

Turnitin LLCC

2213020099_NISAA HUSNIA YAHYA

TI-20

Document Details

Submission ID

trn:oid::2945:395584519

Submission Date

Jun 25, 2026, 2:14 PM GMT+7

Download Date

Jun 25, 2026, 2:17 PM GMT+7

File Name

2213020099_NISAA HUSNIA YAHYA.docx

File Size

2.4 MB

85 Pages

14,855 Words

101,486 Characters

29% Overall Similarity

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

Filtered from the Report

- Bibliography

Top Sources

- 14%  Internet sources
- 15%  Publications
- 26%  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

- 14% Internet sources
- 15% Publications
- 26% 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	
	ejurnal.seminar-id.com	<1%
2	Student papers	
	Universitas Putera Batam on 2026-02-03	<1%
3	Student papers	
	Universitas Pamulang on 2026-06-22	<1%
4	Student papers	
	Udayana University on 2026-01-08	<1%
5	Publication	
	Muhammad Aliq Aulia, Kresno Ario Tri Wibowo, Irfan Nugraha, Ilham Wahyu Anal...	<1%
6	Student papers	
	Universitas Islam Riau on 2026-06-10	<1%
7	Student papers	
	Telkom University on 2026-06-18	<1%
8	Student papers	
	UPN Veteran Yogyakarta on 2026-03-02	<1%
9	Student papers	
	Universitas Khairun on 2026-04-22	<1%
10	Student papers	
	Institut Teknologi Sumatera on 2026-03-10	<1%
11	Internet	
	text-id.123dok.com	<1%

12	Student papers	Universitas Khairun on 2026-05-20	<1%
13	Internet	docplayer.info	<1%
14	Publication	Harliana Harliana, Tito Prabowo, Ady Alzhava Nuary. "Perbandingan Logistic Reg...	<1%
15	Internet	jurnal.polibatam.ac.id	<1%
16	Student papers	Institut Teknologi Sumatera on 2025-09-22	<1%
17	Internet	etheses.uin-malang.ac.id	<1%
18	Internet	kc.umh.ac.id	<1%
19	Student papers	Universitas Islam Indonesia on 2026-06-15	<1%
20	Student papers	Universitas Bengkulu on 2025-10-13	<1%
21	Internet	e-jurnal.pnl.ac.id	<1%
22	Student papers	UPN Veteran Yogyakarta on 2026-05-05	<1%
23	Student papers	Universitas Islam Negeri Sumatera Utara on 2026-06-18	<1%
24	Student papers	Sriwijaya University on 2026-05-04	<1%
25	Student papers	Universitas Muhammadiyah Palembang on 2026-04-23	<1%

26	Internet	repository.unissula.ac.id	<1%
27	Student papers	Universitas Putera Batam on 2026-02-03	<1%
28	Student papers	Universitas Putera Batam on 2026-02-03	<1%
29	Internet	jurnal.itg.ac.id	<1%
30	Student papers	Universitas Muhammadiyah Makassar on 2026-06-25	<1%
31	Internet	repository.widyatama.ac.id	<1%
32	Student papers	Universitas Khairun on 2026-05-11	<1%
33	Student papers	Fakultas Teknik on 2025-08-07	<1%
34	Student papers	Sriwijaya University on 2023-09-19	<1%
35	Internet	es.scribd.com	<1%
36	Student papers	Telkom University on 2026-04-27	<1%
37	Internet	repository.ar-raniry.ac.id	<1%
38	Student papers	Tarumanagara University on 2026-06-21	<1%
39	Student papers	Universitas Darma Persada on 2026-02-08	<1%

40	Student papers	Universitas Muhammadiyah Jember on 2026-03-31	<1%
41	Student papers	Universitas Putera Batam on 2026-02-02	<1%
42	Internet	www.bi.go.id	<1%
43	Student papers	Universitas Pancasila on 2026-02-21	<1%
44	Publication	Yoga - Yapiza, Taufik - Taufik, Tri - Susilowati. "CLUSTERING TINGKAT KERAWANA...	<1%
45	Internet	journal.ilmudata.co.id	<1%
46	Internet	repository.uin-suska.ac.id	<1%
47	Internet	research-repository.griffith.edu.au	<1%
48	Publication	Bagas Diki Saputra, Alfi Zuhriya Khoirunnisaa. "Klasifikasi Penggunaan Listrik Ru...	<1%
49	Student papers	Universitas Islam Riau on 2026-06-11	<1%
50	Student papers	andalas on 2024-08-12	<1%
51	Internet	digilib.unila.ac.id	<1%
52	Publication	Fadhlan Ihsan Lubis, Muhammad Iqbal. "Analisis Waktu Optimal Unggah Instagr...	<1%
53	Student papers	Politeknik Negeri Jember on 2026-06-22	<1%

54	Publication	Teguh Muryanto, Aji Nurrohman, Rachmat Setiabudi, Wibisono Wibisono, Berliya...	<1%
55	Internet	www.scribd.com	<1%
56	Student papers	Konsorsium PTS Indonesia - Small Campus II on 2025-07-16	<1%
57	Student papers	UIN Sunan Ampel Surabaya on 2025-11-29	<1%
58	Student papers	Universitas Bengkulu on 2025-10-06	<1%
59	Student papers	Universitas Islam Riau on 2025-09-10	<1%
60	Student papers	Universitas Islam Riau on 2026-06-15	<1%
61	Student papers	Universitas Jember on 2026-05-26	<1%
62	Publication	Eka Sri Rahayu, Bobi Agustian. "Klasifikasi dan analisis faktor yang mempengaru...	<1%
63	Publication	Heru Teguh Santoso Teguh. "Implementasi Data Mining Clustering Dalam Meng...	<1%
64	Student papers	UPN Veteran Yogyakarta on 2026-05-04	<1%
65	Student papers	Universitas Muhammadiyah Sukabumi on 2025-07-12	<1%
66	Internet	sistemasi.ftik.unisi.ac.id.sistemasi.org	<1%
67	Publication	Anis Athifah, Herlawati Herlawati, Fata Nidaul Khasanah. "Evaluasi Kinerja Cluste...	<1%

68	Student papers	UIN Raden Intan Lampung on 2026-02-12	<1%
69	Student papers	Universitas Islam Riau on 2025-11-27	<1%
70	Student papers	Universitas Tarumanagara on 2025-12-17	<1%
71	Publication	Yasir Hasan. "PENGUKURAN SILHOUETTE SCORE DAN DAVIES-BOULDIN INDEX PA...	<1%
72	Internet	123dok.com	<1%
73	Student papers	Fakultas Teknik on 2025-07-23	<1%
74	Publication	Fransiscus Aditya Wibowo, Mardiani Mardiani. "PREDIKSI PERSEDIAAN DAN PERM...	<1%
75	Student papers	Politeknik Statistika STIS on 2025-07-05	<1%
76	Publication	Sonianto Son, Fatoni Fatoni, Zaki Afwanda, Ahmad Fauzan. "PENGEMBANGAN SIS...	<1%
77	Student papers	UPN Veteran Yogyakarta on 2026-04-02	<1%
78	Internet	lib.unnes.ac.id	<1%
79	Internet	repository.ub.ac.id	<1%
80	Internet	repository.umsu.ac.id	<1%
81	Publication	Cheolhwan Lee, Jongkook Kim, Namjoong Kim, Seil Ki, Jeonggyu Seo, Changhyup ...	<1%

82	Publication	Hasan Ibrohim, Harminto Mulyo, Gentur Wahyu Nyipto Wibowo. "Pemodelan Hyb...	<1%
83	Publication	Ichsan Mursidah, Remi Sanjaya, Bambang Yulianto, Dhian Sweetania, Puji Sularsi...	<1%
84	Publication	Larisa Oriana, Anisa Dwi Sanggar Wati, Anggi Putri Ramahdani, Natasya Nurul Sa...	<1%
85	Student papers	UPN Veteran Jakarta on 2025-12-19	<1%
86	Student papers	UPN Veteran Yogyakarta on 2026-06-02	<1%
87	Student papers	Universitas Budi Luhur on 2026-02-21	<1%
88	Student papers	Universitas Islam Riau on 2025-11-12	<1%
89	Student papers	Universitas Muhammadiyah Purwokerto on 2025-12-12	<1%
90	Student papers	Universitas Tarumanagara on 2025-12-23	<1%
91	Internet	doku.pub	<1%
92	Internet	repository.unpkediri.ac.id	<1%
93	Publication	Ihsan Pratama Putra, Avram Fadhillah. "Perbandingan Metode K-Means dan Hier...	<1%
94	Student papers	Institut Agama Islam Al-Zaytun Indonesia on 2025-08-31	<1%
95	Student papers	Institut Bisnis dan Teknologi Indonesia (INSTIKI) on 2025-06-08	<1%

96	Student papers	Konsorsium PTS Batch 5 on 2026-06-25	<1%
97	Student papers	Universitas Islam Bandung on 2025-08-08	<1%
98	Internet	ejurnal.undipa.ac.id	<1%
99	Internet	jurnal.fikom.umi.ac.id	<1%
100	Internet	repository.uninus.ac.id	<1%
101	Student papers	Institut Teknologi Sumatera on 2025-09-22	<1%
102	Publication	M. Khaerul Ihsan, Muhammad Maulana, Tanwir Tanwir, Abi Mas'ud, Naufal Hanif,...	<1%
103	Student papers	UPN Veteran Yogyakarta on 2026-01-06	<1%
104	Student papers	Universitas Brawijaya on 2017-10-05	<1%
105	Student papers	Universitas Kristen Duta Wacana on 2026-04-14	<1%
106	Student papers	Universitas Kristen Duta Wacana on 2026-05-07	<1%
107	Student papers	Universitas Sanata Dharma on 2026-06-10	<1%
108	Internet	jurnal.ilmubersama.com	<1%
109	Publication	Deny Adi Setyawan, Agustina Purwatiningsih, Febrian Sulistyo Budi. "Analisis Pol...	<1%

110	Student papers	Dominican College on 2014-05-02	<1%
111	Publication	Muhammad Syani, Tundo, Sugiyono, Tri Wahyudi. "Klasterisasi Penggunaan Ban...	<1%
112	Publication	Thomas Brian, Immanuel Freddy Augustino, Parman Parman, Muhamad Sukarn...	<1%
113	Student papers	Universitas Muhammadiyah Purwokerto on 2025-11-13	<1%
114	Publication	Yurika Caesarita, Nanang Wahyudi, Agung Kurniawan Faisol. "Prediksi Kelulusan ...	<1%
115	Publication	Zian Asti Dwiyanti, Cahyo Prianto. "Prediksi Cuaca Kota Jakarta Menggunakan Me...	<1%
116	Internet	eprints.unm.ac.id	<1%
117	Internet	repository.dinamika.ac.id	<1%
118	Internet	repository.itpln.ac.id	<1%
119	Internet	repository.unj.ac.id	<1%
120	Publication	Agustin Agustin, Junadhi Junadhi, Fransiskus Zoromi, Parlindungan Kudadiri. "An...	<1%
121	Publication	Aprilian Gevindo, Syafri Arlis. "Smoker Detection Using the You Only Look Once (Y...	<1%
122	Publication	Muhammad Afif Hidayat, Idham Fatirrohman Syufi. "Implementasi Sistem E-Com...	<1%
123	Publication	Rizki Akbar Pratama, Barry Ceasar Octariadi, Syarifah Putri Agustini Alkadri. "Pen...	<1%

124	Student papers	UPN Veteran Yogyakarta on 2026-01-13	<1%
125	Student papers	Universitas Diponegoro on 2024-12-18	<1%
126	Student papers	Universitas Pamulang on 2026-05-04	<1%
127	Student papers	Universitas Pancasila on 2026-03-04	<1%
128	Internet	ejournal.warunayama.org	<1%
129	Internet	informatika.unpkediri.ac.id	<1%
130	Internet	mdpi-res.com	<1%
131	Student papers	Abdullah Gul University on 2025-11-20	<1%
132	Publication	Diana Sava Salsabila, Arimbi Puspitasari, Dwi Rolliawati. "Klasifikasi Jenis Sampah...	<1%
133	Student papers	Forum Perpustakaan Perguruan Tinggi Indonesia Jawa Timur II on 2025-08-20	<1%
134	Publication	Jelly Helia Putri, Sunarsan Sitohang. "Klasifikasi Kinerja Pegawai Menggunakan Al...	<1%
135	Student papers	Konsorsium Perguruan Tinggi Swasta Indonesia II on 2021-07-13	<1%
136	Publication	Mohamad Sugeng Pangestu, Maulida Ayu Fitriani. "Perbandingan Perhitungan Ja...	<1%
137	Publication	Muhamad Faza Almaliki, Muhammad Chaerur Rijal, Isnawaty Isnawaty, Maudhy S...	<1%

138	Student papers	Politeknik Negeri Jakarta on 2022-04-20	<1%
139	Publication	Saikin Saikin, Mohammad Taufan Asri Zaen, Sofiansyah Fadli, Hairul Fahmi. "Pene...	<1%
140	Student papers	Sriwijaya University on 2021-06-29	<1%
141	Student papers	UPN Veteran Yogyakarta on 2025-10-29	<1%
142	Student papers	Universitas Brawijaya on 2018-01-04	<1%
143	Student papers	Universitas Diponegoro on 2017-02-08	<1%
144	Student papers	Universitas Muhammadiyah Buton on 2026-06-05	<1%
145	Student papers	Universitas Negeri Semarang on 2026-04-13	<1%
146	Student papers	Universitas Pendidikan Ganesha on 2025-07-24	<1%
147	Student papers	Universitas Sanata Dharma on 2026-06-10	<1%
148	Student papers	Universitas Terbuka on 2018-02-02	<1%
149	Internet	journal.fortei7.org	<1%
150	Internet	jurnal.ciptamediaharmoni.id	<1%
151	Internet	koma.or.id	<1%

