada	Tidak Layak
6	Mustakim	37	3	L	SLTP	Rp. 500.000	Petani	Sumur	Ada	Tidak Ada	Ada	Layak	Tidak Ada	Tidak Ada	Layak
7	Misriadi	53	3	L	TIDAK SEKOLAH	Rp. 500.000	Petani	Sumur	Ada	Tidak Ada	Ada	Layak	Tidak Ada	Tidak Ada	Layak
8	Muhlis Abdul Fatah	35	1	L	SD	Rp. 1.000.000	Petani	Sumur	Ada	Tidak Ada	Ada	Layak	Tidak Ada	Tidak Ada	Tidak Layak
9	Supartin	62	5	P	SLTP	Rp. 500.000	Irt	Sumur	Ada	Tidak Ada	Tidak Ada	Kurang Layak	Ada	Tidak Ada	Layak
10	Abdul Hakim	34	2	L	SD	Rp. 1.000.000	Petani	Sumur	Ada	Tidak Ada	Ada	Layak	Ada	Tidak Ada	Tidak Layak
11	Wahidin	46	3	L	SLTP	Rp. 1.000.000	Petani	Sumur	Ada	Tidak Ada	Ada	Layak	Tidak Ada	Tidak Ada	Tidak Layak
12	M. Aminuddin	49	3	L	D4	Rp. 3.000.000	Wiraswasta	Sumur	Ada	Tidak Ada	Ada	Layak	Tidak Ada	Tidak Ada	Tidak Layak
13	M.Thoha	92	3	L	SD	Rp. 1.000.000	Pedagang	Sumur	Ada	Tidak Ada	Ada	Layak	Ada	Tidak Ada	Tidak Layak
14	Muh. Mujib	67	7	L	SLTA	Rp. 1.000.000	Perangkat Desa	Sumur	Ada	Tidak Ada	Ada	Layak	Tidak Ada	Tidak Ada	Tidak Layak
15	Nurwakhid	57	2	L	ALTA	Rp. 1.000.000	Petani	Sumur	Ada	Tidak Ada	Ada	Layak	Tidak Ada	Tidak Ada	Tidak Layak

Sumber : Data Ekonomi Warga Desa Sungai Ambat

262 Berdasarkan tabel 4.1 menyajikan data awal penelitian yang di peroleh dari
Data Ekonomi Warga Desa Sungai Ambat. Dapat dilihat bahwa sebagian besar
masyarakat yang berstatus layak menerima bantuan memiliki karakteristik seperti
pendapatan bulanan rendah, pekerjaan sebagai petani atau buruh, jumlah
tanggungan keluarga yang cukup banyak, serta kondisi rumah yang kurang layak.
89 Selain itu, keberadaan lansia dalam rumah tangga juga menjadi salah satu faktor
yang mempengaruhi tingkat kelayakan penerima bantuan. Sebaliknya, masyarakat
265 dengan kondisi ekonomi yang lebih baik seperti memiliki pekerjaan tetap,
pendapatan lebih tinggi, dan kondisi rumah layak cenderung masuk dalam kategori
tidak layak menerima bantuan.

69 Berdasarkan data pada tabel 4.1 usia keluarga berada pada rentang 27 hingga
92 tahun dengan jumlah tanggungan antara 1 sampai 7 orang. Sebagian besar data
berjenis kelamin laki-laki, sedangkan hanya satu data yang berjenis perempuan.
Tingkat pendidikan data cukup beragam, mulai dari Tidak Sekolah, SD,SLTP,
SLTA,hingga Diploma IV (D4). Sementara itu, jenis kelamin di dominasi oleh
259 Petani, disusul Wiraswasta, Pedagang, Perangkat Desa, Dan Ibu Rumah Tangga.
Dari data tersebut dapat diketahui bahwa penentuan kelayakan penerima BLT tidak
hanya berdasarkan satu variabel saja, tetapi mempertimbangkan berbagai kondisi
sosial ekonomi masyarakat secara menyeluruh. Oleh karena itu, penggunaan
27 algoritma klasifikasi diharapkan mampu membantu pemerintah desa dalam
menentukan penerima bantuan secara lebih objektif, tepat sasaran, dan transparan.

43 4.2 Preprocessing Data

28 Preprocessing data merupakan tahapan penting dalam penelitian yang
dilakukan sebelum data digunakan pada proses klasifikasi menggunakan algoritma

Naive Bayes dan K-Nearest Neighbor (KNN). Tahap ini bertujuan untuk meningkatkan kualitas data agar data menjadi lebih bersih, konsisten, terstruktur, dan siap diproses oleh model model machine learning. Data yang diperoleh dari Kantor Desa Sungai Ambat masih berupa data mentah (raw data), sehingga diperlukan beberapa tahap pengolahan agar hasil klasifikasi yang diperoleh menjadi lebih akurat dan optimal.

Pada penelitian ini, preprocessing data dilakukan terhadap dataset masyarakat Desa Sungai Ambat yang berjumlah 524 data. Dataset tersebut terdiri dari berbagai atribut sosial ekonomi masyarakat yang digunakan sebagai dasar dalam menentukan kelayakan penerima Bantuan Langsung Tunai (BLT). Tahapan preprocessing data yang dilakukan dalam penelitian ini meliputi data cleaning, transformasi data, normalisasi data, pemilihan fitur, dan pembagian dataset.

4.2.1 Data Cleaning

Tahap data cleaning dilakukan untuk membersihkan data dari kesalahan, ketidaksesuaian maupun data yang tidak lengkap. Pada penelitian ini, data yang digunakan berasal Dari Data Ekonomi Warga Desa Sungai Ambat yang berisi informasi mengenai kondisi sosia dan ekonomi setiap kepala keluarga. Proses ini bertujuan agar data yang digunakan benar-benar valid dan tidak mempengaruhi performa model klasifikasi. Adapun proses yang dilakukan pada tahap ini meliputi, yaitu : menghapus data duplikat, menangani missing value atau data kosong, memperbaiki kesalahan penulisan data, menyeragamkan format data.

Melalui proses data cleaning tersebut, data ekonomi warga desa sungai ambat menjadi lebih bersih, konsisten, dan memiliki kualitas yang lebih baik. Hasil dari tahap ini adalah dataset yang siap digunakan pada proses data transformation dan

penerapan algoritma klasifikasi. Dengan demikian, proses data cleaning diharapkan mampu meningkatkan kualitas data sehingga model klasifikasi yang dibangun dapat menghasilkan tingkat akurasi yang lebih baik dalam menentukan status kelayakan penerima bantuan sosial.

4.2.2 Transformasi Data

Transformasi data dilakukan untuk mengubah data kedalam formasi yang sesuai dengan kebutuhan algoritma machine learning. Beberapa atribut pada dataset masih berbentuk kategorikal sehingga perlu di ubah menjadi numerik menggunakan teknik encoding, yaitu mengubah setiap kategori pada atribut menjadi representasi berupa angka. Teknik ini bertujuan agar algoritma machine learning mampu membaca, mengolah dan membedakan setiap kategori berdasarkan nilai numerik yang telah ditentukan. Sebagai contoh, pada atribut jenis kelamin, kategori laki-laki dapat dikodekan menjadi nilai 1, sedangkan perempuan menjadi nilai 0. Demikian pula pada atribut listrik, kategori ada dapat diberikan nilai 1 dan tidak ada diberi nilai 0. Selain itu, atribut yang memiliki lebih dari dua kategori, seperti pendidikan, status pekerjaan, listrik, kondisi rumah, kepemilikan aset, keberadaan lansia, keberadaan disabilitas juga ditransformasikan ke dalam bentuk numerik.

Hasil dari tahap transformasi adalah dataset yang seluruh atributnya telah berada dalam format numerik dan siap digunakan pada proses pembentukan model klasifikasi. Dataset yang telah ditransformasikan dihadapkan mampu meningkatkan efektivitas proses komputasi, mempermudah algoritma dalam mengenali pola dari data ekonomi warga desa sungai ambat, serta menghasilkan model klasifikasi yang memiliki tingkat akurasi yang lebih baik dalam menentukan status kelayakan penerima bantuan sosial.

Tabel 4.2 Data Yang Telah Di Encoding

No	Usia	Jumlah Tanggungan	L/P	Pendidikan	Pendapatan	Status Pekerjaan	Akses Air Bersih	Listrik	Penerima Bantuan Lain	Kepemilikan Aset	Kondisi Rumah	Keberadaan Lansia	Keberadaan Penyandang Disabilitas	Status Kelayakan
1	39	3	0	4	Rp. 1.000.000	14	0	0	0	0	1	1	1	1
2	49	3	0	4	Rp. 1.000.000	9	0	0	0	0	1	1	1	1
3	27	2	0	4	Rp. 1.000.000	14	0	0	0	0	1	1	1	1
4	64	2	0	4	Rp. 500.000	9	0	0	0	1	1	1	1	0
5	47	4	0	4	Rp. 1.000.000	9	0	0	0	0	1	0	1	1
6	37	3	0	5	Rp. 500.000	9	0	0	0	0	1	1	1	0
7	53	3	0	6	Rp. 500.000	9	0	0	0	0	1	1	1	0
8	35	1	0	3	Rp. 1.000.000	9	0	0	0	0	1	1	1	1
9	62	5	1	5	Rp. 500.000	3	0	0	0	1	0	0	1	0
10	34	2	0	3	Rp. 1.000.000	9	0	0	0	0	1	0	1	1
11	46	3	0	5	Rp. 1.000.000	9	0	0	0	0	1	1	1	1
12	49	3	0	1	Rp. 3.000.000	14	0	0	0	0	1	1	1	1
13	92	3	0	3	Rp. 1.000.000	7	0	0	0	0	1	1	1	1
14	67	7	0	4	Rp. 1.000.000	8	0	0	0	0	1	1	1	1
15	57	2	0	4	Rp. 1.000.000	9	0	0	0	0	1	1	1	1

Sumber : Hasil encoding di google colab

Tabel 4.2 menyajikan hasil proses *encoding* terhadap data ekonomi warga Desa Sungai Ambat. Proses *encoding* dilakukan sebagai bagian dari tahapan transformasi data untuk mengubah atribut yang semula berbentuk kategorikal menjadi representasi numerik sehingga dapat diproses oleh algoritma klasifikasi. Pada penelitian ini, proses *encoding* dilakukan menggunakan Google Colab dengan bantuan pustaka Python, sehingga setiap kategori pada atribut diberikan nilai numerik secara konsisten. Pemberian kode numerik tidak mengubah makna maupun karakteristik dari setiap atribut, tetapi hanya mengubah bentuk penyimpanannya agar dapat dikenali oleh sistem. Berdasarkan Tabel 4.2, atribut kategorikal seperti Jenis Kelamin, Pendidikan, Status Pekerjaan, Akses Air Bersih, Listrik, Penerima Bantuan Lain, Kepemilikan Aset, Kondisi Rumah, Keberadaan Lansia, dan Keberadaan Penyandang Disabilitas telah dikonversi menjadi nilai numerik. Sementara itu, atribut numerik seperti Usia dan Jumlah Tanggungan tidak mengalami perubahan karena telah berada dalam format numerik dan dapat langsung digunakan dalam proses klasifikasi. Atribut Pendapatan pada tahap ini masih ditampilkan dalam bentuk nominal rupiah dan selanjutnya dapat dilakukan proses pengkodean atau normalisasi sesuai kebutuhan algoritma yang digunakan. Hasil proses encoding menghasilkan dataset yang telah direpresentasikan dalam bentuk numerik sehingga siap digunakan sebagai data masukan pada proses klasifikasi menggunakan algoritma Naïve Bayes dan K-Nearest Neighbor (K-NN).

4.2.3 Normalisasi Data

Normalisasi data dilakukan khusus pada algoritma KNN karena metode ini sensitif terhadap skala data. Tujuan normalisasi adalah untuk menyamakan rentang nilai antar atribut sehingga tidak ada atribut yang memiliki pengaruh lebih besar dibandingkan atribut lainnya.