152	Internet	nurchang21.blogspot.com	<1%
153	Internet	repository.unilak.ac.id	<1%
154	Publication	Arif Rahman, Andysah Putera Utama Siahaan, Khairul Khairul, Muhammad Iqbal, ...	<1%
155	Publication	Febriansyah Febriansyah, Nurmaleni Nurmaleni. "Implementasi Algoritma Rando...	<1%
156	Publication	Hayatul Ilmi, Khairuman Khairuman, Rossiana Br Ginting. "Perancangan Sistem A...	<1%
157	Publication	Laila Salsabilla Hanifa, Herlawati Herlawati, Rakhmat Purnomo. "Perbandingan K...	<1%
158	Publication	Nurul Azwanti, Yulia Yulia. "Perbandingan Algoritma Clustering K-Means dan DBS...	<1%
159	Student papers	Perpustakaan on 2025-10-08	<1%
160	Student papers	Politeknik Ilmu Pelayaran Semarang on 2026-05-20	<1%
161	Publication	Rinakit Estu Waluyo, Muis Muhtadi, Nofri Ramadan, Muhammad Zulfiqar Rafi. "A...	<1%
162	Student papers	Telkom University on 2026-01-21	<1%
163	Student papers	Telkom University on 2026-03-27	<1%
164	Student papers	UM Surabaya on 2025-07-08	<1%
165	Student papers	UPN Veteran Yogyakarta on 2026-04-22	<1%

166	Student papers	Universitas Brawijaya on 2016-12-22	<1%
167	Student papers	Universitas Dinamika on 2026-06-11	<1%
168	Student papers	Universitas Islam Riau on 2026-01-07	<1%
169	Student papers	Universitas Islam Riau on 2026-03-05	<1%
170	Student papers	Universitas Khairun on 2026-05-04	<1%
171	Student papers	Universitas Khairun on 2026-05-25	<1%
172	Student papers	Universitas Merdeka Malang on 2025-01-06	<1%
173	Student papers	Universitas Pancasila on 2026-01-29	<1%
174	Student papers	Universitas Pancasila on 2026-05-18	<1%
175	Student papers	Universitas Sumatera Utara on 2024-11-15	<1%
176	Student papers	Universitas Tarumanagara on 2025-12-22	<1%
177	Internet	digilib.uin-suka.ac.id	<1%
178	Internet	ejournal.unuja.ac.id	<1%
179	Internet	eprints.dinus.ac.id	<1%

180	Internet	eprints.umsida.ac.id	<1%
181	Internet	journalcenter.org	<1%
182	Internet	kirim.ai	<1%
183	Internet	ojsiibn1.indobarunasional.ac.id	<1%
184	Internet	publikasiilmiah.umsu.ac.id	<1%
185	Internet	repositori.uin-alaudidin.ac.id	<1%
186	Internet	repository.ipb.ac.id:8080	<1%
187	Student papers	Academic Library Consortium on 2024-11-26	<1%
188	Publication	Agim Sahrija Asha Yuda, Muhammad Desfriyan Arif Rosady, Nabil Ibrahim Faisal, ...	<1%
189	Student papers	Boston University on 2026-05-13	<1%
190	Publication	Cokorde Dhoio Pranamyaditia. "RISIKO KESELAMATAN DAN KESEHATAN KERJA PAD...	<1%
191	Publication	I Gede Haga Olat Tyamarta -, I Nyoman Sukajaya -, Putu Kartika Dewi -. "DIAGNO...	<1%
192	Student papers	STT PLN on 2023-07-27	<1%
193	Student papers	Sekolah Tinggi Seminari Katolik on 2025-12-13	<1%

194	Publication	Shifa Hairul Nisah, Nova Dwi Maelani, Niken Ayu Dara, Sefrika Entas. "Analisis Pol...	<1%
195	Student papers	Tarumanagara University on 2025-04-24	<1%
196	Student papers	UIN Batusangkar on 2025-12-02	<1%
197	Student papers	UNIVERSITAS BUDI LUHUR on 2026-02-10	<1%
198	Student papers	UPN Veteran Yogyakarta on 2025-10-28	<1%
199	Student papers	UPN Veteran Yogyakarta on 2026-05-04	<1%
200	Student papers	Universitas Andalas on 2025-05-08	<1%
201	Student papers	Universitas Brawijaya on 2017-07-13	<1%
202	Student papers	Universitas Pamulang on 2026-05-21	<1%
203	Student papers	Universitas Pendidikan Ganesha on 2025-08-02	<1%
204	Student papers	Universitas Sangga Buana YPKP on 2026-02-21	<1%
205	Student papers	Universitas Sangga Buana YPKP on 2026-02-23	<1%
206	Student papers	Universitas Sumatera Utara on 2026-04-15	<1%
207	Student papers	Universitas Tarumanagara on 2025-12-21	<1%

208	Publication	Yohanes Soge Tukan, Bernadete Deta, Alvian Nara Weking. "Perancangan Aplikas...	<1%
209	Internet	adoc.pub	<1%
210	Internet	eprints.umg.ac.id	<1%
211	Internet	eprints.upj.ac.id	<1%
212	Internet	garuda.ristekdikti.go.id	<1%
213	Internet	jurnalmahasiswa.unipasby.ac.id	<1%
214	Internet	repository.bunghatta.ac.id	<1%
215	Internet	wapikweb.org	<1%
216	Publication	Adinda Riska Safitri, Andi Arniaty Arsyad. "Pemodelan Emosi Pengguna Berbasis ...	<1%
217	Publication	Ananda Rizki Dani, Irma Handayani. "Classification of Yogyakarta Batik Motifs Usi...	<1%
218	Publication	Andrean Maulana Artiarno, Pratomo Setiaji, Fajar Nugraha. "K-Means Clustering ...	<1%
219	Publication	Chaerul Hidayat, Agung Nugroho, Asep Suprianto. "Prediksi Risiko Stroke Berdas...	<1%
220	Publication	Christian Bautista, Daniel Udjulawa. "Penerapan Random Forest Dan Borderline S...	<1%
221	Publication	Danda Bagaskoro, Pita Penengah, Muh. Junaidi, Muhamad Ilham Afandi. "Optima...	<1%

222	Publication	Dea Karlina Pali, Audy Aldrin Kenap, Efraim Ronald S. Moningkey. "Website Klasifi...	<1%
223	Publication	Dwito Julian Tambunan, Agustina Simbolon, Sri Ulina Br Marmata, Sardo Parding...	<1%
224	Student papers	FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM on 2026-01-06	<1%
225	Publication	Fadila Ananda Kartika Hidayat, Neny Sulistianingsih, Rifqi Hammad. "Perbanding...	<1%
226	Student papers	Fakultas Teknik on 2025-08-21	<1%
227	Publication	Fanny Miktami Lady Taxi Waruwu, Audita Nuvriasari. "Strategies for Handling Pr...	<1%
228	Publication	Gita Rohma Utami Asyafiiyah, Artika Widyastuti, Friska Andriani. "Analisis Wilaya...	<1%
229	Student papers	IAIN Bukit Tinggi on 2025-07-31	<1%
230	Student papers	Iain Palopo on 2020-01-31	<1%
231	Publication	Jordan Johan Josafat Simanjuntak, Magdalena A. Ineke Pakereng. "PERBANDINGA...	<1%
232	Publication	Luthfia Nurma Hapsari, Ilham Fannani, Yenny Rahmawati, Ahmad Muhariya. "Ex...	<1%
233	Publication	Patria Gita Laksamana, Willy Prihartono, Fathurrohman .. "IMPLEMENTASI NAÏVE ...	<1%
234	Publication	Riska Aini Putri, Muhammad Abdul Hadi, Fuad Maulana, Dzaky Akram, Nathasa S...	<1%
235	Student papers	School of Business and Management ITB on 2024-12-16	<1%

236	Publication	Sujatmiko Ginting. "Pengaruh Perilaku Konsumen terhadap Penggunaan Media ..."	<1%
237	Student papers	Sultan Agung Islamic University on 2026-02-16	<1%
238	Student papers	Tarumanagara University on 2026-06-25	<1%
239	Student papers	UM Surabaya on 2025-09-12	<1%
240	Student papers	UPN Veteran Yogyakarta on 2026-02-25	<1%
241	Student papers	UPN Veteran Yogyakarta on 2026-06-02	<1%
242	Student papers	Universitas Bengkulu on 2024-10-08	<1%
243	Student papers	Universitas Bengkulu on 2026-06-25	<1%
244	Student papers	Universitas Brawijaya on 2016-11-03	<1%
245	Student papers	Universitas Brawijaya on 2017-07-18	<1%
246	Student papers	Universitas Brawijaya on 2019-07-08	<1%
247	Student papers	Universitas Dinamika on 2026-06-11	<1%
248	Student papers	Universitas Diponegoro on 2025-06-16	<1%
249	Student papers	Universitas Islam Negeri Sumatera Utara on 2026-05-18	<1%

250	Student papers	Universitas Islam Riau on 2025-11-24	<1%
251	Student papers	Universitas Khairun on 2026-05-18	<1%
252	Student papers	Universitas Kristen Wira Wacana Sumba on 2026-06-24	<1%
253	Student papers	Universitas Muslim Indonesia on 2026-01-02	<1%
254	Student papers	Universitas Negeri Surabaya on 2026-05-18	<1%
255	Student papers	Universitas Sanata Dharma on 2026-06-15	<1%
256	Student papers	Universitas Satya Negara Indonesia on 2026-05-18	<1%
257	Student papers	Universitas Tarumanagara on 2025-12-15	<1%
258	Student papers	Universitas Tarumanagara on 2025-12-18	<1%
259	Publication	Virginia Prisela Mantiri, Stanly W. Alexander, Syermi S.E. Mintalangi. "Analysis of t...	<1%
260	Publication	Yudhi Selvina Selvina. "PEMETAAN DAN KLASIFIKASI WILAYAH RAWAN GEMPA DI ...	<1%
261	Internet	core.ac.uk	<1%
262	Internet	ejournal.unkhair.ac.id	<1%
263	Internet	greentech.fateta.unand.ac.id	<1%

264	Student papers	iGroup on 2015-08-06	<1%
265	Internet	indeksprestasi.blogspot.com	<1%
266	Internet	journal.stekom.ac.id	<1%
267	Internet	jurnal.stmikroyal.ac.id	<1%
268	Internet	jurnal.untan.ac.id	<1%
269	Internet	kkn.unnes.ac.id	<1%
270	Internet	lib.ibs.ac.id	<1%
271	Internet	repositori.unwira.ac.id	<1%
272	Internet	repository.uisu.ac.id	<1%
273	Internet	www.ejournalwiraraja.com	<1%
274	Internet	www.iicls.org	<1%
275	Publication	Jiyan Suhada, Hari Setyawan. "Analisis Perbandingan Kinerja Metode Load Balan...	<1%
276	Student papers	Sekolah Teknik Elektro & Informatika on 2025-07-23	<1%
277	Student papers	UIN Sultan Syarif Kasim Riau on 2020-07-15	<1%

278	Student papers	UNIVERSITAS BUDI LUHUR on 2026-01-27	<1%
279	Student papers	Universitas Islam Riau on 2025-11-24	<1%
280	Student papers	Universitas Islam Riau on 2026-06-04	<1%
281	Student papers	Universitas Islam Riau on 2026-06-15	<1%
282	Student papers	Universitas Sanata Dharma on 2026-06-15	<1%
283	Student papers	Abdullah Gul University on 2025-09-29	<1%
284	Student papers	Academic Library Consortium on 2026-01-25	<1%
285	Publication	Achmad Harpin Asrori, Nanang Kurnia Wahab. "Analisis Komparatif Model Decisi...	<1%
286	Student papers	Boston University on 2026-05-13	<1%
287	Publication	Delgio Liem Sorongan, Quido C Kainde, SONDY C KUMAJAS. "Web-Based Decision S...	<1%
288	Student papers	Fakultas Teknik on 2025-07-16	<1%
289	Publication	Kiki Putriani Siregar, Budi Triandi, Roslina Roslina. "Perbandingan Algoritma Mac...	<1%
290	Student papers	Sriwijaya University on 2021-12-03	<1%
291	Student papers	Tarumanagara University on 2026-06-21	<1%

292	Student papers	Universitas Bangka Belitung on 2026-01-19	<1%
293	Student papers	Universitas Bengkulu on 2025-10-27	<1%
294	Student papers	Universitas Bina Darma on 2022-07-18	<1%
295	Student papers	Universitas Brawijaya on 2016-11-04	<1%
296	Student papers	Universitas Islam Riau on 2026-03-06	<1%
297	Student papers	Universitas Islam Sumatera Utara on 2026-04-20	<1%
298	Student papers	Universitas Muhammadiyah Palembang on 2026-06-10	<1%
299	Student papers	Universitas Pamulang on 2026-05-04	<1%
300	Student papers	Universitas Putera Batam on 2026-02-03	<1%
301	Student papers	Universitas Sanata Dharma on 2026-06-12	<1%
302	Student papers	Universitas Tarumanagara on 2025-12-30	<1%
303	Internet	repo.ikipgribali.ac.id	<1%

BAB 1

PENDAHULUAN

A. Latar Belakang

Koperasi Simpan Pinjam (KSP) berperan strategis sebagai lembaga keuangan mikro yang memberikan akses pembiayaan bagi UMKM yang tidak terlayani oleh perbankan formal (Indriani, 2022). Sebagai penyangga ekonomi masyarakat, keberlanjutan operasional koperasi sangat bergantung pada kelancaran pengembalian pinjaman. Namun, agar peran strategis ini tetap terjaga dan risiko kerugian dapat diminimalisir, diperlukan proses penilaian risiko kredit yang objektif, konsisten, dan mampu menggambarkan kondisi nasabah secara menyeluruh.

Dalam praktiknya, proses penilaian risiko kredit di Koperasi Sri Hadi Dharma masih dilakukan secara manual. Penilaian masih mengandalkan data pribadi dan jaminan tanpa standar baku, sehingga keputusan cenderung subjektif dan kurang konsisten. Data nasabah juga masih tersimpan dalam Excel tanpa analisis lanjutan. Penggunaan Excel untuk manajemen risiko kredit rentan terhadap *human error* dan sulit digunakan untuk analisis prediktif (Padriyansyah & Pratiwi, 2025). Keterbatasan ini menyebabkan evaluasi risiko menjadi lambat dan tidak akurat, yang berdampak pada peningkatan potensi kredit macet. Kondisi ini sejalan dengan temuan bahwa banyak koperasi di Indonesia belum mengelola data secara optimal (Kementerian Koperasi dan UKM, 2023). Oleh karena itu, diperlukan sistem penilaian risiko berbasis data dengan pendekatan *hybrid* untuk menghasilkan segmentasi risiko yang objektif dan prediksi yang lebih akurat.