Dalam penelitian ini digunakan metode Min-Max Normalization dengan rumus sebagai berikut :

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Keterangan

X_{norm} = Nilai hasil normalisasi

X = Nilai data asli

X_{min} = Nilai minimum atribut

X_{max} = Nilai maksimal atribut

Normalisasi data input bertujuan untuk menyesuaikan nilai range data dengan fungsi aktivasi dalam sistem Extreme Learning Mechine(ELM). Ini berarti nilai kurat input harus berada pada range 0 sampai 1. Sehingga range input yang emenuhi syarat adalah nilai data input dari 0 sampai 1. Oleh karena itu output yang dihasilkan pun akan berada pada range 0 sampai 1. Kemudian untuk mendapatkan nilai sebenarnya dari output perlu dilakukan normalisasi. Setelah data sampel diketahui, berikutnya dilakukan normalisasi data pada masing-masing variabel dengan tujuan agar menyamakan data dalam bentuk rataaan sehingga antara data yang satu dengan data yang lain memiliki range yang berbedekatan.

Data yang digunakan dalam penelitian ini merupakan data kondisi ekonomi masyarakat Desa Sungai Ambat yang telah melalui tahap normalisasi. Hasil normalisasi menghasilkan nilai setiap atribut berada pada rentang tertentu sehingga lebih sesuai untuk digunakan dalam proses pelatihan dan pengujian algoritma klasifikasi. Adapun data sampel kondisi ekonomi Desa Sungai Ambat yang telah dinormalisasikan dapat dilihat pada Tabel 4.4 di bawah ini

Tabel 4.3 Data Telah Di Normalisasi

No	Usia	Jumlah Tanggungan	L/P	Pendidikan	Pendapatan	Status Pekerjaan	Akses Air Bersih	Listrik	Penerima Bantuan Lain	Kepemilikan Aset	Kondisi Rumah	Keberadaan Lansia	Keberadaan Penyandang Disabilitas	Status Kelayakan
1	0.197183	0.375	0	4	0.2	14	0	0	0	0	1	1	1	1
2	0.338028	0.375	0	4	0.2	9	0	0	0	0	1	1	1	1
3	0.028169	0.25	0	4	0.2	14	0	0	0	0	1	1	1	1
4	0.549296	0.25	0	4	0	9	0	0	0	1	1	1	1	0
5	0.309859	0.5	0	4	0.2	9	0	0	0	0	1	0	1	1
6	0.169014	0.375	0	5	0	9	0	0	0	0	1	1	1	0
7	0.394366	0.375	0	6	0	9	0	0	0	0	1	1	1	0
8	0.140845	0.125	0	3	0.2	9	0	0	0	0	1	1	1	1
9	0.521127	0.625	1	5	0	3	0	0	0	1	0	0	1	0
10	0.126761	0.25	0	3	0.2	9	0	0	0	0	1	0	1	1
11	0.295775	0.375	0	5	0.2	9	0	0	0	0	1	1	1	1
12	0.338028	0.75	0	1	1	14	0	0	0	0	1	1	1	1
13	0.943662	0.375	0	3	0.2	7	0	0	0	0	1	1	1	1
14	0.591549	0.375	0	4	0.2	8	0	0	0	0	1	1	1	1
15	0.450704	0.875	0	4	0.2	9	0	0	0	0	1	1	1	1

Sumber : Hasil encoding di google colab

281
5

Data sampel yang digambarkan pada tabel 4.4 diatas merupakan hasil dari normalisasi data kondisi ekonomi Desa Sungai Ambat. Pada penelitian ini, proses normalisasi tidak diterapkan pada seluruh atribut. Normalisasi hanya dilakukan pada atribut Usia Kepala Keluarga, Jumlah Tanggungan, dan Pendapatan Bulanan karena ketiga atribut tersebut merupakan data numerik yang memiliki rentang nilai yang berbeda dan cukup besar. Adapun perhitungan normalisasi dapat dilakukan dengan cara sebagai berikut :

Perhitungan normalisasi:

Data 1

$$\text{Usia} = 39 \quad X' = \frac{39-25}{96-25} = \frac{14}{71} = 0,197183$$

$$\text{Tanggungan} = 3 \quad X' = \frac{3-0}{8-0} = \frac{3}{8} = 0,375$$

$$\text{Pendapatan} = \text{Rp } 1.000.000 \quad X' = \frac{1.000.000-500.000}{3.000.000-500.000} = \frac{500.000}{2.500.000} = 0,2$$

Data 2

$$\text{Usia} = 49 \quad X' = \frac{49-25}{96-25} = \frac{24}{71} = 0,338028$$

$$\text{Tanggungan} = 3 \quad X' = \frac{3-0}{8-0} = \frac{3}{8} = 0,375$$

$$\text{Pendapatan} = \text{Rp } 1.000.000 \quad X' = \frac{1.000.000-500.000}{3.000.000-500.000} = \frac{500.000}{2.500.000} = 0,2$$

Data 3

$$\text{Usia} = 27 \quad X' = \frac{27-25}{96-25} = \frac{2}{71} = 0,028169$$

$$\text{Tanggungan} = 2 \quad X' = \frac{2-0}{8-0} = \frac{2}{8} = 0,25$$

$$\text{Pendapatan} = \text{Rp } 1.000.000 \quad X' = \frac{1.000.000 - 500.000}{3.000.000 - 500.000} = \frac{500.000}{2.500.000} = 0,2$$

Data 4

$$\text{Usia} = 64 \quad X' = \frac{64 - 25}{96 - 25} = \frac{39}{71} = 0,549296$$

$$\text{Tanggungan} = 2 \quad X' = \frac{2 - 0}{8 - 0} = \frac{2}{8} = 0,25$$

$$\text{Pendapatan} = \text{Rp } 500.000 \quad X' = \frac{500.000 - 500.000}{3.000.000 - 500.000} = \frac{0}{2.500.000} = 0$$

Data 5

$$\text{Usia} = 47 \quad X' = \frac{47 - 25}{96 - 25} = \frac{22}{71} = 0,309859$$

$$\text{Tanggungan} = 4 \quad X' = \frac{4 - 0}{8 - 0} = \frac{4}{8} = 0,5$$

$$\text{Pendapatan} = \text{Rp } 1.000.000 \quad X' = \frac{1.000.000 - 500.000}{3.000.000 - 500.000} = \frac{500.000}{2.500.000} = 0,2$$

Data 6

$$\text{Usia} = 37 \quad X' = \frac{37 - 25}{96 - 25} = \frac{12}{71} = 0,169014$$

$$\text{Tanggungan} = 3 \quad X' = \frac{3 - 0}{8 - 0} = \frac{3}{8} = 0,375$$

$$\text{Pendapatan} = \text{Rp } 500.000 \quad X' = \frac{500.000 - 500.000}{3.000.000 - 500.000} = \frac{0}{2.500.000} = 0$$

Data 7

$$\text{Usia} = 53 \quad X' = \frac{53 - 25}{96 - 25} = \frac{28}{71} = 0,394366$$

$$\text{Tanggungan} = 3 \quad X' = \frac{3 - 0}{8 - 0} = \frac{3}{8} = 0,375$$

$$\text{Pendapatan} = \text{Rp } 500.000 \quad X' = \frac{500.000 - 500.000}{3.000.000 - 500.000} = \frac{0}{2.500.000} = 0$$

Data 8

$$\text{Usia} = 35 \quad X' = \frac{35 - 25}{96 - 25} = \frac{10}{71} = 0,140845$$

$$\text{Tanggungan} = 1 \quad X' = \frac{1 - 0}{8 - 0} = \frac{1}{8} = 0,125$$

$$\text{Pendapatan} = \text{Rp } 1.000.000 \quad X' = \frac{1.000.000 - 500.000}{3.000.000 - 500.000} = \frac{500.000}{2.500.000} = 0,2$$

Data 9

$$\text{Usia} = 62 \quad X' = \frac{62 - 25}{96 - 25} = \frac{37}{71} = 0,521127$$

$$\text{Tanggungan} = 5 \quad X' = \frac{5 - 0}{8 - 0} = \frac{5}{8} = 0,625$$

$$\text{Pendapatan} = \text{Rp } 500.000 \quad X' = \frac{500.000 - 500.000}{3.000.000 - 500.000} = \frac{0}{2.500.000} = 0$$

Data 10

$$\text{Usia} = 34 \quad X' = \frac{34 - 25}{96 - 25} = \frac{9}{71} = 0,126761$$

$$\text{Tanggungan} = 2 \quad X' = \frac{2 - 0}{8 - 0} = \frac{2}{8} = 0,25$$

$$\text{Pendapatan} = \text{Rp } 1.000.000 \quad X' = \frac{1.000.000 - 500.000}{3.000.000 - 500.000} = \frac{500.000}{2.500.000} = 0,2$$

4.2.4 Feature Selection

Tahap feature selection dilakukan untuk memilih atribut yang paling relevan terhadap proses klasifikasi kelayakan penerima BLT. Pemilihan fitur bertujuan untuk meningkatkan performa model serta mengurangi atribut yang tidak memiliki pengaruh signifikan terhadap hasil klasifikasi.

Variabel yang digunakan dalam penelitian ini meliputi:

1. Usia Kepala Keluarga
2. Jumlah Anggota Rumah Tangga
3. Jenis Kelamin Kepala Keluarga
4. Pendidikan Kepala Keluarga
5. Pendapatan Bulanan
6. Status Pekerjaan Kepala Keluarga
7. Penerima Bantuan Sosial Lain
8. Kepemilikan Aset Dasar
9. Kelayakan Kondisi Rumah
10. Akses Air Bersih
11. Akses Listrik
12. Keberadaan Lansia
13. Keberadaan Penyandang Disabilitas

Sedangkan variabel output yang digunakan yaitu status kelayakan penerima BLT dengan dua kategori:

1. Layak
2. Tidak Layak

4.2.5 Pembagian Dataset

Setelah seluruh proses preprocessing selesai dilakukan, dataset kemudian dibagi menjadi dua bagian yaitu data latih (training data) dan data uji (testing data). Perbandingan data dilakukan dengan perbandingan 80% data latih 20% data uji. Data latih digunakan untuk membangun model klasifikasi dan mempelajari pola

hubungan antar atribut, sedangkan data uji digunakan untuk mengukur kemampuan model dalam memprediksi data baru secara objektif.

Melalui tahapan preprocessing data ini, dataset yang digunakan dalam penelitian menjadilebih siap dan optimal untuk diterapkan kedalam algoritma Naïve Bayes dan K-Nearest Neighbor (KNN), sehingga diharapkan mampumenghasilkan klasifikasi penerima bantuan langsung tunai (BLT) yang lebih akurat, objektif, dan tepat sasaran di Desa Sungai Ambat.

Sebelum dilakukan proses pelatihan model, data penelitian terlebih dahulu dibagi menjadi data training dan data testing menggunakan metode split data dengan perbandingan 80:20. Dari total 524 data, sebanyak 419 data digunakan sebagai data training dan 105 data digunakan sebagai data testing.

Tabel 4.4 Distribusi Kelas Keseluruhan

Kelas	Jumlah data
Tidak layak	453
Layak	71
Total	524

Berdasarkan distribusi data, diketahui bahwa jumlah data pada setiap kelas masih belum seimbang. Kelas tidak layak memiliki jumlah data paling banyak yaitu 453 data, sedangkan kelas layak sebanyak 71 data. Ketidakseimbangan data ini dapat mempengaruhi performa model dalam melakukan klasifikasi karena model cenderung lebih mudah mengenali kelas mayoritas dibandingkan kelas minoritas.