Sejumlah penelitian telah menawarkan pendekatan untuk mendukung proses penilaian kredit. Sistem Pendukung Keputusan berbasis metode *SMART* telah digunakan untuk memberikan rekomendasi kelayakan kredit, namun, proses penentuan bobot kriterianya masih bersifat subjektif (Prawiratama dkk., 2024). Berdasarkan hasil evaluasi model pada penelitian lain, algoritma *Random Forest* memperoleh akurasi tertinggi yaitu 89%,

disusul *Decision Tree* sebesar 84%, serta *Logistic Regression* dan *K-Nearest Neighbors* yang masing-masing sebesar 81% (Aziz & Rianto, 2024). Temuan tersebut diperkuat oleh penelitian lain yang menyatakan bahwa algoritma *Random Forest* mencapai akurasi 78,60% dan nilai *AUC* sebesar 0,907, yang dikategorikan sebagai performa prediksi baik sekali (Pahlevi dkk., 2023). Namun demikian, metode *supervised learning* seperti *Random Forest* tetap memerlukan data berlabel, sedangkan Koperasi Sri Hadi Dharma belum memiliki riwayat label tersebut karena pencatatan sebelumnya tidak terstruktur.

Metode *unsupervised* digunakan ketika data tidak memiliki label. Penelitian sebelumnya menunjukkan bahwa metode *K-Means* mampu mengelompokkan nasabah berdasarkan karakteristik tertentu untuk keperluan pemetaan profil (Tamba & Ekawati, 2024). Penelitian lain menunjukkan bahwa model *hybrid K-Means + KNN* dapat digunakan untuk segmentasi dan prediksi dengan hasil yang valid (Mandala dkk., 2022). Namun, metode *clustering* seperti *K-Means* hanya menghasilkan pengelompokan tanpa menyediakan skor risiko, bobot faktor risiko, atau nilai numerik risiko yang dapat digunakan langsung dalam keputusan kredit.

Berdasarkan kondisi tersebut, penelitian ini mengembangkan model *hybrid* yang menggabungkan metode *unsupervised* dan *supervised*. Metode *K-Means* digunakan untuk mengelompokkan nasabah berdasarkan atribut seperti besar pinjaman, sisa pinjaman, angsuran, jasa, tenor, tunggakan, serta atribut tambahan hasil *feature engineering* berupa tunggakan maksimum untuk membentuk label asumsi risiko. Label ini kemudian digunakan untuk melatih algoritma *Random Forest* sehingga menghasilkan *risk scoring* dan bobot pengaruh setiap variabel terhadap risiko kredit. Permasalahan utama dalam penilaian kredit koperasi bukan hanya potensi kredit macet, tetapi ketiadaan mekanisme penilaian risiko yang objektif dan berbasis data akibat tidak tersedianya label risiko historis, yang menyebabkan belum tersedianya sistem penilaian risiko kredit yang terstruktur dan dapat digunakan sebagai dasar pengambilan keputusan.

80

B. Identifikasi Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka dapat diidentifikasi beberapa permasalahan sebagai berikut:

1. Proses penilaian risiko kredit di Koperasi Sri Hadi Dharma masih dilakukan secara manual tanpa pedoman yang terstandar.
2. Koperasi tidak memiliki data berlabel seperti “lunas” atau “gagal bayar”, sehingga metode *supervised* tidak dapat langsung digunakan untuk membangun model prediktif.

11

C. Rumusan Masalah

Berdasarkan identifikasi masalah tersebut, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana merancang proses penilaian risiko kredit yang lebih objektif dan terstruktur pada Koperasi Sri Hadi Dharma ?
2. Bagaimana membangun model prediktif menggunakan kombinasi metode *unsupervised* dan *supervised* ketika data tidak memiliki label risiko?

117

D. Batasan Masalah

Agar penelitian lebih terarah dan fokus, terdapat beberapa batasan yang ditetapkan, di antaranya:

1. Objek penelitian: Penelitian ini berfokus pada data nasabah Koperasi Sri Hadi Dharma yang digunakan untuk proses pengelompokan dan pemodelan risiko kredit. Data yang digunakan dibatasi hanya pada nasabah lama yang telah memiliki riwayat pinjaman dan pembayaran angsuran, sehingga tidak mencakup nasabah baru yang belum memiliki rekam jejak pembayaran.
2. Variabel penelitian: Variabel penelitian mencakup besar pinjaman, sisa pinjaman, angsuran, jasa, tenor, tunggakan, serta atribut tambahan hasil *feature engineering* berupa tunggakan maksimum.
3. Metode penelitian: Penelitian ini menggunakan kombinasi metode *unsupervised* dan *supervised*, yaitu algoritma *K-Means* untuk

233

pembentukan label asumsi risiko (*pseudo-label*) dan *Random Forest* untuk pembangunan model prediksi risiko.

4. **Lingkup sistem:** Sistem dikembangkan dalam bentuk aplikasi berbasis web menggunakan *Streamlit* yang berfungsi sebagai *risk scoring system* dan visualisasi informasi risiko kredit.
5. **Jenis data:** Data yang digunakan berupa data historis nasabah yang tersimpan dalam bentuk (Excel) dan tidak mencakup data berlabel seperti “lunas” atau “gagal bayar”.
6. **Evaluasi model:** Evaluasi performa model menggunakan metrik prediksi, seperti *accuracy*, *precision*, *recall*, dan *F1-score*.
7. **Perangkat dan perangkat lunak:** Proses pemodelan dilakukan menggunakan Python dengan *library* seperti *pandas*, *scikit-learn*, dan *matplotlib*, serta dikerjakan melalui VS Code untuk pemrosesan dan visualisasi.

E. Tujuan Penelitian

Dengan rumusan masalah diatas, ditetapkan beberapa tujuan utama penelitian ini sebagai berikut:

1. Mengelompokkan nasabah Koperasi Sri Hadi Dharma menggunakan metode *K-Means* untuk membentuk label asumsi risiko berdasarkan karakteristik tertentu.
2. Membangun model prediksi risiko menggunakan algoritma *Random Forest* yang dilatih berdasarkan label asumsi risiko dari *K-Means*, sekaligus menghasilkan *risk scoring system* untuk menilai tingkat risiko setiap nasabah.

F. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan beberapa manfaat, sebagai berikut:

1. Manfaat Teoritis

- a. Memberikan kontribusi berupa penerapan pendekatan *hybrid* (*K-Means* + *Random Forest*) untuk menghasilkan *pseudo-label* dan model prediksi risiko pada data yang tidak berlabel.

205

- b. Menjadi referensi bagi penelitian selanjutnya terkait pengembangan sistem penilaian risiko berbasis *machine learning* pada lembaga keuangan mikro.

2. Manfaat Praktis

2

- a. Membantu Koperasi Sri Hadi Dharma dalam melakukan penilaian risiko kredit secara lebih objektif dan konsisten.
- b. Mendukung proses pengambilan keputusan kredit yang lebih cepat, lebih objektif, dan berbasis data.
- c. Menjadi dasar awal bagi koperasi untuk memanfaatkan data secara lebih optimal dalam *monitoring* risiko dan manajemen portofolio kredit.

BAB II

LANDASAN TEORI

A. Teori dan Penelitian Terdahulu

1. Landasan Teori

a. Koperasi Simpan Pinjam (KSP)

Koperasi Simpan Pinjam (KSP) merupakan lembaga ekonomi yang berperan penting dalam memberdayakan ekonomi masyarakat melalui penyediaan layanan simpanan dan pinjaman yang mudah diakses oleh nasabah maupun masyarakat yang membutuhkan. Keberadaan KSP membantu masyarakat memperoleh dana pinjaman secara cepat dan relatif mudah, sehingga mampu menghindarkan masyarakat dari praktik riba serta lembaga keuangan informal yang merugikan (Hasan et al., 2023). Selain itu, KSP memiliki keterkaitan yang erat dengan Usaha Mikro, Kecil, dan Menengah (UMKM) karena menjadi salah satu sumber pembiayaan utama bagi pelaku usaha dalam mengembangkan aktivitas ekonominya (Hasanah & Hanifah, 2020).

Peran Koperasi Simpan Pinjam juga terlihat dalam upaya pengembangan UMKM melalui penyediaan pembiayaan permodalan serta sebagai alternatif lembaga simpanan selain perbankan formal. KSP tidak hanya memberikan pinjaman modal, tetapi juga berkontribusi dalam pemberdayaan ekonomi masyarakat melalui kegiatan penyuluhan, pendampingan usaha, dan peningkatan kualitas pelayanan kepada nasabah (Fadliansyah dkk., 2024). Dengan pengelolaan koperasi yang berorientasi pada pelayanan dan peningkatan kepercayaan nasabah, KSP mampu mendorong peningkatan kesejahteraan masyarakat secara berkelanjutan.

Meskipun memiliki peran strategis dalam pembiayaan masyarakat, Koperasi Simpan Pinjam juga menghadapi risiko kredit yang relatif tinggi akibat karakteristik nasabahnya yang didominasi

134 oleh sektor informal dan UMKM. Keterbatasan jaminan, fluktuasi pendapatan, serta minimnya data riwayat kredit formal menyebabkan proses penilaian kelayakan pinjaman menjadi lebih kompleks dibandingkan lembaga perbankan. Oleh karena itu, diperlukan sistem penilaian risiko kredit yang lebih objektif dan berbasis data untuk mendukung keberlanjutan operasional KSP.

b. Sistem Penilaian Resiko Kredit

165 Sistem penilaian risiko kredit (*credit risk assessment*) merupakan proses sistematis untuk mengukur probabilitas kegagalan debitur dalam memenuhi kewajiban pembayaran pinjaman sesuai dengan perjanjian yang telah ditetapkan. Proses ini umumnya diwujudkan dalam bentuk *credit scoring*, yaitu pemberian skor numerik yang merepresentasikan tingkat risiko kredit seorang peminjam. Seiring dengan meningkatnya ketersediaan data operasional koperasi, pendekatan berbasis *machine learning* menjadi alternatif yang relevan dalam penilaian risiko kredit. *Machine learning* memungkinkan pemanfaatan data sederhana yang dimiliki koperasi, seperti karakteristik demografis dan pinjaman, untuk membentuk pola risiko secara lebih objektif dan konsisten. (Ampountolas, 2021). Penggunaan algoritma *Random Forest* dinilai efektif dalam sistem penilaian risiko kredit karena mampu mengprediksikan peminjam ke dalam beberapa kelas risiko dan memberikan hasil prediksi yang andal bagi lembaga keuangan mikro seperti KSP.

225 Pada praktiknya, penilaian risiko kredit di banyak Koperasi Simpan Pinjam masih dilakukan secara manual dan berbasis pengalaman petugas, sehingga berpotensi menimbulkan subjektivitas dan inkonsistensi dalam pengambilan keputusan. Kondisi ini menjadi semakin menantang ketika data kredit tidak memiliki label historis yang jelas, seperti informasi gagal bayar atau pelunasan. Oleh karena itu, pendekatan *machine learning* menjadi

253

61

alternatif yang relevan karena mampu mengekstraksi pola risiko dari data yang terbatas namun beragam.

Penerapan sistem penilaian risiko berbasis *machine learning* pada KSP juga relevan dalam pembiayaan sektor produktif, termasuk nasabah koperasi yang bergerak di bidang pertanian dan UMKM. Model *credit scoring* berbasis *Random Forest* dapat membantu KSP mengidentifikasi variabel penting yang memengaruhi risiko kredit, seperti biaya usaha, pendapatan, produktivitas, dan luas lahan, sehingga keputusan pemberian pinjaman menjadi lebih tepat sasaran (Bahri dkk., 2023). Selain itu, integrasi *machine learning* dengan pengelolaan data yang baik memungkinkan KSP meningkatkan efektivitas manajemen risiko pinjaman melalui tahapan praproses data, rekayasa fitur, dan evaluasi model, sehingga mampu meminimalkan risiko kredit bermasalah dan meningkatkan keberlanjutan koperasi (Nuthalapati, 2022).

c. Algoritma K-Means Clustering

K-Means adalah algoritma *clustering* yang digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik sehingga pola tersembunyi dalam dataset dapat teridentifikasi tanpa memerlukan label awal. Algoritma ini banyak diterapkan pada analisis finansial, pemasaran, dan manajemen risiko karena kemampuannya dalam melakukan segmentasi secara efisien (Ikotun dkk., 2023). Selain itu, algoritma K-Means telah diterapkan dalam berbagai bidang untuk mengelompokkan data berdasarkan karakteristik tertentu, salah satunya pada klasifikasi emosi wajah (Zuhal dkk., 2021).

Dalam penelitian komparatif, *K-Means* dan *K-Means++* terbukti mampu melakukan segmentasi data secara akurat untuk analisis kelompok nasabah, sehingga mempermudah proses identifikasi risiko dasar (Bangkalang dkk., 2023). Pada sektor keuangan, *K-*

Means banyak digunakan untuk segmentasi awal sebelum membangun model *supervised learning* (Chitra & Heikal, 2024).

Dalam konteks penelitian ini, algoritma *K-Means* digunakan sebagai pendekatan awal untuk mengelompokkan data pinjaman nasabah yang belum memiliki label risiko kredit. Hasil pengelompokan ini selanjutnya diperlakukan sebagai *pseudo-label* risiko, yang merepresentasikan tingkat risiko relatif antar kelompok nasabah berdasarkan kemiripan karakteristik data.

Rumus dan Tahapan Algoritma K-Means (Farida & Lubis, 2025)

1) *Euclidean Distance*

Euclidean Distance digunakan untuk mengukur jarak antara suatu data dengan *centroid cluster*. Jarak ini menjadi dasar dalam menentukan kedekatan suatu data terhadap masing-masing cluster. Secara matematis, *Euclidean Distance* dirumuskan pada Persamaan (2.1):

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots\dots\dots(2.1)$$

x_i dan y_i masing-masing merepresentasikan nilai atribut ke- i dari data dan *centroid*, sedangkan n menunjukkan jumlah atribut. Nilai jarak yang paling kecil menunjukkan bahwa data tersebut paling dekat dengan centroid tertentu dan akan dikelompokkan ke dalam cluster tersebut.

2) *Normalisasi Data*

Sebelum proses *clustering* menggunakan algoritma *K-Means*, data numerik perlu dinormalisasi untuk menyamakan skala antar atribut. Hal ini dilakukan karena algoritma *K-Means* menggunakan perhitungan jarak *Euclidean* yang sensitif terhadap perbedaan rentang nilai data. Atribut dengan skala yang lebih besar berpotensi mendominasi hasil pengelompokan apabila tidak dilakukan normalisasi terlebih dahulu.

Salah satu metode normalisasi yang umum digunakan adalah *Min-Max Normalization*, yang memetakan nilai data ke dalam rentang $[0,1]$, sebagaimana ditunjukkan pada Persamaan (2.2):

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \dots \dots \dots (2.2)$$

Keterangan:

x' adalah nilai hasil normalisasi

x adalah nilai asli data

$x_{\min}$ dan $x_{\max}$ masing-masing adalah nilai minimum dan maksimum pada atribut terkait

Dengan normalisasi ini, seluruh atribut memiliki skala yang sebanding sehingga proses *clustering* menggunakan *K-Means* dapat menghasilkan pengelompokan yang lebih objektif dan stabil.

3) Fungsi Objektif (*Within-Cluster Sum of Squares* / WCSS)

Fungsi objektif pada algoritma *K-Means* bertujuan untuk meminimalkan variasi data di dalam *cluster*. Ukuran yang digunakan adalah *Within-Cluster Sum of Squares* (WCSS), yang merepresentasikan jumlah kuadrat jarak antara setiap data dengan *centroid* cluster-nya. Fungsi objektif ini dirumuskan pada Persamaan (2.3):

$$WCSS = \sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - \mu_j\|^2 \dots \dots \dots (2.3)$$

K menyatakan jumlah *cluster*, C_j adalah himpunan data pada *cluster* ke- j , dan μ_j adalah *centroid cluster* ke- j . Nilai WCSS yang lebih kecil menunjukkan bahwa data dalam *cluster* semakin homogen, sehingga kualitas pengelompokan menjadi lebih baik.

4) Pembaruan *Centroid*

Setelah data dikelompokkan ke dalam *cluster* sementara, posisi *centroid* perlu diperbarui agar merepresentasikan pusat *cluster* yang baru. Pembaruan *centroid* dilakukan dengan menghitung nilai rata-rata dari seluruh data yang tergabung dalam *cluster* tersebut, sebagaimana ditunjukkan pada Persamaan (2.4):

$$\mu_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i \quad \dots\dots\dots(2.4)$$

$|C_j|$ menyatakan jumlah data dalam *cluster* ke- j . Proses pembaruan *centroid* ini memastikan bahwa posisi *centroid* selalu berada di titik tengah distribusi data pada masing-masing *cluster*.

5) Uji Stabilitas Cluster

Uji stabilitas *cluster* bertujuan untuk mengevaluasi konsistensi hasil pengelompokan yang dihasilkan oleh algoritma *clustering*. Salah satu ukuran yang umum digunakan adalah *Adjusted Rand Index (ARI)*. ARI mengukur tingkat kesamaan antara dua hasil *clustering* dengan mempertimbangkan kemungkinan kesamaan yang terjadi secara acak. Nilai ARI berada pada rentang -1 hingga 1, dimana nilai mendekati 1 menunjukkan bahwa hasil *clustering* memiliki tingkat konsistensi yang tinggi, sedangkan nilai mendekati 0 menunjukkan kesamaan yang terjadi secara acak.

6) Tahapan Algoritma K-Means

- a) Tentukan jumlah *cluster* k .
- b) Inisialisasi *centroid* awal secara acak atau heuristik.
- c) Hitung jarak setiap data ke *centroid* (misal menggunakan *Euclidean Distance*).
- d) Kelompokkan data ke *centroid* terdekat, membentuk *cluster* sementara.
- e) Hitung ulang *centroid* berdasarkan rata-rata nasabah *cluster*.
- f) Ulangi langkah 3–5 hingga *centroid* stabil (konvergen).

d. Algoritma Random Forest

15

Random Forest merupakan algoritma *ensemble learning* yang menggabungkan banyak *decision tree* untuk meningkatkan stabilitas dan akurasi prediksi pada data yang beragam dan kompleks (Firmansah dkk., 2023). Dalam konteks penilaian kredit, *Random Forest* mampu memproses variabel yang tidak linear dan beragam sehingga mendukung keputusan kelayakan kredit secara lebih objektif, terutama pada lembaga keuangan mikro dan koperasi.

95

Model ini terbukti efektif untuk memprediksi risiko kredit pada program pembiayaan UMKM seperti Kredit Usaha Rakyat (KUR), khususnya dalam mengidentifikasi debitur berisiko tinggi dan menekan potensi *Non-Performing Loan* (Wulansari & Purwitasari, 2023a). *Random Forest* juga memberikan informasi bobot variabel (*feature importance*) untuk mengetahui faktor risiko utama (Ulandari dkk., 2024). Selain itu, *hyperparameter tuning* dapat meningkatkan akurasi model secara signifikan (Bhakti dkk., 2024).

3

Pemilihan *Random Forest* dalam penelitian ini didasarkan pada kemampuannya dalam menangani data yang heterogen, tidak linier, serta relatif tahan terhadap *noise* dan *overfitting*. Selain itu, *Random Forest* dapat memanfaatkan informasi tambahan hasil *clustering* untuk meningkatkan kualitas prediksi, sehingga cocok digunakan dalam pendekatan *hybrid* pada penilaian risiko kredit koperasi.

Rumus dan Tahapan Algoritma *Random Forest* (Yuda dkk., 2025).

1) Rumus *Gini Indeks*

Gini Index digunakan untuk mengukur tingkat *impurity* atau ketidakhomogenan data pada suatu *node* dalam *decision tree*. Nilai Gini yang semakin kecil menunjukkan bahwa data dalam *node* tersebut semakin *homogen*. Secara matematis, *Gini Index* dirumuskan sebagaimana ditunjukkan pada Persamaan (2.5) :

60

$$Gini(D) = 1 - \sum_{i=1}^n p_i^2 \dots\dots\dots(2.5)$$

dimana p_i adalah proporsi kelas ke- i dalam dataset D .

2) Rumus *Entropy*

Entropy digunakan untuk mengukur tingkat ketidakpastian atau *impurity* dalam suatu *node*. Semakin besar nilai *Entropy*, semakin tinggi tingkat ketidakpastian data pada *node* tersebut. Rumus *Entropy* secara matematis ditunjukkan pada Persamaan (2.6) :

$$Entropy(D) = - \sum_{i=1}^n p_i \log_2 p_i \dots\dots\dots(2.6)$$

3) Rumus *Majority Voting*

Pada *Random Forest*, hasil prediksi akhir ditentukan melalui mekanisme *majority voting*, yaitu kelas yang paling banyak dipilih oleh seluruh *decision tree* akan menjadi hasil prediksi akhir. Secara matematis, mekanisme *majority voting* dirumuskan pada Persamaan (2.7) :

$$\hat{y} = \text{mode}(h_1(x), h_2(x), \dots, h_k(x)) \dots\dots\dots(2.7)$$

dimana $h_1, h_2, \dots, h_k$ adalah hasil prediksi masing-masing *decision tree*, dan $\hat{y}$ adalah prediksi akhir *Random Forest*.

4) Probabilitas

Selain menghasilkan label prediksi, algoritma *Random Forest* juga mampu memberikan nilai *probabilitas* untuk setiap kelas. *Probabilitas* ini menunjukkan tingkat keyakinan model terhadap hasil prediksi yang diberikan.

Probabilitas kelas dihitung berdasarkan proporsi jumlah *decision tree* yang memilih kelas tertentu dibandingkan dengan jumlah seluruh pohon dalam model. Secara matematis, *probabilitas* prediksi dirumuskan pada Persamaan (2.8):

$$P(\text{kelas}) = \frac{\text{jumlah pohon yang memilih kelas}}{\text{total pohon}} \dots\dots\dots (2.8)$$

Nilai *probabilitas* yang lebih besar menunjukkan tingkat keyakinan model yang lebih tinggi terhadap kelas tersebut. Informasi *probabilitas* ini dapat digunakan sebagai bahan pertimbangan tambahan dalam pengambilan keputusan penilaian risiko kredit.

5) *Feature Importance*

Selain menghasilkan prediksi, algoritma *Random Forest* juga dapat digunakan untuk mengetahui tingkat kepentingan setiap atribut melalui *feature importance*. *Feature importance* menunjukkan besarnya kontribusi masing-masing variabel terhadap hasil prediksi, sehingga dapat digunakan untuk mengidentifikasi atribut yang paling berpengaruh dalam pembentukan model (Newaz dkk., 2022).

6) Tahapan Algoritma *Random Forest*

- a) Melakukan *bootstrap* sampling untuk membuat *subset* data dari dataset asli.
- b) Membentuk *decision tree* pada setiap *subset* data.
- c) Memilih fitur secara acak pada setiap node untuk *split*.
- d) Menentukan *split* terbaik menggunakan ukuran *impurity* seperti Gini atau *Entropy*.
- e) Menggabungkan hasil prediksi semua *tree* dengan *majority voting* untuk menentukan prediksi akhir.

e. Pendekatan Hybrid



Gambar 2. 1 Alur Proses *Hybrid*

Pendekatan *hybrid learning* dalam penilaian risiko kredit merupakan strategi yang mengintegrasikan metode *unsupervised learning* dan *supervised learning* untuk meningkatkan akurasi serta stabilitas model prediksi. Pada pendekatan ini, teknik *unsupervised*

seperti *clustering* digunakan pada tahap awal untuk mengidentifikasi pola dan segmentasi nasabah berdasarkan kemiripan karakteristik, sebelum dilanjutkan dengan metode *supervised* untuk melakukan prediksi risiko kredit. (Xu, 2024) menjelaskan bahwa penggabungan kedua pendekatan tersebut memungkinkan model memanfaatkan struktur laten data yang tidak berlabel, sehingga proses penilaian risiko menjadi lebih adaptif dan mampu menangani kompleksitas data keuangan yang tinggi.

Selanjutnya, hasil segmentasi dari tahap *clustering* dapat digunakan sebagai informasi tambahan atau *pseudo-label* dalam proses prediksi, sehingga model *supervised* seperti *Random Forest* dapat bekerja pada kelompok data yang lebih homogen. Penelitian oleh (Machado & Karray, 2022) menunjukkan bahwa penerapan *hybrid machine learning* dengan tahapan awal *clustering* dan dilanjutkan prediksi mampu meningkatkan kinerja *credit scoring* dibandingkan penggunaan metode tunggal. Hal serupa juga dikemukakan oleh (Atteba dkk., 2025), yang menyatakan bahwa integrasi *unsupervised* dan *supervised learning* menghasilkan evaluasi risiko kredit yang lebih konsisten, interpretatif, dan efektif, khususnya pada kondisi data yang terbatas atau tidak memiliki label risiko yang jelas, sehingga relevan diterapkan pada lembaga keuangan mikro dan pembiayaan UMKM.

Berbeda dengan penelitian sebelumnya yang umumnya menerapkan metode *clustering* dan prediksi secara terpisah, penelitian ini mengintegrasikan kedua pendekatan tersebut dalam satu alur sistem penilaian risiko kredit. Pendekatan *hybrid* ini diharapkan mampu menjembatani keterbatasan data tidak berlabel sekaligus meningkatkan ketepatan prediksi risiko kredit pada Koperasi Simpan Pinjam.

f. Stratified Sampling

Stratified sampling merupakan teknik pembagian data yang dilakukan dengan mempertahankan proporsi setiap kelas pada data pelatihan dan data pengujian. Teknik ini banyak digunakan pada permasalahan klasifikasi, terutama pada dataset yang memiliki distribusi kelas tidak seimbang, karena mampu menghasilkan sampel yang lebih representatif terhadap distribusi data asli. Dengan menjaga proporsi masing-masing kelas, model dapat dilatih dan dievaluasi secara lebih konsisten sehingga mengurangi bias terhadap kelas mayoritas (Newaz dkk., 2022). Pada penelitian ini, *stratified sampling* digunakan untuk membagi data menjadi data pelatihan sebesar 80% dan data pengujian sebesar 20% dengan tetap mempertahankan proporsi kategori Risiko Rendah, Risiko Sedang, dan Risiko Tinggi.

g. Evaluasi Model

Evaluasi model merupakan tahap penting untuk menilai kinerja algoritma dalam melakukan prediksi dan menentukan seberapa baik model mampu memprediksikan objek sesuai dengan kategori yang benar (Opitz, 2024). Pada tugas prediksi, seperti penilaian risiko kredit, evaluasi model membantu memastikan bahwa prediksi yang dihasilkan akurat dan dapat diandalkan (Noriega dkk., 2023). Pemilihan metrik evaluasi harus disesuaikan dengan tujuan dan karakteristik data. Misalnya, pada sektor keuangan, kesalahan prediksi dapat menimbulkan kerugian signifikan (Rainio dkk., 2024).

Dalam konteks penilaian risiko kredit, pemilihan metrik evaluasi tidak hanya berfokus pada akurasi, tetapi juga pada kemampuan model dalam meminimalkan kesalahan yang berisiko tinggi bagi koperasi. Kesalahan dalam memprediksikan nasabah berisiko tinggi sebagai layak kredit dapat berdampak langsung pada meningkatnya kredit bermasalah, sehingga metrik seperti *precision*, *recall*, dan *F1-score* menjadi sangat penting.