Tabel 4.5 Pembagian Data Training dan Data Testing

Jenis Data	Jumlah Data
Training	419
Testing	105

68

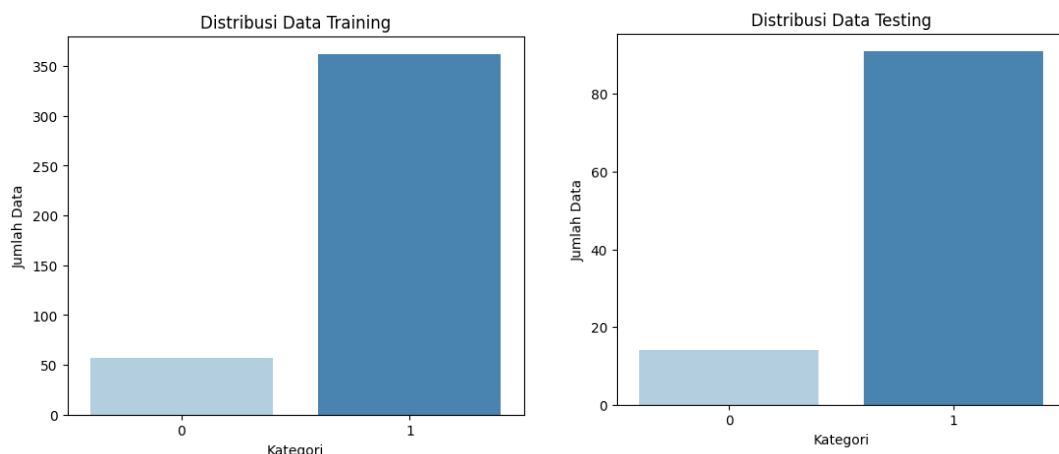
Tabel 4.6 Distribusi Data Training

Kelas Data	Jumlah Data
Tidak Layak	362
Layak	57
Total	419

68

Tabel 4.7 Distribusi Data Testing

Kelas Data	Jumlah Data
Tidak Layak	91
Layak	14
Total	105



100

Gambar 4.1 Diagram Data Training dan Data Testing

256

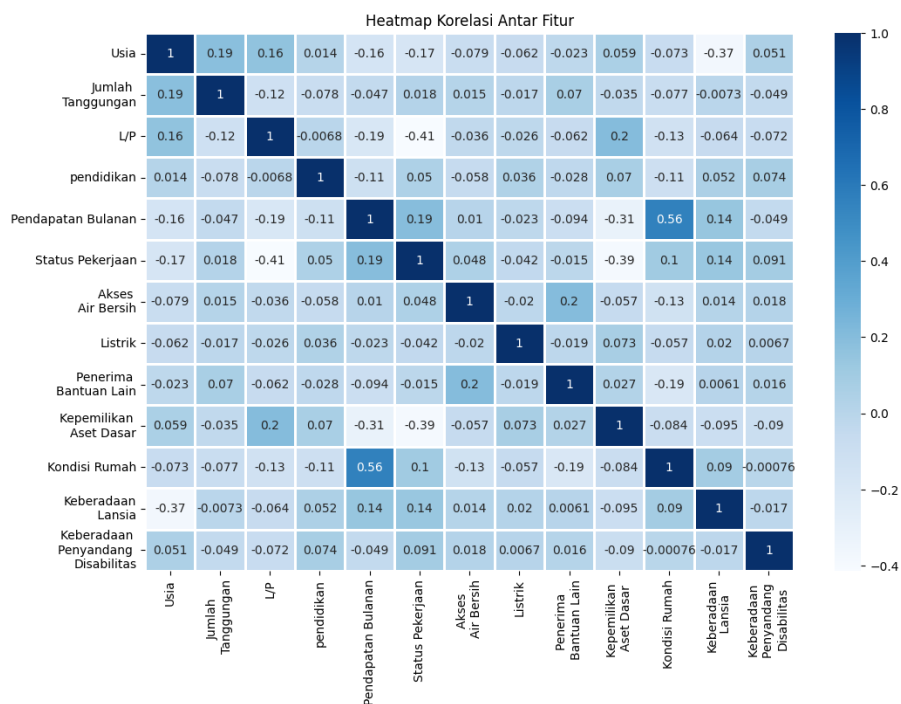
22

35

43

Berdasarkan Gambar 4.1 dapat dilihat distribusi data training dan data testing yang digunakan dalam penelitian klasifikasi penerima Bantuan Langsung Tunai (BLT). Dataset penelitian dibagi menjadi dua bagian yaitu data training sebesar 80% dan data testing sebesar 20%. Pembagian data ini bertujuan agar model machine learning dapat mempelajari pola data dari data training dan kemudian diuji menggunakan data testing untuk mengetahui performa model dalam melakukan prediksi terhadap data baru.

Pada grafik Distribusi Data Training, terlihat bahwa kategori **1** memiliki jumlah data yang jauh lebih banyak dibandingkan kategori **0**. Kategori **1** memiliki jumlah sekitar 362 data, sedangkan kategori **0** hanya sekitar 57 data. Sementara itu, pada grafik Distribusi Data Testing, pola distribusi data juga menunjukkan kondisi yang hampir sama. Kategori **1** (Tidak Layak) memiliki jumlah data sekitar 91 data, sedangkan kategori **0** (Layak) hanya sekitar 14 data. Distribusi tersebut menunjukkan bahwa data testing tetap mempertahankan proporsi kelas yang hampir sama dengan data training.



Gambar 4.2 Heatmap Korelasi Antar Fitur

Berdasarkan heatmap korelasi antar fitur, sebagai besar variabel menunjukkan hubungan yang lemah hingga sedang. Korelasi positif tertinggi ditemukan antar pendapatan bulanan dan kondisi rumah sebesar 0,56, yang menunjukkan bahwa peningkatan pendapatan cenderung diikuti oleh kondisi rumah yang lebih baik. Sementara itu, korelasi negatif terbesar terjadi antara variabel L/P

dan status pekerjaan sebesar -0,41. Secara keseluruhan tidak terdapat korelasi yang sangat kuat antara variabel, sehingga risiko multikolinearitas dalam dataset tergolong rendah dan seluruh fitur masih layak dipertimbangkan dalam proses analisis lebih lanjut.

4.3 Implementasi Algoritma

Tahap implementasi algoritma merupakan proses penerapan metode klasifikasi terhadap data yang telah melalui tahap preprocessing. Pada penelitian ini digunakan dua algoritma klasifikasi yaitu Naive Bayes dan K-Nearest Neighbor (KNN). Implementasi dilakukan menggunakan bahasa pemrograman Python pada Google Colab dengan memanfaatkan beberapa library seperti Pandas, NumPy, Scikit-Learn, dan Matplotlib.

Tujuan dari implementasi algoritma ini adalah untuk mengetahui kemampuan masing-masing metode dalam mengklasifikasikan kelayakan penerima Bantuan Langsung Tunai (BLT) berdasarkan data sosial ekonomi masyarakat Desa Sungai Ambat. Data yang telah dibersihkan, ditransformasikan, dan dinormalisasi kemudian dibagi menjadi data training dan data testing dengan perbandingan 80% : 20%.

4.3.1 Algoritma Naive Bayes

Algoritma Naive Bayes merupakan metode klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes dalam proses pengambilan keputusan. Algoritma ini bekerja dengan menghitung probabilitas setiap atribut terhadap kelas target, kemudian menentukan kelas dengan probabilitas tertinggi sebagai hasil prediksi. Dalam penelitian penentuan penerima Bantuan Langsung Tunai (BLT), algoritma Naive Bayes digunakan untuk mengklasifikasikan masyarakat ke dalam

27 kategori layak menerima bantuan atau “tidak layak menerima bantuan”. Proses klasifikasi dilakukan berdasarkan beberapa kriteria yang telah ditentukan, seperti tingkat penghasilan, jumlah tanggungan keluarga, kondisi rumah, status pekerjaan, kepemilikan aset, dan kondisi ekonomi lainnya. Data-data tersebut akan dijadikan sebagai atribut yang digunakan dalam proses perhitungan probabilitas.

243 Melalui data training yang berasal dari data penerima bantuan sebelumnya, algoritma Naive Bayes akan mempelajari pola-pola tertentu untuk menentukan peluang kelayakan penerima BLT. Sebagai contoh, masyarakat dengan penghasilan rendah, jumlah tanggungan yang banyak, dan kondisi rumah tidak layak huni memiliki probabilitas lebih besar untuk dikategorikan sebagai penerima bantuan. Sebaliknya, masyarakat dengan penghasilan tinggi dan kondisi ekonomi yang stabil cenderung masuk ke dalam kategori tidak layak menerima BLT. Berikut ini langkah- langkah menghitung menggunakan Algoritma Naive Bayes :

1. Hitung Probabilitas Awal

140 Probabilitas awal menunjukkan seberapa besar kemungkinan suatu kelas muncul dalam keseluruhan data sebelum mempertimbangkan dengan fitur lainnya. Prior probability merupakan atribut-atribut yang dimiliki oleh data. Pada algoritma naïve bayes, nilai prior diperoleh dari proporsi jumlah data pada masing-masing kelas terhadap total data training. Berdasarkan hasil perbagian data, diperoleh data training sebanyak 419 data yang terdiri dari 362 data berstatus layak (1) 57 data berstatus tidak layak (0).

Prior Kelas Layak

Probabilitas awal kelas layak dihitung menggunakan rumus:

$$P(\text{Layak}) = \frac{\text{Jumlah Data Layak}}{\text{Jumlah Seluruh Data Training}}$$

$$P(\text{Layak}) = \frac{362}{419} = 0,8640 \text{ atau } P(\text{Layak}) = 86,40\%$$

Artinya, dari seluruh data training yang digunakan dalam penelitian, sebesar 86,40% data termasuk kedalam katagori layak menerima BLT

Prior Kelas Tidak Layak

Priobabilitas awal kelas tidak layak dihitung menggunakan rumus:

$$P(\text{Tidak Layak}) = \frac{\text{Jumlah Data Tidak Layak}}{\text{Jumlah Seluruh Data Training}}$$

$$P(\text{Tidak Layak}) = \frac{57}{419} = 0,1360 \text{ atau } P(\text{Tidak Layak}) = 13,60\%$$

Artinya, dari seluruh data training yang digunakan, sebesar 13,60% data termasuk ke dalam katagori tidak layak menerma BLT.

Berdasarkan hasil perhitungan prior probability, diketahui bahwa kelas Layak memiliki probabilitas awal sebesar 86,40%, sedangkan kelas Tidak Layak memiliki probabilitas awal sebesar 13,60%. Hal ini menunjukkan bahwa sebagian besar data pada dataset penelitian merupakan warga yang berstatus layak menerima Bantuan Langsung Tunai (BLT). Nilai prior ini akan digunakan oleh algoritma Naive Bayes bersama nilai likelihood untuk menghitung probabilitas posterior setiap kelas. Kelas dengan probabilitas posterior terbesar akan dipilih sebagai hasil prediksi status kelayakan penerima BLT.

2. Menghitung Likelihood

Likelihood merupakan probabilitas kemunculan suatu atribut terhadap kelas tertentu. Nilai likelihood digunakan untuk mengetahui seberapa besar peluang suatu atribut muncul pada kelas layak maupun tidak layak. Berikut ini hasil dari perhitungan likelihood pada metode Gaussia Naive Bayes pada tabel 4.8

Tabel.4.8 Hasil Likelihood

Atribut	1	0
Usia	1,982	0,601
Jumlah Tanggunagn	1,450	0,980
L/P	0,990	0,910
Pendidikan	0,720	0,510
Pendapatan Bulanan	1,210	0,880
Status Pekerjaan	0,810	0,640
Akses Air Bersih	0,980	0,920
Listrik	0,990	0,960
Penerima Bantuan Lain	0,970	0,930
Kepemilikan Aset	0,850	0,720
Kondisi Rumah	0,920	0,850
Keberadaan Lansia	0,940	0,900
Keberadaan Disabilitas	0,950	0,910

Berikut ini perhitungan setiap atribut, data tersebut sudah di normalisasi dan berupa angka kontinu, sehingga nilai likelihood dihitung menggunakan Gaussian Naive Bayes sebelum menghitung setiap atribut maka terlebih dahulu menghitung mean dan standar deviasinya sebagai berikut :

Mean kelas 1

$$\mu_1 = \frac{x_1 + x_2 + x_3 + \dots + x_{362}}{362} = 0,315$$

Mean kelas 0

$$\mu_0 = \frac{x_1 + x_2 + x_3 + \dots + x_{57}}{57} = 0,487$$

Standar Deviasi Kelas 1

$$\sigma_1 = \sqrt{\frac{\sum_{i=1}^{362} (x_i - \mu_1)^2}{362}} = \sqrt{\frac{7,617}{362}} = 0,145$$

Standar Deviasi Kelas 0

$$\sigma_0 = \sqrt{\frac{\sum_{i=1}^{57} (x_i - \mu_0)^2}{57}} = \sqrt{\frac{1,8468}{57}} = 0,180$$

Melalui perhitungan tersebut diperoleh nilai likelihood masing-masing atribut numerik terhadap setiap kelas. Berikut perhitungan atribut numerik :

Perhitungan atribut usia kelas 0:

a. Menghitung penyebut

$$\begin{aligned}\sigma^2 &= (0,145)^2 = 0,021025 \\ 2\pi(0,021025) &= 0,13210 \\ \sqrt{0,13210} &= 0,36346 \\ \frac{1}{0,36346} &= 2,751\end{aligned}$$

b. Menghitung eksponen

$$\begin{aligned}(0,197183 - 0,315)^2 &= 0,013881 \\ 2(0,021015)^2 &= 0,04205 \\ \frac{0,013881}{0,04205} &= 0,3301 \\ e^{-0,3301} &= 0,7204\end{aligned}$$

c. Menghitung likelihood

$$P(Usia | 0) = 2,751 \times 0,7204 = 1,982$$

Sehingga diperoleh

$$P(Usia | 0) = 1,982$$

a. Menghitung penyebut

$$\begin{aligned}\sigma^2 &= (0,180)^2 = 0,0324 \\ 2\pi(0,0324) &= 0,20358 \\ \sqrt{0,20358} &= 0,451208 \\ \frac{1}{0,45120} &= 2,216\end{aligned}$$

b. Menghitung eksponen

$$\begin{aligned}(0,197183 - 0,487)^2 &= 0,08399 \\ 2(0,180)^2 &= 0,0648 \\ \frac{0,08399}{0,0648} &= 1,296 \\ e^{-1,296} &= 0,271\end{aligned}$$

c. Menghitung likelihood

$$P(Usia | 0) = 2,216 \times 0,271 = 0,601$$

Sehingga diperoleh

$$P(Usia | 0) = 0,601$$

Pada penelitian ini, atribut seperti L/P, Pendidikan, Status Pekerjaan, Listrik, dan atribut lainnya memang merupakan data kategorikal. Pada penelitian ini semua dihitung dengan langkah yang sama menggunakan GaussianNB dari scikit-learn, semua atribut tersebut diperlakukan sebagai data numerik, sehingga nilai likelihood tetap dihitung menggunakan rumus Gaussian (berdasarkan mean dan standar deviasi), bukan berdasarkan frekuensi kemunculan kategori.