Beberapa metrik evaluasi yang umum digunakan antara lain (Ma'rufudin & Yudhistira, 2025):

1) *Accuracy*

Accuracy mengukur persentase prediksi yang benar dari keseluruhan data uji. Metrik ini dihitung berdasarkan perbandingan antara jumlah prediksi benar (*True Positive* dan *True Negative*) terhadap total seluruh prediksi, sebagaimana ditunjukkan pada Persamaan (2.9). Meskipun demikian, pada dataset yang tidak seimbang, nilai *accuracy* dapat memberikan gambaran yang menyesatkan karena dominasi kelas mayoritas dapat meningkatkan nilai akurasi secara semu.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad \dots\dots\dots(2.9)$$

2) *Precision*

Precision mengukur tingkat ketepatan model dalam memprediksi kelas positif, yaitu perbandingan antara jumlah prediksi positif yang benar terhadap seluruh prediksi positif yang dihasilkan model. Perhitungan *precision* dirumuskan pada Persamaan (2.10). Metrik ini menjadi sangat penting ketika kesalahan prediksi positif (*False Positive*) memiliki konsekuensi yang signifikan, seperti pada penilaian risiko kredit.

$$\text{Precision} = \frac{TP}{TP + FP} \quad \dots\dots\dots(2.10)$$

3) *Recall*(*Sensitivity*)

Recall menunjukkan kemampuan model dalam mengidentifikasi seluruh kasus positif yang sebenarnya. Nilai *recall* diperoleh dari perbandingan antara jumlah prediksi positif yang benar terhadap seluruh data yang benar-

benar termasuk dalam kelas positif, sebagaimana dirumuskan pada Persamaan (2.11). Metrik ini penting ketika kesalahan negatif (*False Negative*) berpotensi menimbulkan risiko yang tinggi.

$$\text{Recall} = \frac{TP}{TP + FN} \dots\dots\dots(2.11)$$

4) *F1-Score*

F1-Score merupakan ukuran gabungan yang menyeimbangkan nilai *precision* dan *recall* melalui rata-rata harmonik. Perhitungan *F1-Score* ditunjukkan pada Persamaan (2.12). Metrik ini lebih representatif untuk mengevaluasi model pada kondisi dataset yang tidak seimbang karena mempertimbangkan dua aspek kesalahan prediksi sekaligus.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \dots\dots\dots(2.12)$$

5) *Confusion Matrix*

Confusion matrix merupakan tabel yang menyajikan ringkasan hasil prediksi model dengan membandingkan label aktual dan label hasil prediksi. Struktur *confusion matrix* ditunjukkan pada Persamaan (2.13), yang menjadi dasar perhitungan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Melalui *confusion matrix*, jenis kesalahan prediksi yang dilakukan model dapat dianalisis secara lebih rinci.

$$\text{Confusion Matrix} = \begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix} \dots\dots\dots(2.13)$$

Keterangan simbol:

TP (*True Positive*): jumlah prediksi positif yang benar

TN (*True Negative*): jumlah prediksi negatif yang benar

FP (*False Positive*): jumlah prediksi positif yang salah

FN (*False Negative*): jumlah prediksi negatif yang salah

6) *Classification report*

Classification report merupakan ringkasan hasil evaluasi model yang memuat nilai *precision*, *recall*, *F1-score*, dan *support* untuk masing-masing kelas. *Classification report* digunakan untuk memberikan informasi yang lebih rinci mengenai performa model pada setiap kategori dibandingkan hanya menggunakan nilai *accuracy* secara keseluruhan.

2. Kajian Pustaka

Penelitian (Irawan dkk., 2025) memanfaatkan algoritma *K-Means clustering* untuk mengelompokkan nasabah bank berdasarkan karakteristik demografis dan finansial sebagai upaya memahami keragaman profil nasabah yang selama ini sulit dipetakan secara sistematis. Dengan menggunakan 600 data nasabah dan variasi jumlah kluster, penelitian ini menunjukkan bahwa pemilihan nilai K yang tepat mampu menghasilkan segmentasi yang jelas dan bermakna. Hasil pengelompokan memperlihatkan perbedaan karakteristik nasabah dari sisi preferensi produk, tingkat keterlibatan layanan, dan kecenderungan risiko, sehingga dapat dimanfaatkan sebagai dasar penyusunan strategi pemasaran dan pengelolaan risiko. penelitian ini berfokus pada segmentasi deskriptif tanpa menghasilkan nilai risiko numerik atau model prediktif, serta diterapkan pada data perbankan yang relatif terstruktur dan berskala menengah, sehingga hasilnya lebih bersifat pemetaan karakteristik.

Pendekatan serupa juga diterapkan oleh (Rosadi et al, 2025), Namun dalam konteks yang berbeda, yaitu pengelompokan siswa SMK berdasarkan nilai praktik dan minat karier di bidang Teknologi Informasi. Dengan jumlah data yang lebih kecil dan variabel *non-finansial*, *K-Means* digunakan untuk mengeksplorasi pola kemampuan dan kecenderungan siswa guna mendukung pengambilan keputusan di

274 bidang pendidikan. Hasil penelitian menunjukkan terbentuknya beberapa kelompok dengan karakteristik akademik yang berbeda, sehingga memberikan gambaran awal mengenai kebutuhan pendampingan dan strategi pembelajaran. Penelitian ini menggunakan data *non-finansial* di bidang pendidikan, dan memiliki skala data yang kecil serta tujuan eksploratif, sehingga tidak diarahkan pada analisis risiko atau prediksi.

58 Keterbatasan tersebut mulai diatasi dalam penelitian (Hasibuan & Mahyuni, 2022) yang mengaplikasikan *K-Means clustering* pada data nasabah kredit bermasalah di sektor perbankan mikro. Dengan memanfaatkan variabel kredit seperti jumlah tunggakan, sisa tunggakan, dan limit kredit, penelitian ini berhasil mengelompokkan nasabah ke dalam beberapa kategori kondisi kredit yang lebih operasional. Hasil segmentasi memberikan gambaran yang lebih spesifik mengenai tingkat permasalahan kredit nasabah dan memudahkan pihak manajemen dalam pengambilan keputusan. Penelitian ini secara khusus berfokus pada nasabah kredit bermasalah, serta dilengkapi dengan implementasi sistem aplikasi, meskipun analisis yang dihasilkan masih bersifat statis dan belum prediktif.

58 Perkembangan pendekatan analisis risiko kredit selanjutnya ditunjukkan oleh (Rahmawati & Handayani, 2025) melalui penerapan algoritma *Random Forest* untuk mengprediksikan kelayakan kredit nasabah. Berbeda dari metode *clustering* yang bersifat eksploratif, *Random Forest* digunakan untuk menghasilkan keputusan prediksi yang lebih tegas antara nasabah layak dan tidak layak kredit. Hasil penelitian menunjukkan tingkat akurasi yang cukup tinggi dengan performa prediksi yang stabil. Penelitian ini memfokuskan penilaian risiko pada kelayakan kredit awal nasabah, diterapkan pada konteks BPR daerah dengan variabel finansial yang relatif terbatas, sehingga model lebih diarahkan pada keputusan operasional kredit dibandingkan analisis risiko yang bersifat komprehensif dan berkelanjutan.

Pendekatan prediktif tersebut dikembangkan lebih lanjut oleh (Wulansari & Purwitasari, 2023) dengan memfokuskan analisis pada prediksi kredit macet KUR menggunakan *Random Forest* pada data yang tidak seimbang. Dengan menerapkan teknik *SMOTE*, penelitian ini mampu meningkatkan kualitas model dan menghasilkan akurasi yang lebih tinggi dibandingkan metode pembandingan. Selain menghasilkan prediksi yang lebih akurat, penelitian ini juga mengidentifikasi faktor-faktor dominan penyebab kredit macet. Penelitian ini menangani permasalahan ketidakseimbangan data, serta memanfaatkan *feature importance* sebagai dasar analisis risiko, sehingga memberikan kontribusi analitis yang lebih mendalam.

Skala dan kompleksitas analisis semakin diperluas dalam penelitian (Hazizah dkk., 2025) yang menerapkan *Random Forest* pada prediksi risiko gagal bayar kartu kredit menggunakan dataset berskala besar. Dengan memanfaatkan variabel perilaku finansial nasabah, model mampu mempertahankan performa yang stabil meskipun dihadapkan pada data yang kompleks. penelitian ini menggunakan dataset berskala besar dengan variabel perilaku finansial, serta menekankan ketahanan *Random Forest* terhadap *outlier* dan *missing value*, sehingga relevan untuk pengambilan keputusan strategis di perbankan modern. Namun, penelitian-penelitian tersebut umumnya masih menerapkan pendekatan *clustering* dan prediksi atau prediksi secara terpisah, sehingga peluang penggabungan keduanya dalam satu sistem penilaian risiko kredit yang terintegrasi masih terbuka.

Berdasarkan kajian pustaka tersebut, dapat disimpulkan bahwa penelitian terdahulu telah memberikan kontribusi signifikan dalam penggunaan *K-Means* dan *Random Forest* pada berbagai konteks. Namun, masih terbatas penelitian yang mengintegrasikan kedua metode tersebut dalam satu sistem penilaian risiko kredit yang terstruktur, khususnya pada Koperasi Simpan Pinjam dengan keterbatasan data

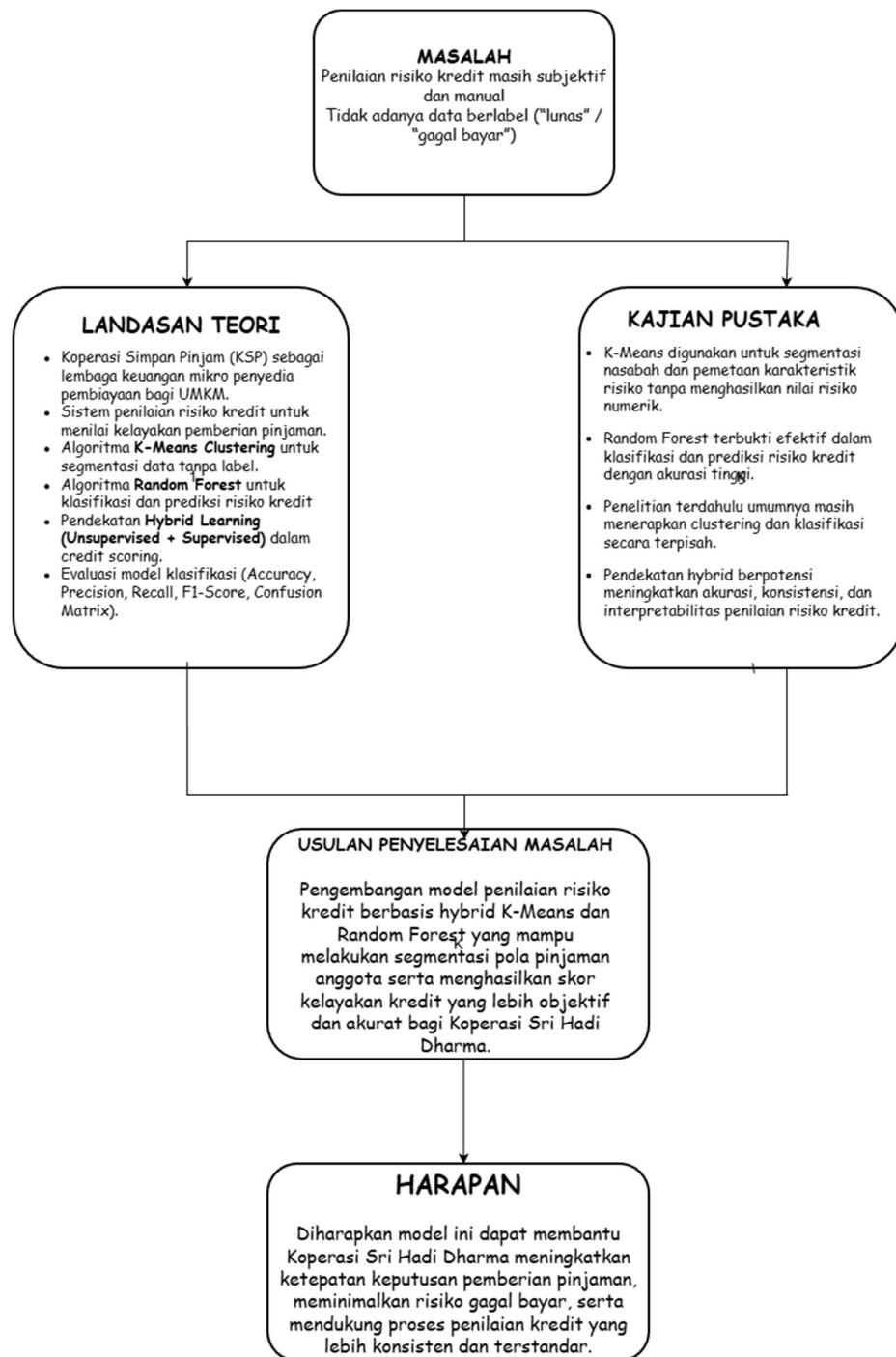
162

berlabel. Oleh karena itu, penelitian ini berupaya mengisi celah tersebut melalui pendekatan *hybrid learning*

B. Kerangka Berpikir

283

Kerangka berpikir pada Gambar 2.2 merepresentasikan alur logis penelitian dalam menangani permasalahan penilaian risiko kredit di Koperasi Sri Hadi Dharma. Kerangka ini menggambarkan hubungan antara masalah, landasan teori, kajian pustaka, dan alur pemikiran menuju solusi berupa pengembangan model penilaian risiko kredit yang lebih objektif, terukur, dan terstandar.



Gambar 2. 2 Kerangka Berpikir

Gambar 2.2 menunjukkan bagaimana masalah penilaian risiko kredit yang masih subjektif dan tidak memiliki data berlabel dapat ditangani melalui pemanfaatan landasan teori dan kajian pustaka yang relevan. Alur kerja sistem model yang diusulkan diawali dari pemetaan

masalah, dilanjutkan dengan pemilihan konsep dan metode, yaitu *K-Means* untuk segmentasi nasabah dan *Random Forest* untuk prediksi kelayakan kredit kemudian dilanjutkan dengan pemilihan konsep dan metode seperti *K-Means* untuk segmentasi nasabah serta *Random Forest* untuk prediksi kelayakan kredit.

1. Masalah

Penelitian ini berangkat dari dua permasalahan utama pada Koperasi Sri Hadi Dharma:

- a. Penilaian risiko kredit masih bersifat subjektif dan manual, sehingga keputusan pemberian pinjaman rentan bias dan tidak konsisten antar petugas.
- b. Tidak adanya data berlabel seperti keterangan “lunas” atau “gagal bayar”, sehingga penggunaan metode supervised learning secara langsung menjadi tidak memungkinkan.

Kondisi ini menyebabkan analisis risiko sulit dilakukan secara akurat dan terstandar.

2. Landasan Teori

Bagian ini berisi konsep-konsep dasar yang menjadi fondasi penelitian, meliputi:

- a. Koperasi Simpan Pinjam (KSP): karakteristik, mekanisme pinjaman, dan risiko gagal bayar.
- b. Sistem Penilaian Risiko Kredit: definisi, indikator risiko, dan pentingnya penilaian yang objektif.
- c. Algoritma *K-Means Clustering* digunakan untuk mengelompokkan pola pinjaman nasabah berdasarkan karakteristik tertentu (*unsupervised*).
- d. Algoritma *Random Forest* digunakan sebagai model prediksi untuk menghasilkan skor atau bobot risiko (*supervised*).
- e. Pendekatan *Hybrid* menggabungkan hasil segmentasi *K-Means* dengan prediksi *Random Forest*.