3. Menghitung Probabilitas Posterior

Setelah diperoleh nilai probabilitas prior dan likelihood dari setiap atribut, langkah selanjutnya adalah menghitung probabilitas posterior menggunakan metode Gaussian Naive Bayes. Probabilitas posterior merupakan probabilitas suatu data termasuk ke dalam kelas tertentu setelah mempertimbangkan seluruh informasi yang terdapat pada atribut-atribut yang digunakan dalam proses klasifikasi.

Pada metode Naive Bayes, probabilitas posterior dihitung berdasarkan Teorema Bayes dengan menggabungkan probabilitas awal (prior) dan probabilitas kemunculan setiap atribut terhadap masing-masing kelas (likelihood). Asumsi utama yang digunakan adalah bahwa setiap atribut bersifat independen (conditional

independence), sehingga probabilitas gabungan dapat diperoleh dengan mengalikan nilai likelihood dari seluruh atribut. Berikut ini hasil dari perhitungan probabilitas posterior dapat di lihat pada tabel 4.9

Tabel 4.9 Hasil Prdikasi

No	Data Asli	Hasil Prediksi
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	1	1
8	1	1
9	1	1
10	1	0
11	1	1
12	1	1
13	1	1
14	1	1
15	1	1

Secara matematis, probabilitas posterior dapat dihitung seperti di bawah ini agar dapat hasil prediksi seperti tabel 4.9 sebagai berikut:

Kelas 1

$$P(1|X) = 0,8640 \times 1,982 \times 1,450 \times 0,990 \times 0,720 \times 1,210 \times 0,810 \times 0,980 \\ \times 0,990 \times 0,970 \times 0,850 \times 0,920 \times 0,940 \times 0,950 = 1,1400$$

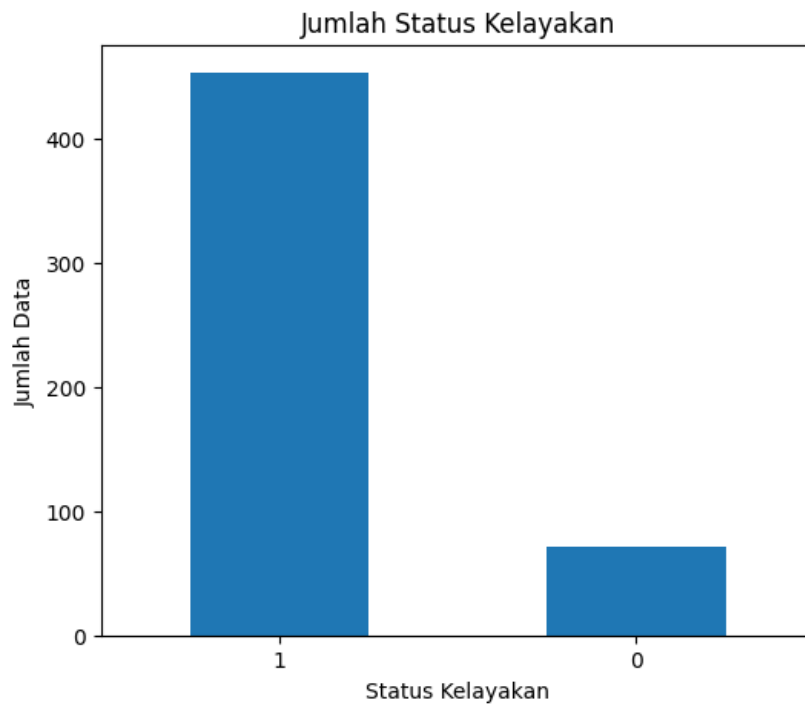
$$P(1 | X) = 1,1400$$

Kelas 0

$$P(0|X) = 0,1360 \times 0,601 \times 0,980 \times 0,910 \times 0,510 \times 0,880 \times 0,640 \times 0,920 \\ \times 0,960 \times 0,930 \times 0,720 \times 0,850 \times 0,900 \times 0,910 = 0,0086$$

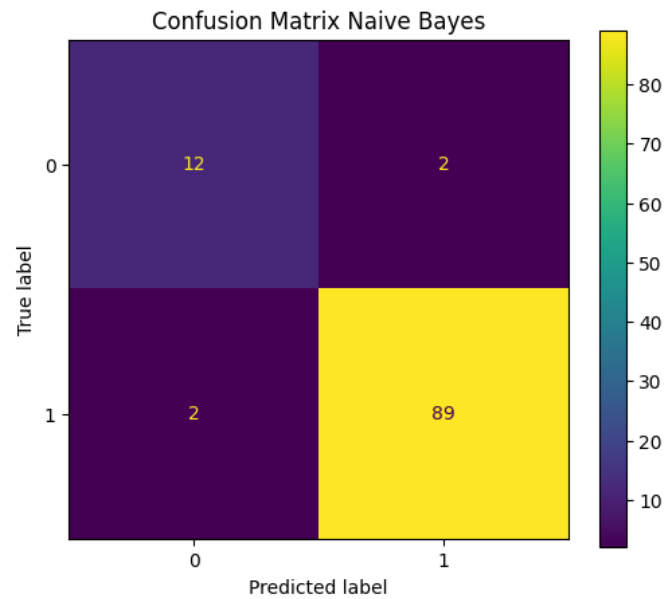
$$P(0 | X) = 0,0086$$

$$1,1400 > 0,0086 = 1$$



Gambar 4.3 Diagram Jumlah Status Kelayakan

Berdasarkan Gambar 4.3, terlihat bahwa distribusi data berdasarkan status kelayakan menunjukkan adanya perbedaan yang cukup signifikan antar kedua kelas. Kelas status kelayakan 1 (tidak layak) memiliki jumlah data yang jauh lebih banyak dibandingkan dengan status kelayakan 0 (layak). Dari diagram terlihat bahwa terdapat sekitar 453 data pada kelas layak, sedangkan kelas tidak layak hanya sekitar 71 data. Hal ini menunjukkan bahwa sebagian besar data dalam dataset termasuk ke dalam kategori tidak layak menerima bantuan BLT. Perbedaan jumlah data yang cukup signifikan ini menunjukkan adanya ketidakseimbangan kelas (class imbalance) pada dataset. Kondisi tersebut dapat memengaruhi kinerja model klasifikasi karena model cenderung lebih mudah memprediksi kelas mayoritas dibandingkan kelas minoritas. Oleh karena itu, pada tahap pemodelan perlu dilakukan evaluasi menggunakan metrik yang sesuai agar performa model dapat dinilai secara lebih objektif.



Gambar 4.4 Confusion Matrix Naive Bayes

Berdasarkan gambar *Confusion Matrix Naive Bayes*, dapat diketahui bahwa hasil klasifikasi algoritma Naive Bayes dalam menentukan penerima Bantuan Langsung Tunai (BLT) menunjukkan performa yang sangat baik. *Confusion matrix* digunakan untuk mengevaluasi tingkat ketepatan model klasifikasi dengan membandingkan antara data aktual (*true label*) dan hasil prediksi sistem (*predicted label*). Pada *confusion matrix* tersebut terdapat empat komponen utama, yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*. Nilai-nilai tersebut digunakan untuk mengetahui seberapa baik model dalam melakukan prediksi terhadap setiap kelas.

Berdasarkan hasil pengujian, diperoleh nilai *True Negative (TN)* sebesar 12, yang berarti terdapat 12 data yang sebenarnya termasuk kategori tidak layak menerima BLT (kelas 0) dan berhasil diprediksi dengan benar oleh sistem sebagai tidak layak. Hal ini menunjukkan bahwa algoritma mampu mengenali data masyarakat yang memang tidak memenuhi syarat penerima bantuan. Selanjutnya, terdapat nilai *False Positive (FP)* sebesar 2, yang menunjukkan bahwa terdapat 2

data yang sebenarnya tidak layak menerima bantuan, tetapi diprediksi oleh sistem sebagai layak menerima BLT. Kesalahan ini menunjukkan adanya data yang salah klasifikasi sehingga dapat menyebabkan bantuan diberikan kepada masyarakat yang kurang tepat sasaran.

Pada bagian *False Negative (FN)* diperoleh nilai sebesar 2, yang berarti terdapat 2 data yang sebenarnya layak menerima bantuan, namun diprediksi sebagai tidak layak oleh sistem. Kesalahan ini menunjukkan bahwa masih terdapat sebagian kecil masyarakat yang seharusnya menerima bantuan tetapi tidak teridentifikasi oleh model. Sementara itu, nilai *True Positive (TP)* sebesar 89 menunjukkan bahwa terdapat 89 data yang benar-benar layak menerima BLT dan berhasil diprediksi dengan benar oleh algoritma Naive Bayes. Nilai ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali masyarakat yang layak menerima bantuan.

Secara keseluruhan, hasil confusion matrix menunjukkan bahwa jumlah prediksi yang benar jauh lebih besar dibandingkan jumlah prediksi yang salah. Dari total 105 data pengujian, hanya terdapat 4 kesalahan klasifikasi, sedangkan 101 data berhasil diprediksi dengan benar. Hal ini membuktikan bahwa algoritma Naive Bayes memiliki tingkat akurasi yang tinggi dan mampu melakukan klasifikasi penerima BLT dengan baik. Dengan demikian, algoritma Naive Bayes dapat dikatakan efektif digunakan dalam sistem pendukung keputusan untuk menentukan penerima Bantuan Langsung Tunai (BLT). Model ini mampu membantu proses seleksi penerima bantuan secara lebih objektif, cepat, dan tepat sasaran sehingga dapat meminimalkan kesalahan dalam penyaluran bantuan sosial kepada masyarakat.

Untuk memperoleh Classification Report, tidak perlu menghitung satu per satu 105 data testing secara manual. Yang dihitung manual adalah hasil evaluasi dari prediksi model terhadap 105 data testing berikut ini penyelesaian nya :

1. Menghitung Precisin Kelas 0

Precision merupakan metrik yang digunakan untuk mengukur tingkat ketepatan model dalam memprediksi suatu kelas. Untuk kelas 0 (Tidak Layak), nilai yang digunakan adalah True Positive (TP) = 12, yaitu jumlah data tidak layak yang berhasil diprediksi sebagai tidak layak. False Positive (FP) = 2, yaitu jumlah data layak yang salah diprediksi sebagai tidak layak. Sehingga perhitungan precision kelas 0 yaitu:

$$Precision_0 = \frac{12}{12+2} = \frac{12}{14} = 0,8571 \approx 0,86$$

2. Menghitung Recall Kelas 0

Recall merupakan metrik evaluasi yang digunakan untuk mengukur kemampuan model dalam menemukan seluruh data yang sebenarnya termasuk ke dalam suatu kelas. Untuk menghitung menggunakan TP (True Positive) = 12, jumlah data tidak layak yang berhasil diprediksi sebagai tidak layak. FN (False Negative)= 2, jumlah data tidak layak yang salah diprediksi sebagai layak. Sehingga perhitungan recal kelas 0 adalah:

$$Recall_0 = \frac{12}{12+2} = \frac{12}{14} a = 0,8571 \approx 0,86$$

3. Menghitung F1- Score Kelas 0

F1-Score merupakan metrik evaluasi yang digunakan untuk mengukur keseimbangan antara nilai precision dan recall. Nilai F1-Score akan tinggi apabila model memiliki nilai precision dan recall yang sama-sama tinggi.

Berdasarkan hasil perhitungan sebelumnya di peroleh precision kelas 0 = 0,86 , recall kelas 0 = 0,86, sehingga perhitungan F1-Score kelas 0 adalah:

$$F1_0 = 2 \times \frac{0,8571 \times 0,8571}{0,8571 + 0,8571} = 0,8571 \approx 0,86$$

4. Menghitung Precision Kelas 1

Pada penelitian ini, precision kelas 1 menunjukkan seberapa besar ketepatan model dalam memprediksi data sebagai Layak dibandingkan dengan seluruh data yang diprediksi sebagai Layak. Untuk menghitung menggunakan TN (True Negative) = 89, FP (False Positive) = 2, sehingga perhitungan Precision kelas 1 adalah :

$$Precision_1 = \frac{89}{89+2} = \frac{89}{91} = 0,9780 \approx 0,98$$

5. Menghitung Recall Kelas 1

$$Precision_1 = \frac{89}{89+2} = \frac{89}{91} = 0,9780 \approx 0,98$$

6. Menghitung F1- Score Kelas 1

$$F1_1 = 2 \times \frac{0,9780 \times 0,9780}{0,9780 + 0,9780} = 0,9780 \approx 0,98$$

7. Menghitung Accuracy

$$Accuracy = \frac{89 + 12}{89+12+2+2} = \frac{101}{105} = 0,9619 \approx 96,19\%$$

8. Menghitung Macro Average

$$\text{Precision macro: } \frac{0,86+0,98}{2} = 0,92$$

$$\text{Recall macro: } \frac{0,86+0,98}{2} = 0,92$$

$$\text{F1 macro: } \frac{0,86+0,98}{2} = 0,92$$

9. Menghitung Weighted Average

Support kelas 0 = 14, kelas 1= 91 total 105

Weighted precision

$$\begin{aligned} & \frac{(0,86 \times 14) + (0,98 \times 91)}{105} \\ &= \frac{12,04 + 89,18}{105} \\ &= 0,964 \\ &= 0,96 \end{aligned}$$

Weighted recall

$$\begin{aligned} & \frac{(0,86 \times 14) + (0,98 \times 91)}{105} \\ &= \frac{12,04 + 89,18}{105} \\ &= 0,964 \\ &= 0,96 \end{aligned}$$

Weighted F1-Score

$$\begin{aligned} & \frac{(0,86 \times 14) + (0,98 \times 91)}{105} \\ &= \frac{12,04 + 89,18}{105} \\ &= 0,964 \\ &= 0,96 \end{aligned}$$

Tabel 4.10 classification report Naïve bayes

Classification Report:					
	precision	recall	f1-score	support	
0	0.86	0.86	0.86	14	
1	0.98	0.98	0.98	91	
accuracy			0.96	105	
macro avg	0.92	0.92	0.92	105	
weighted avg	0.96	0.96	0.96	105	

Berdasarkan hasil pengujian menggunakan Classification Report, diperoleh nilai evaluasi performa algoritma yang meliputi *precision*, *recall*, *f1-score*, dan *accuracy*. Hasil pengujian menunjukkan bahwa model klasifikasi yang digunakan memiliki tingkat kinerja yang sangat baik dalam menentukan kategori data penerima Bantuan Langsung Tunai (BLT).

Pada kelas 0 yang merepresentasikan kategori tidak layak menerima bantuan, diperoleh nilai *precision* sebesar 0,86, *recall* sebesar 0,86, dan *f1-score* sebesar 0,86 dengan jumlah data (*support*) sebanyak 14 data. Nilai *precision* sebesar 0,86 menunjukkan bahwa dari seluruh data yang diprediksi sebagai kategori tidak layak, sebanyak 86% berhasil diprediksi dengan benar oleh sistem. Nilai *recall* sebesar 0,86 menunjukkan bahwa sistem mampu mengenali 86% data tidak layak dari keseluruhan data tidak layak yang sebenarnya. Sedangkan nilai *f1-score* sebesar 0,86 menunjukkan keseimbangan yang baik antara *precision* dan *recall* pada kelas tersebut.

Pada kelas 1 yang merepresentasikan kategori layak menerima bantuan, diperoleh nilai *precision* sebesar 0,98, *recall* sebesar 0,98, dan *f1-score* sebesar 0,98 dengan jumlah data sebanyak 91 data. Nilai tersebut menunjukkan bahwa algoritma memiliki kemampuan yang sangat tinggi dalam mengidentifikasi masyarakat yang layak menerima BLT. *Precision* sebesar 0,98 menunjukkan bahwa 98% data yang diprediksi layak merupakan data yang benar-benar layak menerima bantuan. Selain itu, *recall* sebesar 0,98 menunjukkan bahwa sistem berhasil mengenali hampir seluruh data layak secara tepat. Nilai *f1-score* sebesar 0,98 juga menandakan keseimbangan yang sangat baik antara ketepatan prediksi dan kemampuan sistem dalam menemukan data yang sesuai. Secara keseluruhan, model klasifikasi

menghasilkan nilai *accuracy* sebesar 0,96 atau 96%, yang menunjukkan bahwa sistem mampu mengklasifikasikan data dengan tingkat ketepatan yang sangat tinggi. Dari total 105 data pengujian, sebagian besar data berhasil diklasifikasikan dengan benar sesuai kategorinya. Selain itu, nilai *macro average* sebesar 0,92 menunjukkan rata-rata performa model pada setiap kelas tanpa mempertimbangkan jumlah data pada masing-masing kelas. Sedangkan nilai *weighted average* sebesar 0,96 menunjukkan rata-rata performa yang mempertimbangkan proporsi jumlah data di setiap kelas. Tingginya nilai *weighted average* mengindikasikan bahwa model memiliki performa yang stabil dan baik dalam menangani keseluruhan data. Berdasarkan hasil evaluasi tersebut, dapat disimpulkan bahwa algoritma yang digunakan memiliki kemampuan klasifikasi yang sangat baik dalam menentukan penerima Bantuan Langsung Tunai (BLT). Model mampu memberikan prediksi secara akurat sehingga dapat membantu proses penyaluran bantuan menjadi lebih tepat sasaran dan objektif.

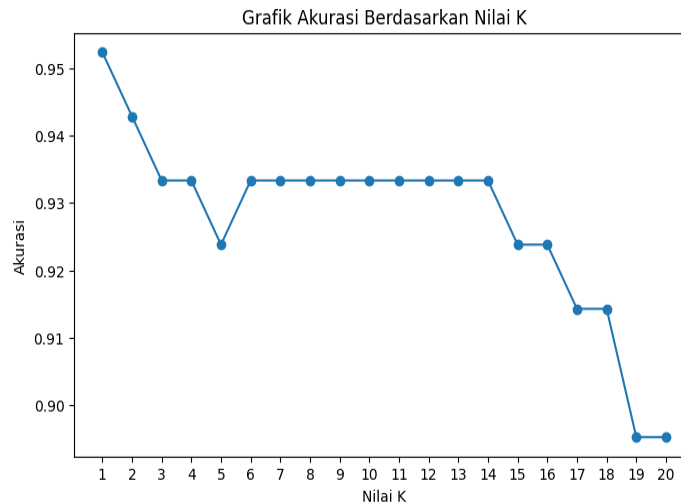
4.3.2 Algoritma K- Nearest Neighbor (KNN)

Algoritma K- Nearest Neighbor (KNN) digunakan sebagai algoritma pembandingan dalam penelitian ini. Algoritma K-Nearest Neighbor (KNN) digunakan sebagai algoritma pembandingan dalam penelitian ini untuk mengetahui tingkat performa dan akurasi metode klasifikasi dalam menentukan penerima Bantuan Langsung Tunai (BLT). Penggunaan KNN sebagai metode pembandingan bertujuan untuk mengevaluasi apakah algoritma Naive Bayes atau KNN memiliki kemampuan yang lebih baik dalam melakukan klasifikasi data masyarakat berdasarkan kriteria yang telah ditentukan. KNN dipilih karena merupakan salah satu algoritma klasifikasi yang cukup populer dan sering digunakan dalam

45 penelitian data mining. Algoritma ini bekerja dengan cara menentukan kelas suatu data berdasarkan kedekatan jarak terhadap data tetangga terdekat pada data training. Semakin dekat jarak suatu data dengan kelompok tertentu, maka semakin besar kemungkinan data tersebut masuk ke dalam kelompok yang sama.

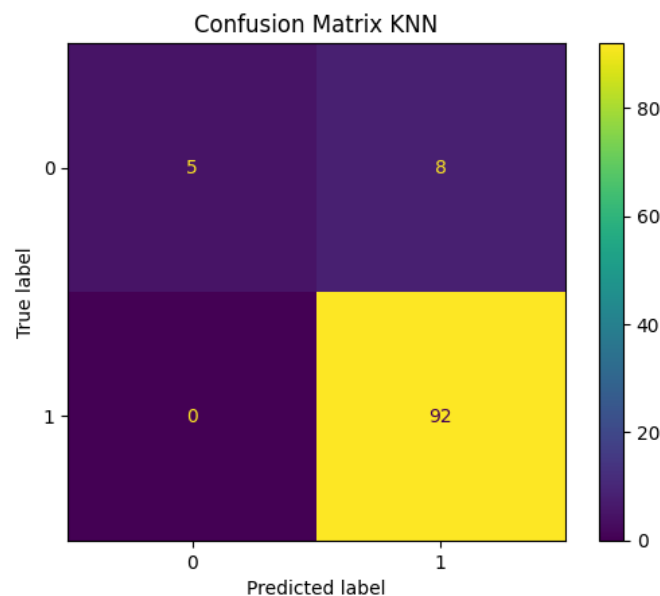
21 Dalam penelitian ini, algoritma KNN digunakan untuk mengklasifikasikan masyarakat ke dalam kategori “layak menerima bantuan” dan “tidak layak menerima bantuan” berdasarkan beberapa atribut, seperti penghasilan, jumlah tanggungan, kondisi rumah, status pekerjaan, dan kepemilikan aset. Proses klasifikasi dilakukan dengan menghitung jarak antara data baru dengan seluruh data training menggunakan metode perhitungan jarak tertentu, seperti Euclidean Distance. Selanjutnya, sistem akan mengambil sejumlah tetangga terdekat sesuai nilai K yang telah ditentukan untuk menentukan hasil klasifikasi berdasarkan mayoritas kelas.

12 Hasil klasifikasi dari algoritma KNN kemudian dibandingkan dengan hasil klasifikasi algoritma Naive Bayes menggunakan beberapa parameter evaluasi, seperti accuracy, precision, recall, f1-score, dan confusion matrix. Melalui proses perbandingan tersebut, dapat diketahui algoritma yang memiliki performa terbaik dalam membantu proses penentuan penerima Bantuan Langsung Tunai (BLT). 25 Penggunaan algoritma KNN sebagai metode pembanding diharapkan dapat memberikan hasil analisis yang lebih objektif dan memperkuat kesimpulan penelitian terkait efektivitas algoritma klasifikasi dalam sistem pendukung keputusan penyaluran bantuan sosial kepada masyarakat. 63 285



Gambar 4.5 Grafik Akurasi Berdasarkan Nilai k

Berdasarkan grafik akurasi K-NN, nilai $K = 1$ menghasilkan akurasi tertinggi sebesar 95,2% Namun, seiring bertambah nilai K , akurasi cenderung menurun dan stabil pada rentang $K=6$ hingga $K = 14$ dengan nilai sekitar 93,3%. Setelah $K > 14$, akurasi kembali mengalami penurunan hingga mencapai 89,9% $K=19$ dan $K = 20$. Oleh karena itu, meskipun $K = 1$ memberikan akurasi tertinggi, rentang K 6-14 dapat dipilih sebagai nilai yang lebih optimal karena memberikan performa yang stabil dan mengurangi risiko overfitting.