- f. Evaluasi model digunakan untuk menilai performa prediksi seperti akurasi, *precision*, *recall*, dan sebagainya.

Landasan teori ini menjadi pijakan metodologis bagi pengembangan model dalam penelitian.

3. Kajian Pustaka

Beberapa penelitian sebelumnya menunjukkan pemanfaatan **K-Means** dan *Random Forest*:

a. **K-Means Clustering**:

- 1) *Irawan dkk. (2025)* mengelompokkan nasabah bank berdasarkan karakteristik demografis dan finansial, menghasilkan segmentasi bermakna meski bersifat deskriptif tanpa skor risiko.
- 2) *Rosadi dkk. (2025)* menerapkan K-Means pada siswa SMK untuk mengeksplorasi pola kemampuan dan minat karier, meski skala data kecil dan non-finansial.
- 3) *Hasibuan & Mahyuni (2022)* mengelompokkan nasabah kredit bermasalah di perbankan mikro, menghasilkan segmentasi operasional untuk keputusan manajemen, meskipun analisis masih statis.

b. **Random Forest**:

- 1) *Rahmawati & Handayani (2025)* menilai kelayakan kredit nasabah BPR dengan akurasi tinggi, fokus pada keputusan operasional.
- 2) *Wulansari & Purwitasari (2023)* memprediksi kredit macet KUR dengan mengatasi ketidakseimbangan data menggunakan SMOTE, menghasilkan prediksi lebih akurat dan mengidentifikasi faktor dominan risiko.
- 3) *Hazizah dkk. (2025)* memprediksikan risiko gagal bayar kartu kredit pada dataset besar, menunjukkan ketahanan metode terhadap outlier dan missing value, relevan untuk keputusan strategis.

Meskipun demikian, penelitian terdahulu umumnya masih menerapkan *clustering* dan prediksi secara terpisah, sehingga potensi penggabungan kedua metode dalam pendekatan *hybrid* untuk penilaian risiko yang terintegrasi masih terbuka.

4. Usulan Penyelesaian Masalah

Berdasarkan masalah dan teori yang mendukung, penelitian ini mengusulkan:

- a. Mengelompokkan pola pinjaman nasabah menggunakan *K-Means (segmentation)*.
- b. Melakukan prediksi tingkat risiko kredit dan menghasilkan *risk score* menggunakan algoritma Random Forest.
- c. Menyediakan hasil penilaian yang lebih objektif, akurat, dan bisa dijadikan standar oleh Koperasi Sri Hadi Dharma.

Pendekatan *hybrid* membantu mengatasi masalah data tidak berlabel, karena hasil *clustering* dapat menjadi dasar untuk pembentukan label awal sebelum prediksi.

5. Harapan

Penelitian ini diharapkan dapat:

- a. Membantu koperasi meningkatkan ketepatan keputusan pemberian pinjaman.
- b. Meminimalkan risiko gagal bayar melalui prediksi dan skor risiko yang lebih akurat.
- c. Mendukung proses penilaian kredit yang lebih konsisten, terukur, dan sesuai standar.

Dengan adanya model ini, Koperasi Sri Hadi Dharma dapat beralih dari penilaian subjektif ke penilaian berbasis data dan *machine learning*.

BAB III

METODE PENELITIAN

A. Desain Penelitian

Desain penelitian ini disusun untuk memberikan kerangka sistematis dalam pengembangan dan evaluasi sistem penilaian risiko kredit berbasis *machine learning* dengan pendekatan *hybrid K-Means* dan *Random Forest*. Pendekatan penelitian diarahkan pada tahapan analisis, perancangan, pengembangan, dan evaluasi model sehingga menghasilkan sistem penilaian risiko yang dapat digunakan secara praktis. Desain ini menjadi dasar dalam penentuan jenis penelitian, variabel yang diteliti, serta metode pengumpulan data yang digunakan pada penelitian ini.

1. Jenis Penelitian(*Research Type*)

Penelitian ini menggunakan pendekatan *Development Research* (penelitian pengembangan) karena fokusnya tidak hanya pada pengujian algoritma, tetapi pada perancangan dan pembangunan sistem penilaian risiko kredit yang komprehensif sebagai produk akhir penelitian. Pengembangan dilakukan melalui model *hybrid* yang menggabungkan *K-Means* untuk segmentasi data pinjaman dan *Random Forest* untuk prediksi, sehingga menghasilkan *risk scoring* yang terstruktur dan dapat dioperasikan secara sistematis sesuai dengan kebutuhan operasional koperasi. Penelitian ini termasuk pengembangan perangkat lunak, karena mencakup desain, implementasi, pengujian, dan penyempurnaan sistem berbasis komputer yang dapat digunakan Koperasi Sri Hadi Dharma sebagai alat bantu pengambilan keputusan kredit. Dengan pendekatan ini, penelitian dilakukan secara bertahap mulai dari analisis kebutuhan, perancangan sistem, implementasi model *hybrid*, hingga evaluasi performa, sehingga produk akhir penelitian berupa sistem penilaian risiko kredit yang fungsional dan siap digunakan sebagai alat bantu pengambilan keputusan di koperasi.

2. Variabel Penelitian

a. Variabel Independent(Bebas/Prediktor)

Variabel independen pada penelitian ini merupakan variabel yang dikendalikan atau dimodifikasi oleh peneliti untuk melihat pengaruhnya terhadap kinerja model penilaian risiko kredit. Variabel independen dalam penelitian ini meliputi Besar Pinjaman, Sisa Pinjaman, Angsuran Pokok, Jasa, Tenor dan Tunggakan. Variabel-variabel ini dipilih karena berpotensi memengaruhi performa model secara signifikan sehingga diperlukan pengaturan yang tepat untuk memperoleh konfigurasi model yang optimal.

b. Variabel Dependent(Terikat/Respon)

Variabel dependen dalam penelitian ini merupakan keluaran utama yang dihasilkan oleh model setelah dilakukan manipulasi terhadap variabel independen. Variabel ini mencakup kinerja model dalam memprediksi tingkat risiko kredit nasabah koperasi, yang dievaluasi melalui metrik prediksi seperti akurasi, *precision*, *recall*, dan *F1-score*. Selain itu, variabel dependen juga mencakup output sistem berupa skor risiko dan label risiko kredit (Risiko Rendah, Sedang dan Tinggi) yang dihasilkan oleh model *hybrid* K-Means dan *Random Forest*. Pengukuran variabel ini digunakan untuk menilai sejauh mana model mampu memberikan hasil prediksi yang tepat, stabil, dan dapat diandalkan, sesuai dengan tujuan pengembangan sistem penilaian risiko kredit pada Koperasi Sri Hadi Dharma.

c. Variabel Kontrol (pengendali)

Variabel kontrol dalam penelitian ini mencakup faktor-faktor yang dipertahankan konstan untuk menjaga validitas internal penelitian, sehingga setiap perubahan performa model benar-benar berasal dari variabel independen. Spesifikasi perangkat keras yang digunakan dijaga tetap, yaitu laptop dengan prosesor AMD Ryzen

3 7320U with Radeon Graphics (2.40 GHz), RAM 8 GB, sistem operasi Windows 11 64-bit, serta penyimpanan berbasis SSD. Lingkungan pengembangan yang digunakan juga dikendalikan secara konsisten, meliputi bahasa pemrograman Python, beserta versi *library* yang digunakan seperti *scikit-learn*, *pandas*, *numpy*, dan *matplotlib*. Selain itu, jenis dan ukuran dataset dari Koperasi Sri Hadi Dharma dipertahankan sama selama seluruh proses pelatihan dan pengujian. Prosedur evaluasi model, termasuk metode pembagian data latih dan data uji, juga dijaga tetap agar hasil pengujian dapat dibandingkan secara adil dan bebas dari pengaruh faktor eksternal.

3. Metode Pengumpulan Data

Metode pengumpulan data dalam penelitian ini dilakukan dengan pendekatan dokumentasi, yaitu memperoleh data pinjaman nasabah dalam bentuk file Excel yang disediakan secara langsung oleh pihak Koperasi Sri Hadi Dharma untuk keperluan analisis dan pengembangan model penelitian. Data tersebut berisi variabel-variabel yang relevan untuk proses pengembangan model penilaian risiko kredit, seperti besar pinjaman, tenor, angsuran, dan tunggakan. Seluruh data yang diterima kemudian diolah, dibersihkan, dan disiapkan sebagai dasar untuk proses *clustering* dan prediksi pada tahap pengembangan model.

B. Instrumen Penelitian

Instrumen penelitian merupakan seluruh perangkat, alat bantu, dan sumber data yang digunakan peneliti untuk mengumpulkan informasi, mengolah data, membangun model, serta mengevaluasi kinerja sistem penilaian risiko kredit. Instrumen-instrumen ini memastikan bahwa proses penelitian berlangsung secara sistematis dan dapat dipertanggungjawabkan, baik dari sisi teknis maupun metodologis.

1. Perangkat Keras(Hardware)

Perangkat keras digunakan sebagai media utama untuk menjalankan seluruh proses pengembangan model, mulai dari

preprocessing data, pelatihan algoritma, hingga evaluasi performa. Adapun perangkat keras yang digunakan dalam penelitian ini adalah Laptop dengan spesifikasi:

- a. Processor: AMD Ryzen 3 7320U with Radeon Graphics (2.40 GHz)
- b. RAM: 8 GB
- c. Penyimpanan: SSD 512 GB
- d. Sistem Operasi: Windows 11 64-bit

2. Perangkat Lunak(Software)

Perangkat lunak digunakan untuk mengolah data, membangun model *machine learning*, dan melakukan evaluasi. Perangkat lunak yang digunakan meliputi:

- a. Python 3.12.3 sebagai bahasa pemrograman utama
- b. Library pendukung Python, antara lain:
 - a. Pandas (pengolahan data)
 - b. NumPy (komputasi numerik)
 - c. Scikit-Learn (algoritma K-Means, *Random Forest*, dan evaluasi)
 - d. Matplotlib/Seaborn (visualisasi data)
- c. VS Code sebagai lingkungan pengembangan
- d. Microsoft Excel untuk membuka dan memeriksa dataset awal

3. Data Set

Dataset yang digunakan dalam penelitian ini berupa file Excel yang diperoleh langsung dari Koperasi Sri Hadi Dharma. Dataset tersebut berisi data historis peminjaman seperti besar pinjaman, tenor, angsuran, dan tunggakan yang selanjutnya digunakan untuk proses *preprocessing*, *clustering*, dan prediksi.

4. Analisis Hasil

Analisis hasil dalam penelitian ini dilakukan untuk menilai performa model *hybrid K-Means* dan *Random Forest* dalam memprediksi risiko kredit. Proses analisis meliputi pengolahan data

hasil pelatihan model, penghitungan metrik evaluasi seperti akurasi, *precision*, *recall*, dan *F1-score*, serta interpretasi skor risiko yang dihasilkan. Analisis ini bertujuan memastikan bahwa model bekerja secara konsisten, mampu membedakan tingkat risiko dengan baik, serta layak digunakan sebagai dasar penilaian kredit pada Koperasi Sri Hadi Dharma.

C. Tempat dan Jadwal Penelitian

Penelitian ini dilaksanakan di Koperasi Simpan Pinjam Sri Hadi Dharma yang berlokasi di Jln. Empunala Balongrawe, Kecamatan Magersari. Proses penelitian meliputi pengambilan data, pengembangan model, implementasi, serta evaluasi sistem. Pelaksanaan penelitian direncanakan berlangsung sesuai tahapan yang telah disusun dalam tabel rencana penelitian berikut.

Tabel 3. 1 Rencana Kegiatan Penelitian

No.	Uraian Kegiatan	Bulan 1	Bulan 2	Bulan 3	Bulan 4	Bulan 5	Bulan 6
1	Studi Literatur						
2	Pengumpulan Data						
3	Analisis Sistem						
4	Perancangan Sistem						
5	Desain Sistem						
6	Implementasi						
7	Uji Coba						
8	Debugging						
9	Laporan						

D. Objek Penelitian dan Subjek Penelitian

1. Analisis Kebutuhan Sistem

Analisis kebutuhan sistem dilakukan untuk mengidentifikasi seluruh komponen dan kapabilitas yang diperlukan agar sistem penilaian risiko kredit dapat beroperasi secara efektif dan sesuai dengan tujuan penelitian. Kebutuhan tersebut dibagi menjadi kebutuhan

256 fungsional dan non-fungsional. Secara fungsional, sistem harus mampu menerima masukan berupa variabel-variabel kredit, seperti besar pinjaman, besaran angsuran, dan tenor pinjaman. Selain itu, sistem diharapkan dapat menjalankan proses pra-pemrosesan data (preprocessing), melakukan pengelompokan data menggunakan algoritma *K-Means*, menghasilkan *pseudo-label*, serta melatih model supervised untuk menghasilkan skor risiko kredit. Secara non-fungsional, sistem harus memiliki tingkat keandalan yang baik, mampu menangani data dalam jumlah menengah, serta menyediakan keluaran yang informatif dan mudah dipahami oleh petugas kredit koperasi. Analisis kebutuhan ini menjadi dasar dalam penyusunan arsitektur sistem dan perancangan komponen pendukung lainnya pada tahapan berikutnya

11 2. Objek Penelitian

219 Objek penelitian ini adalah data peminjam pada Koperasi Sri Hadi Dharma yang digunakan sebagai dasar dalam pengembangan model penilaian risiko kredit. Data tersebut mencakup berbagai atribut, antara lain besar pinjaman, angsuran, dan tenor. Data yang digunakan bersifat historis, namun belum memiliki label risiko secara eksplisit (misalnya status lancar atau gagal bayar). Oleh karena itu, penelitian ini menggunakan metode unsupervised *K-Means* untuk menghasilkan *pseudo-label* sebagai representasi kategori risiko awal. *Pseudo-label* tersebut selanjutnya digunakan dalam tahap supervised learning untuk membangun model prediksi risiko kredit berbasis algoritma *Random Forest*.