Gambar 4.6 Confusion Matrix KNN

37 Berdasarkan gambar *Confusion Matrix K-Nearest Neighbor (KNN)*, dapat
diketahui bahwa algoritma KNN memiliki kemampuan yang cukup baik dalam
22 85 melakukan klasifikasi penerima Bantuan Langsung Tunai (BLT). *Confusion matrix*
digunakan untuk mengevaluasi hasil prediksi model dengan membandingkan data
aktual (*true label*) dan data hasil prediksi (*predicted label*). Melalui *confusion*
matrix, dapat diketahui jumlah data yang diklasifikasikan dengan benar maupun
78 data yang mengalami kesalahan klasifikasi. Pada hasil pengujian tersebut terdapat
129 empat komponen utama, yaitu *True Positive (TP)*, *True Negative (TN)*, *False*
112 *Positive (FP)*, dan *False Negative (FN)* yang digunakan untuk mengukur performa
model klasifikasi.

33 Berdasarkan *confusion matrix*, diperoleh nilai *True Negative (TN)* sebesar
5, yang berarti terdapat 5 data yang sebenarnya termasuk kategori tidak layak
44 menerima BLT (kelas 0) dan berhasil diprediksi dengan benar oleh algoritma
19 sebagai tidak layak menerima bantuan. Nilai ini menunjukkan bahwa model masih
memiliki keterbatasan dalam mengenali masyarakat yang tidak memenuhi syarat
penerima bantuan.

33 Selanjutnya, diperoleh nilai *False Positive (FP)* sebesar 8, yang
menunjukkan bahwa terdapat 8 data yang sebenarnya tidak layak menerima
bantuan, namun diprediksi oleh sistem sebagai layak menerima bantuan. Kesalahan
92 klasifikasi ini menunjukkan bahwa model masih cenderung memberikan prediksi
layak pada beberapa data yang seharusnya tidak menerima bantuan. Kondisi
tersebut dapat menyebabkan bantuan menjadi kurang tepat sasaran. Pada bagian
19 *False Negative (FN)* diperoleh nilai sebesar 0, yang berarti tidak terdapat data yang
99 sebenarnya layak menerima bantuan tetapi diprediksi sebagai tidak layak. Hal ini

menunjukkan bahwa algoritma KNN mampu mengenali seluruh data masyarakat yang layak menerima BLT dengan sangat baik.

Sementara itu, nilai *True Positive (TP)* sebesar 92 menunjukkan bahwa terdapat 92 data yang benar-benar layak menerima bantuan dan berhasil diprediksi dengan benar oleh sistem. Nilai ini menunjukkan bahwa algoritma KNN memiliki kemampuan yang sangat tinggi dalam mendeteksi masyarakat yang layak menerima BLT. Secara keseluruhan, dari total 105 data pengujian, sebanyak 97 data berhasil diklasifikasikan dengan benar, sedangkan 8 data mengalami kesalahan klasifikasi. Hasil tersebut menunjukkan bahwa algoritma KNN memiliki tingkat akurasi yang cukup tinggi dalam proses klasifikasi penerima bantuan sosial. Namun demikian, berdasarkan confusion matrix terlihat bahwa algoritma KNN masih memiliki kelemahan dalam membedakan data masyarakat yang tidak layak menerima bantuan. Hal ini ditunjukkan oleh jumlah *False Positive* yang relatif lebih tinggi dibandingkan algoritma Naive Bayes. Oleh karena itu, hasil evaluasi ini dapat digunakan sebagai bahan perbandingan untuk menentukan algoritma yang paling efektif dan akurat dalam sistem pendukung keputusan penentuan penerima Bantuan Langsung Tunai (BLT).

Untuk memperoleh Classification Report, tidak perlu menghitung satu per satu 105 data testing secara manual. Yang dihitung manual adalah hasil evaluasi dari prediksi model terhadap 105 data testing berikut ini penyelesaian nya :

1. Menghitung Precision Kelas 0

$$Precision_0 = \frac{TP}{TP+FP} = 1,00$$

$$\frac{5}{5 + 0} = 1 \text{ maka } = 0$$

2. Menghitung Recall Kelas 0

$$Recall_0 = \frac{TP}{TP+FN} = 0,38$$

$$\frac{TP}{13} = 0,38 \approx 5$$

$$\text{Maka } FN = 13 - 5 = 8$$

3. Menghitung F1-Score Kelas 0

$$\begin{aligned} F1 &= 2 \times \frac{1,00 \times 0,38}{1,00 + 0,38} \\ &= 2 \times \frac{0,38}{1,00} = 0,551 \approx 0,56 \end{aligned}$$

4. Menghitung Precision Kelas 1

$$Precision_1 = \frac{92}{92 + 8} = \frac{92}{100} = 0,92$$

5. Menghitung Recall Kelas 1

$$\begin{aligned} Recall &= \frac{92}{92 + 0} \\ &= \frac{92}{92} \\ &= 1,00 \end{aligned}$$

6. Menghitung F1-Score Kelas 1

$$\begin{aligned} F1 &= 2 \times \frac{0,92 \times 1,00}{0,92 + 1,00} \\ &= 2 \times \frac{0,92}{1,92} \\ &= 0,9583 \approx 0,96 \end{aligned}$$

7. Menghitung Accuracy

$$\begin{aligned} Accuracy &= \frac{92 + 5}{105} \\ &= \frac{97}{105} \end{aligned}$$

$$= 0,9238$$

$$\approx 0,92$$

8. Menghitung Macro Average

$$\text{Precision Macro} \quad \frac{1,00 + 0,92}{2} = 0,96$$

$$\text{Recall Macro} \quad \frac{0,38 + 1,00}{2} = 0,69$$

$$\text{F1 Macro} \quad \frac{0,56 + 0,96}{2} = 0,76$$

9. Menghitung Weighted Average

Support kelas 0 = 13, kelas 1 = 92

Weighted Precision

$$\frac{(1,00 \times 13) + (0,92 \times 92)}{105}$$

$$= \frac{13 + 84,64}{105}$$

$$= 0,93$$

Weighted Recall

$$\frac{(0,38 \times 13) + (1,00 \times 92)}{105}$$

$$= \frac{4,94 + 92}{105}$$

$$= 0,92$$

Weighted F1-Score

$$\frac{(0,56 \times 13) + (0,96 \times 92)}{105}$$

$$= \frac{7,28 + 88,32}{105}$$

$$= 0,91$$

Berikut hasil dari perhitungan classification report Algoritma K-Nearest Neighbor (KNN) yang disajikan pada Tabel 4.10. Classification Report digunakan untuk mengevaluasi kinerja model klasifikasi berdasarkan beberapa metrik, yaitu precision, recall, f1-score, dan support pada masing-masing kelas. Evaluasi ini bertujuan untuk mengetahui seberapa baik model KNN dalam mengklasifikasikan data menjadi kelas Layak dan Tidak Layak penerima bantuan.

Tabel 4.11 Classification Report KNN

Classification Report:				
	precision	recall	f1-score	support
0	1.00	0.38	0.56	13
1	0.92	1.00	0.96	92
accuracy			0.92	105
macro avg	0.96	0.69	0.76	105
weighted avg	0.93	0.92	0.91	105

Berdasarkan hasil pengujian menggunakan Classification Report pada algoritma K-Nearest Neighbor (KNN), diperoleh nilai evaluasi performa model yang meliputi precision, recall, f1-score, dan accuracy. Hasil tersebut menunjukkan bahwa algoritma KNN memiliki kemampuan yang cukup baik dalam melakukan klasifikasi penerima Bantuan Langsung Tunai (BLT), meskipun masih terdapat beberapa kesalahan klasifikasi pada data tertentu. Pada kelas 0 yang merepresentasikan kategori tidak layak menerima bantuan, diperoleh nilai precision sebesar 1.00, recall sebesar 0.38, dan f1-score sebesar 0.56 dengan jumlah data (support) sebanyak 13 data. Nilai precision sebesar 1.00 menunjukkan bahwa seluruh data yang diprediksi sebagai tidak layak merupakan data yang benar-benar tidak layak menerima bantuan. Dengan kata lain, tidak terdapat kesalahan prediksi pada data yang dikategorikan sebagai tidak layak. Namun demikian, nilai

recall sebesar 0.38 menunjukkan bahwa kemampuan model dalam mengenali seluruh data tidak layak masih rendah. Hal ini berarti algoritma hanya mampu mendeteksi sekitar 38% dari total data yang sebenarnya termasuk kategori tidak layak menerima bantuan. Rendahnya nilai *recall* menyebabkan masih banyak data tidak layak yang salah diklasifikasikan ke dalam kategori layak menerima bantuan. Selain itu, nilai *f1-score* sebesar 0.56 menunjukkan bahwa keseimbangan antara *precision* dan *recall* pada kelas ini masih belum optimal.

Pada kelas 1 yang merepresentasikan kategori layak menerima bantuan, diperoleh nilai *precision* sebesar 0.92, *recall* sebesar 1.00, dan *f1-score* sebesar 0.96 dengan jumlah data sebanyak 92 data. Nilai *recall* sebesar 1.00 menunjukkan bahwa algoritma berhasil mengenali seluruh data yang benar-benar layak menerima bantuan tanpa ada data yang terlewat. Sementara itu, nilai *precision* sebesar 0.92 menunjukkan bahwa sebanyak 92% data yang diprediksi layak memang benar-benar layak menerima bantuan. Nilai *f1-score* sebesar 0.96 juga menunjukkan bahwa performa model pada kelas layak tergolong sangat baik. Secara keseluruhan, algoritma KNN menghasilkan nilai *accuracy* sebesar 0.92 atau 92%, yang menunjukkan bahwa model mampu melakukan klasifikasi dengan tingkat ketepatan yang cukup tinggi. Dari total 105 data pengujian, sebagian besar data berhasil diklasifikasikan dengan benar sesuai kategori masing-masing. Nilai *macro average* sebesar 0.76 menunjukkan rata-rata performa model pada setiap kelas tanpa mempertimbangkan jumlah data, sedangkan nilai *weighted average* sebesar 0.91 menunjukkan rata-rata performa keseluruhan model dengan mempertimbangkan proporsi jumlah data pada setiap kelas. Perbedaan nilai tersebut menunjukkan

bahwa model memiliki performa yang lebih baik pada kelas dengan jumlah data lebih banyak, yaitu kelas layak menerima bantuan.

4.4 Perbandingan Performa Algoritma

Berdasarkan hasil pengujian yang telah dilakukan, diperoleh perbandingan performa antar model berdasarkan nilai metrik evaluasi yang digunakan. Hasil tersebut menunjukkan perbedaan kemampuan masing-masing model dalam melakukan klasifikasi dapat dilihat pada tabel 4.11 dibawah sebagai berikut:

Tabel 4.12 Perbandingan Model

Model	Accuracy	Precision (Macro Avg)	Recall (Macro Avg)	F1-Score (Macro Avg)
Naive Bayes	0.96	0.92	0.92	0.92
KNN	0.92	0.96	0.69	0.76

Berdasarkan tabel perbandingan performa model di atas, dapat diketahui bahwa algoritma Naive Bayes dan K-Nearest Neighbor (KNN) sama-sama memiliki kemampuan yang cukup baik dalam melakukan klasifikasi penerima Bantuan Langsung Tunai (BLT). Namun, terdapat perbedaan performa pada beberapa parameter evaluasi yang digunakan. Algoritma Naive Bayes memperoleh nilai accuracy sebesar 0.96 atau 96%, lebih tinggi dibandingkan algoritma KNN yang memperoleh accuracy sebesar 0.92 atau 92%. Hasil tersebut menunjukkan bahwa Naive Bayes memiliki tingkat ketepatan klasifikasi yang lebih baik dalam menentukan data penerima bantuan.

Pada parameter precision (macro average), algoritma KNN memperoleh nilai lebih tinggi yaitu 0.96 dibandingkan Naive Bayes sebesar 0.92. Hal ini menunjukkan bahwa KNN memiliki tingkat ketepatan yang sangat baik dalam memprediksi data pada kelas tertentu. Namun, nilai recall (macro average) pada KNN hanya sebesar 0.69, sedangkan Naive Bayes memperoleh nilai recall sebesar

0.92. Nilai tersebut menunjukkan bahwa Naive Bayes lebih baik dalam mengenali keseluruhan data pada setiap kelas dibandingkan KNN.