272 3. Subjek Penelitian

4 Subjek penelitian adalah petugas kredit dan pengelola data Koperasi Sri Hadi Dharma yang terlibat secara langsung dalam proses pengolahan dan validasi data. Peran subjek penelitian meliputi penyediaan data peminjam, verifikasi informasi yang digunakan dalam analisis, serta peninjauan hasil keluaran model prediksi risiko. Subjek

penelitian juga berperan sebagai *evaluator* untuk menilai relevansi, kesesuaian, dan kegunaan skor risiko yang dihasilkan model terhadap praktik penilaian kredit yang dilakukan sehari-hari. Dengan demikian, subjek penelitian memberikan kontribusi penting terhadap keakuratan model serta kesesuaian sistem dengan kebutuhan operasional koperasi.

E. Prosedur Penelitian

1. Studi Literatur

Tahap studi literatur bertujuan untuk mempelajari teori, konsep, dan penelitian terdahulu yang relevan dengan penilaian risiko kredit pada Koperasi Simpan Pinjam menggunakan metode *machine learning*. Kegiatan yang dilakukan meliputi:

- a. Mengumpulkan referensi dari jurnal, buku, artikel ilmiah, dan sumber digital terkait Koperasi Simpan Pinjam, sistem penilaian risiko kredit, metode *K-Means*, *Random Forest*, serta pendekatan *hybrid K-Means + Random Forest*.
- b. Menganalisis penelitian terdahulu untuk mengidentifikasi kelemahan atau keterbatasan metode penilaian risiko yang telah digunakan.
- c. Merumuskan strategi pengembangan model *hybrid* yang diharapkan dapat meningkatkan objektivitas, akurasi, dan konsistensi dalam prediksi risiko kredit nasabah.

2. Pengumpulan Data

Tahap ini bertujuan memperoleh data yang menjadi dasar analisis dan pembangunan sistem penilaian risiko kredit. Kegiatan yang dilakukan:

- a. Mengambil data langsung dari Koperasi Sri Hadi Dharma, termasuk profil nasabah, besar pinjaman, tenor, dan data relevan lainnya.
- b. Membersihkan data dari nilai hilang (*missing value*) atau kesalahan pencatatan.

- c. Menyusun data dalam format digital yang siap digunakan untuk analisis, pengelompokan dengan *K-Means*, dan prediksi risiko menggunakan *Random Forest*.
- d. Menyimpan data secara aman agar mudah diakses pada tahap analisis dan implementasi sistem.

3. Analisis Sistem

Tahap analisis sistem bertujuan memahami karakteristik data dan kebutuhan sistem. Kegiatan utama:

- a. Menganalisis profil nasabah dan histori pinjaman untuk menentukan faktor-faktor risiko utama sebagai fitur input pada model *K-Means* dan *Random Forest*.
- b. Mengidentifikasi kebutuhan fungsional (misal modul input data, pengolahan risiko, output rekomendasi) dan kebutuhan non-fungsional (misal keamanan data, kecepatan proses, kemudahan penggunaan antarmuka).
- c. Membuat diagram alir data (DFD) dan *Entity Relationship Diagram* (ERD) untuk memvisualisasikan alur data dan hubungan antar tabel.
- d. Menentukan metode pemodelan *hybrid*, yaitu *K-Means* untuk pengelompokan risiko awal dan *Random Forest* untuk prediksi risiko kredit akhir.

4. Perancangan Sistem

Tahap perancangan sistem dilakukan untuk menyusun konsep keseluruhan sistem. Kegiatan yang dilakukan:

- a. Merancang arsitektur sistem mulai dari input data nasabah, *preprocessing*, pengelompokan awal dengan *K-Means*, hingga prediksi akhir menggunakan *Random Forest*.
- b. Menentukan fitur data penting yang akan digunakan sebagai input model.
- c. Menentukan spesifikasi teknis algoritma, termasuk jumlah *cluster K-Means* serta parameter *Random Forest* seperti jumlah pohon (*n_estimators*) dan kedalaman pohon maksimum (*max_depth*).

17

5. Desain Sistem

Tahap desain sistem berfokus pada penyusunan rincian teknis implementasi. Kegiatan yang dilakukan:

- a. Mendesain struktur database nasabah, histori pinjaman, dan hasil analisis risiko.
- b. Membuat *flowchart* alur proses tiap modul sistem.
- c. Menyusun prototipe antarmuka pengguna agar mudah digunakan untuk input data dan menampilkan hasil prediksi risiko.
- d. Menyusun algoritma *hybrid* secara rinci, dengan *K-Means* untuk pengelompokan nasabah berdasarkan profil risiko awal dan *Random Forest* untuk prediksi risiko kredit akhir.

6. Implementasi

Tahap implementasi bertujuan membangun sistem sesuai desain. Kegiatan yang dilakukan:

- a. Mengembangkan modul input data, pengolahan, dan output prediksi risiko.
- b. Mengimplementasikan algoritma *hybrid K-Means + Random Forest* sebagai engine prediksi.
- c. Mengintegrasikan sistem dengan database untuk memastikan aliran data berjalan lancar.
- d. Melakukan uji modul parsial untuk memastikan setiap bagian berfungsi sesuai spesifikasi.

7. Uji Coba

Tahap uji coba bertujuan menilai performa sistem secara keseluruhan. Kegiatan yang dilakukan:

- a. Menjalankan sistem dengan data uji.
- b. Melakukan pengujian berulang (*multi-run*) untuk memperoleh hasil yang stabil.
- c. Menghitung metrik evaluasi, seperti akurasi, presisi, recall, dan F1-score.

153

- d. Mengidentifikasi kesalahan atau masalah yang muncul selama pengujian untuk diperbaiki pada tahap berikutnya.

8. Debugging

Tahap *debugging* bertujuan menyempurnakan sistem agar siap digunakan. Kegiatan:

- a. Memperbaiki seluruh bug atau kesalahan yang ditemukan selama uji coba.
- b. Mengoptimalkan kinerja sistem agar proses prediksi risiko berjalan cepat dan akurat.
- c. Memastikan seluruh fitur sistem berfungsi sesuai desain dan kebutuhan pengguna.

9. Penyusunan Laporan

Tahap terakhir adalah menyusun laporan penelitian secara lengkap. Kegiatan yang dilakukan:

- a. Mendokumentasikan seluruh proses penelitian, mulai dari pengumpulan data, analisis, implementasi, hingga uji coba.
- b. Menyajikan hasil pengujian dalam bentuk tabel, grafik, dan narasi deskriptif.
- c. Merevisi laporan berdasarkan masukan pembimbing hingga siap digunakan untuk seminar hasil dan ujian skripsi.

F. Teknik Analisis Data

1. Desain Sistem

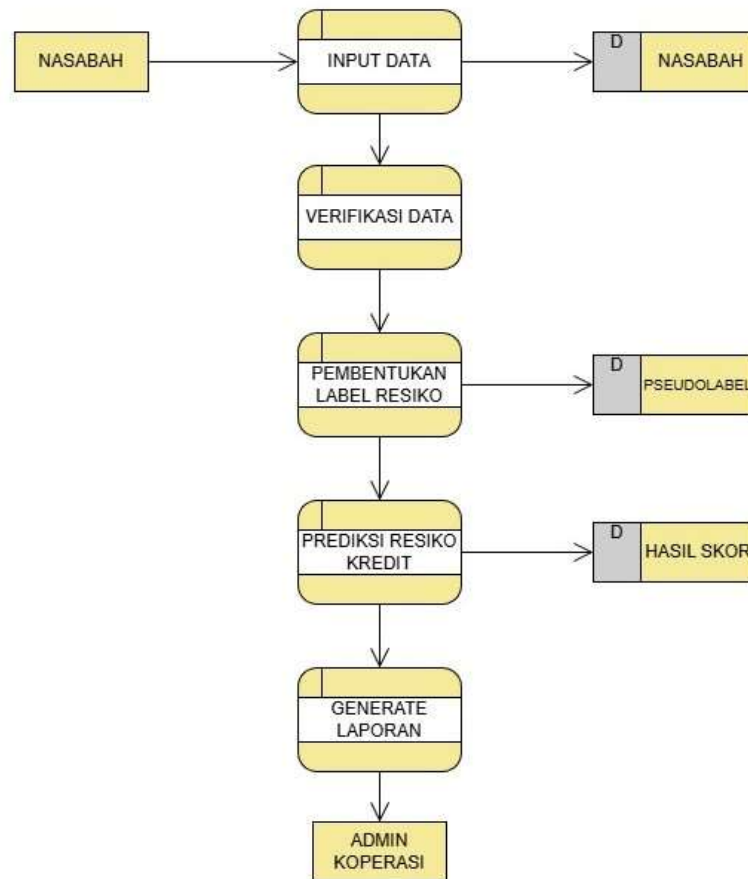
a. Arsitektur Sistem (*System Architecture*)

Arsitektur sistem pada penelitian ini dirancang untuk menggambarkan struktur dan alur kerja sistem penilaian risiko kredit berbasis *machine learning* secara terstruktur. Pendekatan yang digunakan adalah *Structured Systems Analysis and Design (SSAD)*, yang menekankan pada pemodelan proses dan aliran data di dalam sistem. Arsitektur sistem ini bertujuan untuk memperjelas hubungan antara entitas eksternal, proses utama, serta data yang diolah dan dihasilkan oleh sistem.



Gambar 3. 1 DFD level 0

Pada Gambar 3. 1 sistem penilaian risiko kredit digambarkan sebagai satu proses utama yang berinteraksi dengan dua entitas eksternal, yaitu nasabah dan admin koperasi. Nasabah memberikan data peminjaman sebagai input ke dalam sistem, sedangkan admin koperasi menerima keluaran berupa hasil penilaian risiko kredit. Diagram ini menunjukkan batasan sistem serta aliran data masuk dan keluar tanpa menampilkan detail proses internal. Diagram ini digunakan untuk memberikan gambaran umum sistem serta menetapkan batasan interaksi antara sistem penilaian risiko kredit dengan pihak eksternal sebelum proses internal dijabarkan lebih rinci.

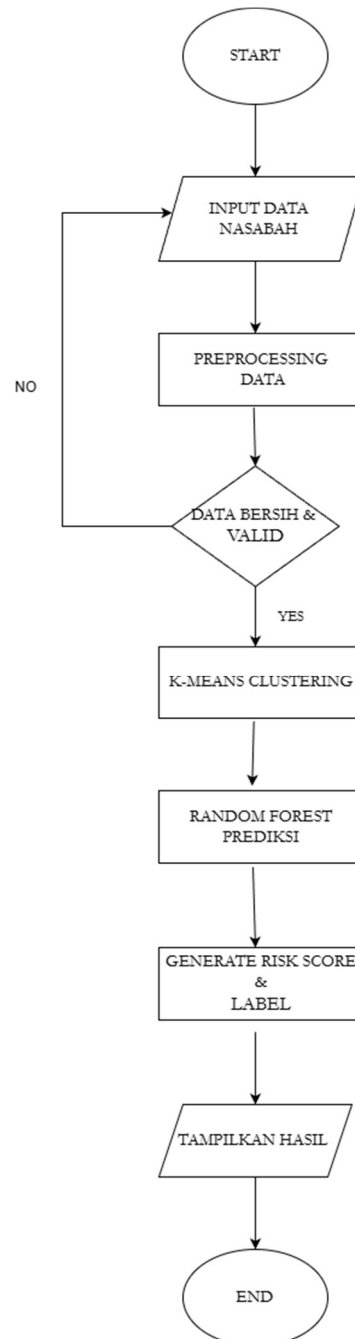


Gambar 3. 2 DFD level 1

Pada Gambar 3. 2 proses utama pada diagram konteks diuraikan menjadi beberapa sub-proses yang saling terhubung. Proses dimulai dari input data nasabah, kemudian dilakukan verifikasi data untuk memastikan kelengkapan dan keakuratan informasi. Setelah data dinyatakan valid, sistem melakukan pembentukan label risiko awal atau *pseudo-label* menggunakan metode *clustering K-Means*. *Pseudo-label* yang dihasilkan pada tahap ini merupakan hasil antara (*intermediate output*) yang berfungsi sebagai representasi awal tingkat risiko, dan bukan merupakan label kebenaran, sehingga digunakan sebagai dasar pembelajaran pada tahap prediksi selanjutnya, yaitu prediksi risiko kredit menggunakan algoritma

4

Random Forest. Hasil prediksi berupa skor risiko kredit kemudian disimpan dan digunakan dalam proses pembuatan laporan yang akan diakses oleh admin koperasi.



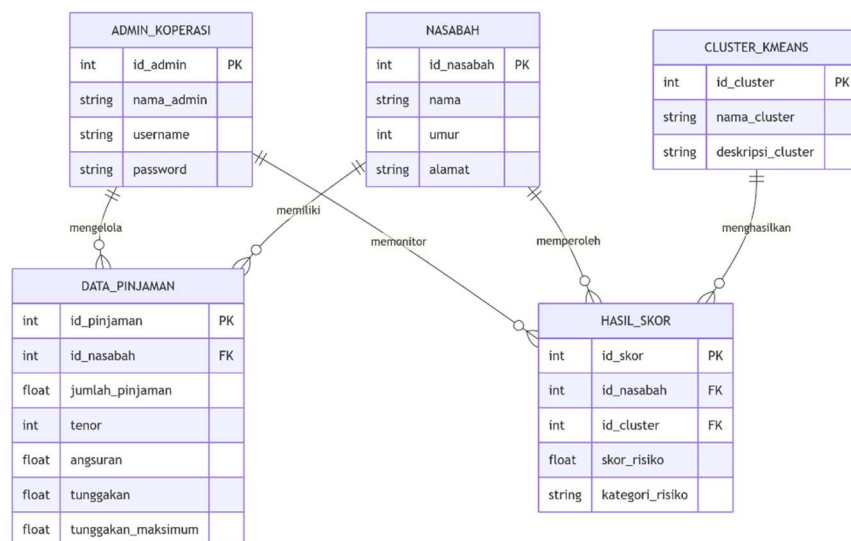
Gambar 3. 3 Flowchart

Flowchart ini disusun berdasarkan proses-proses yang telah dimodelkan pada DFD Level 1, sehingga menggambarkan

implementasi logis dari alur pengolahan data sistem secara lebih rinci dan berurutan. Proses dimulai dari input data nasabah, dilanjutkan dengan tahap *preprocessing* untuk membersihkan dan menyiapkan data. Setelah data dinyatakan bersih dan valid, sistem menjalankan proses *clustering K-Means* untuk membentuk kelompok risiko awal. Hasil *clustering* tersebut kemudian digunakan sebagai dasar dalam proses prediksi risiko kredit menggunakan algoritma *Random Forest* untuk menghasilkan skor dan label risiko. Hasil akhir selanjutnya ditampilkan kepada pengguna sebagai keluaran sistem.

b. Desain Database (*Database Design*)

Perancangan database pada penelitian ini dilakukan untuk mendukung penyimpanan dan pengelolaan data dalam sistem penilaian risiko kredit secara terstruktur dan terintegrasi. Desain database disusun menggunakan *Entity Relationship Diagram* (ERD) yang berfungsi untuk memodelkan entitas, atribut utama, serta hubungan antar entitas yang terlibat dalam sistem. Perancangan ini bertujuan untuk memastikan bahwa data yang digunakan dalam proses analisis risiko kredit dapat dikelola secara konsisten dan mendukung kebutuhan pemodelan *machine learning*.