4.5 Uji Statistik Paired T-Test

Uji statistik *paired t-test* dilakukan untuk mengetahui apakah terdapat perbedaan performa yang signifikan antara algoritma Naive Bayes dan K-Nearest Neighbor (KNN) dalam menentukan penerima Bantuan Langsung Tunai (BLT). Pengujian ini menggunakan data akurasi dari beberapa percobaan yang telah dilakukan pada kedua algoritma.

Tabel 4.13 Uji Statistik Paired T-Test

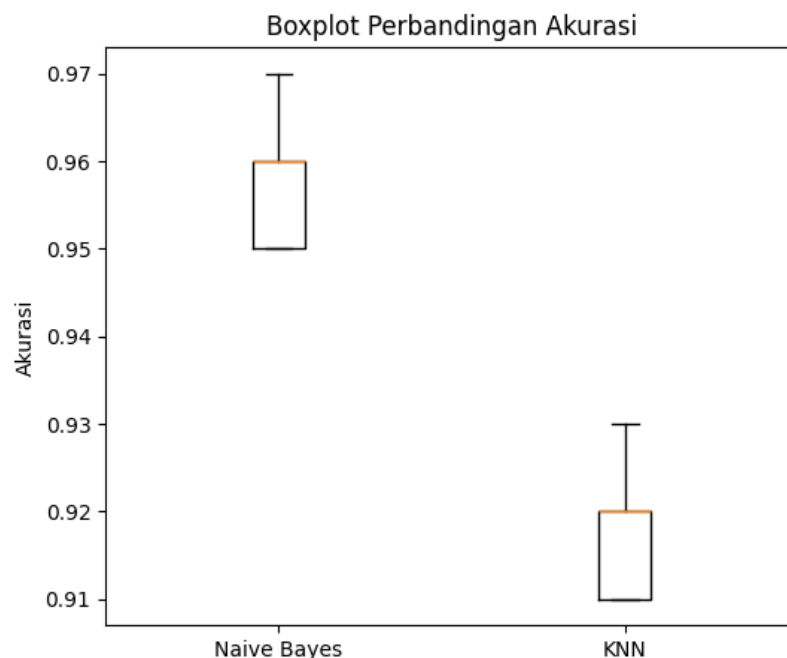
Metode	T- Statistic	P-value	Keterangan
Naive Bayes, KNN	7.5593	< 0.0001	Terdapat perbedaan signifikan

Berdasarkan hasil uji statistik *paired t-test* pada tabel di atas, diperoleh nilai T- Statistic sebesar 7.5593 dengan nilai P-value < 0.0001. Nilai P-value yang lebih kecil dari tingkat signifikansi $\alpha=0.05$ menunjukkan bahwa hipotesis nol (H_0) ditolak dan hipotesis alternatif (H_1) diterima.

Hasil tersebut menunjukkan bahwa terdapat perbedaan performa yang signifikan antara algoritma Naive Bayes dan K-Nearest Neighbor (KNN) dalam proses klasifikasi penerima Bantuan Langsung Tunai (BLT). Dengan kata lain, perbedaan hasil evaluasi kedua algoritma bukan terjadi secara kebetulan, melainkan benar-benar menunjukkan adanya perbedaan kemampuan klasifikasi antara kedua metode.

Berdasarkan hasil pengujian sebelumnya, algoritma Naive Bayes memiliki nilai *accuracy*, *recall*, dan *f1-score* yang lebih baik dibandingkan algoritma KNN. Oleh karena itu, hasil uji statistik ini semakin memperkuat bahwa algoritma Naive

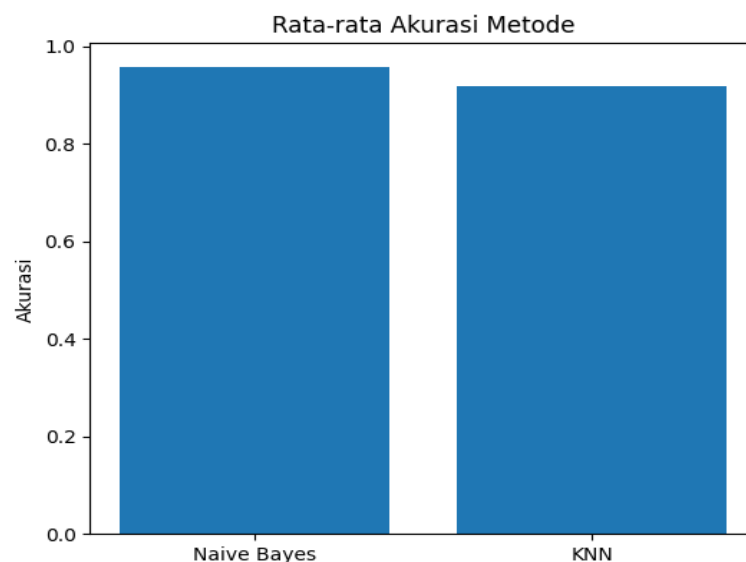
Bayes lebih optimal digunakan dalam sistem pendukung keputusan untuk menentukan penerima Bantuan Langsung Tunai (BLT). Dengan adanya pengujian paired t-test, penelitian ini tidak hanya membandingkan performa algoritma berdasarkan nilai evaluasi model saja, tetapi juga membuktikan secara statistik bahwa perbedaan performa kedua algoritma memiliki tingkat signifikansi yang tinggi. Oleh karena itu, algoritma Naive Bayes dapat direkomendasikan sebagai metode yang lebih efektif dan akurat dalam proses penentuan penerima bantuan sosial.



Gambar 4.7 Boxplot Perbandingan Akurasi

Berdasarkan gambar Boxplot Perbandingan Akurasi, dapat diketahui bahwa algoritma Naive Bayes dan K-Nearest Neighbor (KNN) memiliki perbedaan performa dalam proses klasifikasi penerima Bantuan Langsung Tunai (BLT). Boxplot digunakan untuk menampilkan distribusi data akurasi dari masing-masing algoritma berdasarkan beberapa percobaan yang telah dilakukan. Pada boxplot algoritma Naive Bayes, terlihat bahwa nilai akurasi berada pada rentang 0.95 hingga 0.97 dengan median berada di sekitar 0.96. Hal ini menunjukkan bahwa

algoritma Naive Bayes memiliki tingkat akurasi yang tinggi dan cenderung stabil pada setiap percobaan. Sebaran data yang relatif sempit juga menunjukkan bahwa hasil pengujian Naive Bayes konsisten dan tidak mengalami perubahan yang signifikan antar percobaan. Sementara itu, pada boxplot algoritma KNN, nilai akurasi berada pada rentang 0.91 hingga 0.93 dengan median sekitar 0.92. Hasil tersebut menunjukkan bahwa algoritma KNN juga memiliki performa yang cukup baik, namun nilai akurasinya masih berada di bawah algoritma Naive Bayes. Selain itu, distribusi akurasi KNN menunjukkan bahwa performa algoritma ini lebih rendah dibandingkan Naive Bayes dalam proses klasifikasi data penerima BLT. Jika dibandingkan secara keseluruhan, boxplot menunjukkan bahwa algoritma Naive Bayes memiliki nilai median akurasi yang lebih tinggi dibandingkan KNN. Hal ini menandakan bahwa Naive Bayes mampu memberikan hasil klasifikasi yang lebih baik dan lebih konsisten pada dataset yang digunakan dalam penelitian.



Gambar 4.8 Rata Rata Akurasi Metode

Berdasarkan gambar Rata-rata Akurasi Metode, dapat diketahui bahwa algoritma Naive Bayes dan K-Nearest Neighbor (KNN) memiliki perbedaan tingkat

akurasi dalam proses klasifikasi penerima Bantuan Langsung Tunai (BLT). Diagram batang tersebut menampilkan perbandingan rata-rata nilai akurasi dari kedua algoritma berdasarkan hasil pengujian yang telah dilakukan.

Pada grafik terlihat bahwa algoritma Naive Bayes memperoleh rata-rata akurasi sebesar sekitar 0.96 atau 96%, sedangkan algoritma KNN memperoleh rata-rata akurasi sekitar 0.92 atau 92%. Hasil tersebut menunjukkan bahwa algoritma Naive Bayes memiliki tingkat ketepatan klasifikasi yang lebih tinggi dibandingkan algoritma KNN. Tingginya nilai akurasi pada algoritma Naive Bayes menunjukkan bahwa metode ini mampu melakukan klasifikasi data penerima BLT dengan lebih baik dan lebih konsisten. Algoritma Naive Bayes dapat mengenali pola data masyarakat berdasarkan atribut yang digunakan, seperti penghasilan, jumlah tanggungan, kondisi rumah, dan status pekerjaan, sehingga menghasilkan prediksi yang lebih tepat. Sementara itu, algoritma KNN juga menunjukkan performa yang cukup baik dengan nilai akurasi di atas 90%. Namun, nilai rata-rata akurasi KNN masih berada di bawah Naive Bayes. Hal ini menunjukkan bahwa kemampuan KNN dalam melakukan klasifikasi masih kurang optimal dibandingkan Naive Bayes pada dataset penelitian yang digunakan

30

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan seluruh tahapan analisis, implementasi, dan pengujian komparatif yang telah dilaksanakan, penelitian ini berhasil mengintegrasikan algoritma Naive Bayes dan K-Nearest Neighbor (KNN) ke dalam sistem klasifikasi berbasis Google Colab dengan memanfaatkan 13 variabel sosial ekonomi warga Desa Sungai Ambat. Berdasarkan hasil pengujian terhadap 105 data uji (testing data), algoritma Naive Bayes menunjukkan performa yang sangat signifikan dan lebih unggul dengan meraih tingkat akurasi sebesar 96,19% atau mampu mengklasifikasikan 101 data secara tepat. Di sisi lain, algoritma K-Nearest Neighbor (KNN) menghasilkan tingkat akurasi yang lebih rendah, yaitu sebesar 92,38% dengan total 97 data terklasifikasi secara benar. Melalui analisis mendalam terhadap metrik evaluasi, algoritma Naive Bayes terbukti jauh lebih efektif, stabil, dan optimal dibandingkan dengan KNN dalam konteks penentuan kelayakan ini.

Model Naive Bayes mampu menunjukkan performa yang seimbang dengan perolehan nilai precision, recall, dan f1-score masing-masing sebesar 0,86 untuk kelas tidak layak dan 0,98 untuk kelas layak. Sebaliknya, algoritma KNN memiliki kelemahan mendasar pada nilai recall kelas tidak layak yang sangat rendah, yaitu hanya mencapai 0,38. Angka ini mengindikasikan tingginya bias kemunculan False Positive, di mana sistem berbasis KNN cenderung salah memprediksi warga yang sebenarnya tidak layak menjadi layak menerima bantuan. Secara keseluruhan, penerapan pendekatan data mining menggunakan algoritma Naive Bayes ini

99

52

241

52

9

50

140

156

terbukti mampu memberikan kontribusi nyata bagi tata kelola administrasi desa. Dengan memanfaatkan pemodelan komputasi ini, tingkat subjektivitas dalam pengambilan keputusan dapat ditekan seminimal mungkin. Selain itu, sistem ini juga berhasil meningkatkan aspek transparansi, objektivitas, serta efisiensi aparatur Desa Sungai Ambat dalam menyaring keluarga penerima manfaat secara tepat sasaran, sehingga potensi kesalahan distribusi bantuan sosial di masa mendatang dapat dihindari secara terstruktur dan ilmiah.

5.2 Saran

Demi pengembangan penelitian lanjutan di bidang *data mining* serta optimalisasi implementasi sistem pada masa yang akan datang, terdapat beberapa saran akademis dan praktis yang sangat direkomendasikan adalah sebagai berikut :

1. Pengembangan sistem terintegrasi mengingat batas penelitian ini belum mencakup pembuatan aplikasi, disarankan bagi peneliti berikutnya untuk mengemas model Naive Bayes terbaik ini ke dalam sebuah sistem berbasis web atau mobile yang memiliki antarmuka (*Use Interface*) interaktif bagi aparatur desa.
2. Penanganan Ketidakseimbangan Kelas (*Class Imbalance*): Mengingat adanya ketimpangan proporsi data yang cukup tinggi antara kelas Tidak Layak (453 data) dan kelas Layak (71 data) pada dataset awal, penelitian selanjutnya disarankan untuk menerapkan teknik penyeimbangan data seperti SMOTE (*Synthetic Minority Over-sampling Technique*) guna meningkatkan sensitivitas model terhadap kelas minoritas.
3. Penerapan *Feature Selection* dan *Hyperparameter Tuning* Disarankan untuk mengintegrasikan metode seleksi fitur otomatis (seperti *Information Gain* atau

Chi-Square) serta melakukan optimasi parameter (seperti variasi nilai K dan eksperimen metrik jarak selain *Euclidean Distance* pada KNN) guna mencari kombinasi pemodelan yang lebih presisi.

4. Pembaruan Data Secara Berkala (*Data Updating*) Pihak Pemerintah Desa Sungai Ambat diharapkan melakukan pembaruan data indikator kesejahteraan masyarakat secara periodik. Hal ini penting agar pola probabilistik yang dipelajari oleh algoritma tetap relevan dengan

DAFTAR PUSTAKA

- [1] R. Dewi and H. F. Andrianus, “Analisis pengaruh kebijakan bantuan langsung tunai (BLT) terhadap kemiskinan di indonesia periode 2005-2015,” *MENARA Ilmu* , vol. 15, no. 2, pp. 77–84, 2021.
- [2] C. Iswanaji, M. Z. Nafi’ Hasbi, F. Salekhah, and M. Amin, “Implementasi Analitical Networking Process (Anp) Distribusi Zakat Terhadap Pembangunan Ekonomi Masyarakat Berkelanjutan (Study Kasus Lembaga Baznas Kabupaten Jember Jawa Timur),” *J. Tabarru’ Islam. Bank. Financ.*, vol. 4, no. 1, pp. 195–208, 2021, doi: 10.25299/jtb.2021.vol4(1).6681.
- [3] M Adi Pratama Putra, Alfinda Adellia Dirli, Wilda Andriani, and Reza MauldyRaharja, “Menggali Efektivitas Bantuan Langsung Tunai dalam Membantu Masyarakat,” *Concept J. Soc. Humanit. Educ.*, vol. 2, no. 4, pp. 189–205, 2023, doi: 10.55606/concept.v2i4.781.
- [4] I. Arfanda, W. Ramdhan, and R. A. Yusda, “Naive Bayes Dalam Menentukan Penerima Bantuan Langsung Tunai,” *Digit. Transform. Technol.*, vol. 1, no. 1, pp. 9–16, 2021, doi: 10.47709/digitech.v1i1.1091.
- [5] P. Ramadani, R. Fadillah, Q. Adawiyah, and B. R. Al Ghazali, “Perbandingan Algoritma Naïve Bayes, C4.5, dan K-Nearest Neighbor untuk Klasifikasi Kelayakan Program Keluarga Harapan,” *J. Media Inform.*, vol. 6, no. 2, pp. 775–782, 2024.
- [6] H. Putri, A. I. Purnamasari, A. R. Dikananda, O. Nurdiawan, and S. Anwar, “Penerima Manfaat Bantuan Non Tunai Kartu Keluarga Sejahtera

- Menggunakan Metode NAÏVE BAYES dan KNN,” *Build. Informatics, Technol. Sci.*, vol. 3, no. 3, pp. 331–337, 2021, doi: 10.47065/bits.v3i3.1093.
- [7] K. Masyarakat, D. Bone, K. Bajeng, and K. Gowa, “Kemiskinan Masyarakat Petani di Desa Bone, Kec. Bajeng, Kab. Gowa,” vol. 1, no. 3, pp. 102–121, 2023.
- [8] R. Indra Yudha, L. Hakim, and M. Sabandi, “Studi Kasus Perlindungan Sosial Melalui PKH, BLT, dan BSU: Upaya Mengurangi Ketergantungan Pinjaman pada Masyarakat Rentan di Kota Jambi,” *SJEE (Scientific Journals Econ. Educ.*, vol. 9, no. 2, p. 243, 2025, doi: 10.33087/sjee.v9i2.247.
- [9] Tessa and K. Indarto, “Implementasi Program Bantuan Langsung Tunai (BLT) Di Kampung Geleo Baru Kecamatan Barong Tongkok Kabupaten Kutai Barat,” *J. Inov. dan Kreat.*, vol. 4, no. 2, pp. 29–42, 2024, doi: 10.30656/jika.v4i2.9395.
- [10] F. A. Prasetyo, N. Dyah Gianawati, K. Wulandari, W. Mayangsari, and S. Purwantini, “Jember Problems and Recommendations for Social Assistance: Case Study of the Fuel Compensation Social Assistance Program in Jember District,” *Soc. Work J.*, vol. 13, no. 1, pp. 52–59, 2023, [Online]. Available: <https://doi.org/10.45814/share.v13i1.45720>
- [11] R. Limia Budiarti, L. Puad, and R. -, “Penerapan Metode K-Means Untuk Proses Penentuan Bantuan Langsung Tunai (Blt) Pada Desa Mekar Sari Kecamatan Kumpeh Kabupaten Muaro Jambi,” *J. Akad.*, vol. 16, no. 2, pp. 57–63, 2024, doi: 10.53564/akademika.v16i2.1190.
- [12] H. Zia Rahimi, S. Defit, and J. Veri, “Implementasi Decision Tree Dalam

- Pengambilan Keputusan Untuk Pemberian Beasiswa,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 3, pp. 4175–4180, 2025, doi: 10.36040/jati.v9i3.13580.
- [13] M. Fajar, S. Adam, B. Putra, S. I. Puteri, A. Fajrissiddiq, and L. Sani, “Eksplorasi dan Analisis Data Mining untuk Prediksi Pola Konsumen Menggunakan Teknik Klasifikasi dan Clustering,” 2025.
- [14] N. Riyanah and F. Fatmawati, “Penerapan Algoritma Naive Bayes Untuk Klasifikasi Penerima Bantuan Surat Keterangan Tidak Mampu,” *JTIM J. Teknol. Inf. dan Multimed.*, vol. 2, no. 4, pp. 206–213, 2021, doi: 10.35746/jtim.v2i4.117.
- [15] Chandra Wisnu Nugroho, Sigrid Tamriesfatno, and Sulkifli Sulkifli, “Implementasi Algoritma Naive Bayes untuk Sistem Pendukung Keputusan Penerima Bantuan Pendidikan,” *J. Publ. Ilmu Komput. dan Multimed.*, vol. 4, no. 1, pp. 1–10, 2025, doi: 10.55606/jupikom.v4i1.4276.
- [16] M. Muqorobin, K. Kusrini, and E. T. Luthfi, “Optimasi Metode Naive Bayes Dengan Feature Selection Information Gain Untuk Prediksi Keterlambatan Pembayaran Spp Sekolah,” *J. Ilm. SINUS*, vol. 17, no. 1, p. 1, 2019, doi: 10.30646/sinus.v17i1.378.
- [17] J. H. Saviola and N. D. Hendrawan, “IMPLEMENTASI KLASIFIKASI KUALITAS SUSU MENGGUNAKAN ALGORITMA DECISION TREE , K-NEAREST NEIGHBORS DAN NAIVE BAYES,” vol. 9, no. 5, pp. 8953–8960, 2025.
- [18] D. Ardiyansyah, N. Oktafiani, and U. T. Madura, “PERBANDINGAN

METODE PENGUKURAN JARAK PADA K-NEAREST,” vol. 1, no. 2, pp. 116–122, 2024.

[19] E. E. Lorencislubis, A. Priansyah, R. W. N. Yolanda, and T. Meutia, “PENERAPAN DATA MINING UNTUK MEMPREDIKSI HARGA BAHAN PANGAN DI INDONESIA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR,” vol. 4, pp. 223–235, 2023.

[20] S. Octarina, F. M. Puspita, E. Yuliza, and G. Meeting, “Jurnal Pepadu Jurnal Pepadu,” vol. 6, no. 1, pp. 56–66, 2025.

[21] P. Website, P. Sebagai, M. Pembelajaran, M. Hasil, and K. Python, “Sainmatika unja |,” vol. 6, no. 1, pp. 1–10, 2025.

[22] B. Marga, “Journal of University Community Engagement,” vol. 1, no. 1, pp. 61–69, 2026.

[23] T. Carneiro, R. Victor, M. Da, T. Nepomuceno, G. Bian, and V. H. C. D. E. Albuquerque, “Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications,” pp. 61677–61685, 2018, doi: 10.1109/ACCESS.2018.2874767.

[24] K. Akuntansi and S. I. Cirebon, “ALGORITMA REGRESI LINEAR MENGGUNAKAN GOOGLE COLLAB,” vol. 8, no. 1, pp. 827–837, 2024.

Lampiran Dokumentasi



DAFTAR KELUARAN PERUSAHAAN KAWASAN PERUSAHAAN RUMAH TANGGA (KPR)									
No. Urut	Nama Perusahaan	Alamat	Telepon	Fax	Website	Produk	Volume	Unit	Unit
1	PT. BUKIT BARU	Jl. Raya Kertajaya No. 100	0812 3456 7890		www.bukitbaru.co.id	Kayu	1000	m3	1000
2	PT. BUKIT BARU	Jl. Raya Kertajaya No. 100	0812 3456 7890		www.bukitbaru.co.id	Kayu	1000	m3	1000
3	PT. BUKIT BARU	Jl. Raya Kertajaya No. 100	0812 3456 7890		www.bukitbaru.co.id	Kayu	1000	m3	1000
4	PT. BUKIT BARU	Jl. Raya Kertajaya No. 100	0812 3456 7890		www.bukitbaru.co.id	Kayu	1000	m3	1000
5	PT. BUKIT BARU	Jl. Raya Kertajaya No. 100	0812 3456 7890		www.bukitbaru.co.id	Kayu	1000	m3	1000
6	PT. BUKIT BARU	Jl. Raya Kertajaya No. 100	0812 3456 7890		www.bukitbaru.co.id	Kayu	1000	m3	1000
7	PT. BUKIT BARU	Jl. Raya Kertajaya No. 100	0812 3456 7890		www.bukitbaru.co.id	Kayu	1000	m3	1000
8	PT. BUKIT BARU	Jl. Raya Kertajaya No. 100	0812 3456 7890		www.bukitbaru.co.id	Kayu	1000	m3	1000
9	PT. BUKIT BARU	Jl. Raya Kertajaya No. 100	0812 3456 7890		www.bukitbaru.co.id	Kayu	1000	m3	1000
10	PT. BUKIT BARU	Jl. Raya Kertajaya No. 100	0812 3456 7890		www.bukitbaru.co.id	Kayu	1000	m3	1000

DAFTAR KELUARAN PERUSAHAAN KAWASAN PERUSAHAAN RUMAH TANGGA (KPR)									
No. Urut	Nama Perusahaan	Alamat	Telepon	Fax	Website	Produk	Volume	Unit	Unit
1	PT. BUKIT BARU	Jl. Raya Kertajaya No. 100	0812 3456 7890		www.bukitbaru.co.id	Kayu	1000	m3	1000
2	PT. BUKIT BARU	Jl. Raya Kertajaya No. 100	0812 3456 7890		www.bukitbaru.co.id	Kayu	1000	m3	1000
3	PT. BUKIT BARU	Jl. Raya Kertajaya No. 100	0812 3456 7890		www.bukitbaru.co.id	Kayu	1000	m3	1000
4	PT. BUKIT BARU	Jl. Raya Kertajaya No. 100	0812 3456 7890		www.bukitbaru.co.id	Kayu	1000	m3	1000
5	PT. BUKIT BARU	Jl. Raya Kertajaya No. 100	0812 3456 7890		www.bukitbaru.co.id	Kayu	1000	m3	1000
6	PT. BUKIT BARU	Jl. Raya Kertajaya No. 100	0812 3456 7890		www.bukitbaru.co.id	Kayu	1000	m3	1000
7	PT. BUKIT BARU	Jl. Raya Kertajaya No. 100	0812 3456 7890		www.bukitbaru.co.id	Kayu	1000	m3	1000
8	PT. BUKIT BARU	Jl. Raya Kertajaya No. 100	0812 3456 7890		www.bukitbaru.co.id	Kayu	1000	m3	1000
9	PT. BUKIT BARU	Jl. Raya Kertajaya No. 100	0812 3456 7890		www.bukitbaru.co.id	Kayu	1000	m3	1000
10	PT. BUKIT BARU	Jl. Raya Kertajaya No. 100	0812 3456 7890		www.bukitbaru.co.id	Kayu	1000	m3	1000

[illegible]