Gambar 3. 4 ERD

204

Berdasarkan Gambar 3.4, sistem penilaian risiko kredit terdiri dari beberapa entitas utama yang saling berhubungan dalam mendukung proses pengolahan data dan penilaian risiko kredit.

1) Entitas Nasabah

Entitas Nasabah merepresentasikan individu yang mengajukan pinjaman pada koperasi. Entitas ini menyimpan informasi identitas dan karakteristik dasar peminjam yang digunakan sebagai variabel input dalam proses pengelompokan dan penilaian risiko kredit.

2) Entitas Data Pinjaman

Entitas Data Pinjaman menyimpan informasi terkait pengajuan pinjaman yang dilakukan oleh nasabah. Data pada entitas ini merepresentasikan kondisi finansial pinjaman dan menjadi salah satu faktor utama dalam proses analisis risiko kredit.

3) Entitas Cluster K-Means

Entitas *Cluster K-Means* merepresentasikan hasil pengelompokan awal nasabah berdasarkan kemiripan karakteristik data menggunakan metode *K-Means*. *Cluster* ini berfungsi sebagai *pseudo-label* yang digunakan sebagai dasar pembelajaran pada tahap prediksi risiko kredit.

4) Entitas Hasil Skor

Entitas Hasil Skor menyimpan hasil akhir penilaian risiko kredit yang dihasilkan oleh sistem. Informasi pada entitas ini mencerminkan tingkat risiko kredit nasabah dan digunakan sebagai dasar dalam pengambilan keputusan oleh pihak koperasi.

5) Entitas Admin Koperasi

Entitas Admin Koperasi merepresentasikan pengguna sistem yang memiliki kewenangan dalam mengelola data

11

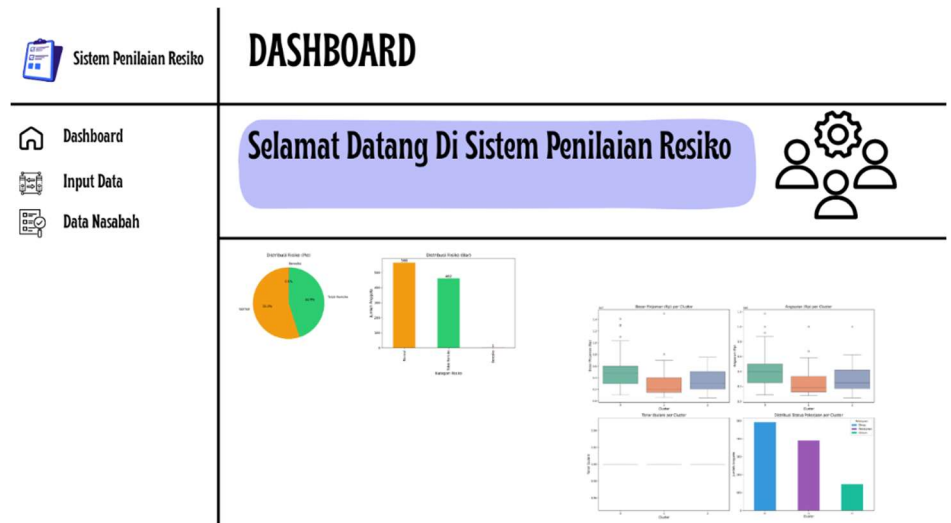
pinjaman dan memantau hasil penilaian risiko kredit yang dihasilkan oleh sistem.

Entitas-entitas tersebut dirancang untuk merepresentasikan data nasabah, data pinjaman, hasil pengelompokan risiko, serta hasil penilaian risiko kredit yang dihasilkan oleh sistem.

c. Desain Antarmuka Pengguna (UI/UX Design)

Desain antarmuka pengguna (UI) dan pengalaman pengguna (UX) pada sistem penilaian risiko kredit dirancang untuk memudahkan petugas koperasi dalam melakukan input data nasabah, memantau distribusi risiko, serta memperoleh hasil penilaian risiko kredit secara cepat, jelas, dan informatif. Antarmuka sistem dikembangkan dengan prinsip kesederhanaan, konsistensi, dan kemudahan navigasi, sehingga dapat digunakan oleh pengguna non-teknis.

1) Dashboard



Gambar 3. 5 Dashboard

Pada Gambar 3.5 ditampilkan halaman Dashboard, yang merupakan halaman utama ketika pengguna mengakses sistem. Dashboard berfungsi sebagai pusat informasi ringkas terkait kondisi risiko kredit nasabah koperasi secara keseluruhan.

Pada bagian atas halaman ditampilkan judul “*Selamat Datang di Sistem Penilaian Risiko*” sebagai penanda konteks sistem yang sedang digunakan. Menu navigasi berada di sisi kiri halaman, yang terdiri dari tiga menu utama, yaitu Dashboard, Input Data, dan Data Nasabah, sehingga pengguna dapat berpindah halaman dengan mudah.

Dashboard menyajikan beberapa visualisasi data, antara lain:

- a) Diagram pie distribusi risiko, yang menunjukkan proporsi nasabah dalam kategori Berisiko Rendah, Sedang dan Berisiko Tinggi.
- b) Diagram batang jumlah nasabah per kategori risiko, untuk memberikan gambaran kuantitatif.
- c) Tabel contoh nasabah dengan kategori .

Visualisasi ini membantu pengambil keputusan dalam memahami pola risiko kredit secara cepat tanpa harus melihat data mentah satu .

2) Input Data Nasabah

The screenshot shows a web application interface for 'Sistem Penilaian Risiko'. On the left is a sidebar with a home icon and three menu items: 'Dashboard', 'Input Data', and 'Data Nasabah'. The main content area is titled 'Input Data Pinjaman'. It contains several input fields: 'Besar Pinjaman :', 'Sisa Pinjaman :', 'Tenor :', 'Tunggakan :', and 'Tunggakan Maksimum:'. Below these fields are two more fields: 'Skor resiko :' and 'Kategori Resiko :'. A blue button labeled 'Prediksi' is located at the bottom right of the form area.

Gambar 3. 6 Input Data

4

32

Pada Gambar 3.6 ditunjukkan halaman Input Data Pinjaman, yang digunakan untuk memasukkan data pinjaman nasabah. Halaman ini dirancang dengan struktur formulir yang sederhana dan berurutan agar meminimalkan kesalahan input.

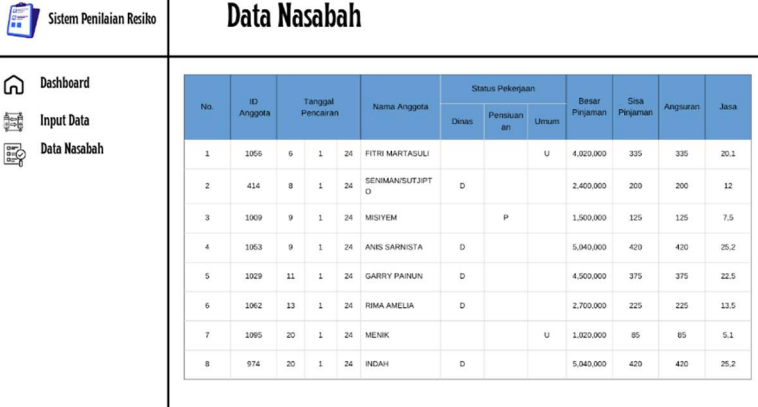
Formulir data pinjaman meliputi :

1. Besar pinjaman
2. Sisa Pinjaman
3. Tenor
4. Tunggalan
5. Tunggalan Maksimum

Setiap bagian dilengkapi dengan tombol Prediksi Risiko untuk menjalankan proses analisis dan penilaian risiko kredit berdasarkan data nasabah yang telah dimasukkan. Desain ini bertujuan untuk meningkatkan kenyamanan pengguna serta memastikan data yang dimasukkan lengkap dan konsisten sebelum digunakan dalam proses clustering K-Means dan prediksi Random Forest.

294

3) Data Nasabah



No	ID Anggota	Tanggal Pencarian	Nama Anggota	Status Pelanggaran			Besar Pinjaman	Sisa Pinjaman	Angsuran	Jasa
				Dinas	Pemakaian	Utang				
1	1055	9 1 24	FITRI MARTASULI			U	4,020,000	335	335	20.1
2	414	8 1 24	SENIMAN/SUTJPT O	D			2,400,000	200	200	12
3	1009	9 1 24	MISIYEM		P		1,500,000	125	125	7.5
4	1053	9 1 24	ANIS SARKISTA	D			5,040,000	420	420	25.2
5	1029	11 1 24	GARRY PAMUN	D			4,500,000	375	375	22.5
6	1062	13 1 24	RIMA AMELIA	D			2,700,000	225	225	13.5
7	1095	20 1 24	MENIK			U	1,020,000	85	85	5.1
8	974	20 1 24	INDAH	D			5,040,000	420	420	25.2

Gambar 3. 7 Data Nasabah

4

206

Pada Gambar 3.7 diperlihatkan halaman Data Nasabah, yang digunakan untuk menampilkan seluruh data nasabah yang tersimpan di dalam sistem. Halaman ini berfungsi sebagai media

utama bagi petugas koperasi untuk melihat, memeriksa, dan mengelola data nasabah yang digunakan dalam proses penilaian risiko kredit.

Pada halaman ini ditampilkan tabel data nasabah yang berisi beberapa informasi penting, seperti:

1. ID nasabah
2. Nama nasabah
3. Besar pinjaman
4. Besar angsuran
5. Tenor
6. Tunggalan
7. Tunggalan maksimum

Selain menampilkan data nasabah, halaman ini juga membantu pengguna dalam melakukan pengecekan data sebelum proses *clustering K-Means* dan prediksi *Random Forest* dilakukan. Penyajian data dalam bentuk tabel memudahkan pengguna dalam melakukan pencarian, verifikasi, dan evaluasi data secara lebih terstruktur. Dengan adanya halaman Data Nasabah, proses pengelolaan data menjadi lebih efektif, konsisten, dan mendukung proses analisis risiko kredit pada koperasi.

2. Simulasi Proses Penyelesaian Masalah

a. Data Input

Data yang digunakan dalam penelitian ini berupa data nasabah koperasi yang mengajukan pinjaman. Data disusun dalam bentuk tabel dan terdiri dari atribut numerik yang mencerminkan karakteristik finansial nasabah dalam proses penilaian risiko kredit. Atribut-atribut tersebut meliputi:

- 1) Besar_Pinjaman (numerik) – Total pinjaman
- 2) Sisa_Pinjaman (numerik) – Sisa pinjaman yang harus dibayar
- 3) Angsuran (numerik) – Angsuran per bulan

- 4) Tenor (numerik) – Lama pinjaman (bulan)
- 5) Tunggakan (numerik) – Jumlah tunggakan pembayaran nasabah

Data ini digunakan untuk membentuk label risiko awal dengan K-Means ($k=3$), kemudian digunakan sebagai target pelatihan *Random Forest*.

Tabel 3. 2 Contoh Data

Besar pinjaman	Sisa Pinjaman	Angsuran Pokok	Tunggakan	Tenor
2,400,000	200,000	200,000	270,000	1
1,500,000	250,000	125,000	270,000	2
1,500,000	375,000	125,000	148,500	3
3,000,000	1,000,000	250,000	1,080,000	4
2,040,000	850,000	170,000	1,953,900	5

Tabel 3.2 hanya menampilkan sebagian kecil data nasabah yang digunakan untuk simulasi. Data ini merepresentasikan variasi kondisi finansial nasabah, seperti besar pinjaman, sisa pinjaman, angsuran, jasa, tunggakan, dan tenor yang digunakan sebagai faktor dalam penentuan risiko kredit

Preprocessing Data

Preprocessing data dilakukan untuk mempersiapkan data historis pinjaman anggota sebelum digunakan pada proses analisis menggunakan algoritma *K-Means* dan *Random Forest*. Tahapan preprocessing yang dilakukan adalah sebagai berikut:

a) Pembersihan Data (*Data Cleaning*)

Tahap ini dilakukan untuk memastikan kualitas data yang digunakan dalam penelitian. Proses pembersihan data meliputi pengecekan data kosong, data duplikat, serta penyesuaian format data sehingga diperoleh data yang siap digunakan pada proses analisis.

b) Pembentukan Atribut Tunggakan Maksimum

Tahap ini dilakukan menggunakan teknik *feature engineering* dengan membentuk atribut baru berupa Tunggakan Maksimum.

Atribut ini diperoleh dari nilai tunggakan tertinggi yang dimiliki setiap anggota berdasarkan data historis pembayaran pinjaman. Penambahan atribut ini bertujuan untuk memberikan representasi risiko kredit yang lebih informatif.

c) Seleksi Atribut (*Feature Selection*)

Tahap ini dilakukan dengan memilih atribut yang digunakan dalam proses analisis, yaitu Besar Pinjaman, Sisa Pinjaman, Angsuran Pokok, Tenor, dan Tunggakan Maksimum. Pemilihan atribut dilakukan karena variabel tersebut memiliki keterkaitan dengan karakteristik pembayaran pinjaman anggota.

d) Normalisasi Data (*Min-Max Normalization*)

Tahap terakhir dilakukan dengan mengubah nilai setiap atribut ke dalam rentang 0–1 menggunakan metode *Min-Max Normalization* sebagaimana ditunjukkan pada Persamaan (2.2). Normalisasi bertujuan untuk menghindari dominasi atribut yang memiliki nilai besar sehingga setiap atribut memiliki kontribusi yang seimbang pada proses *clustering* menggunakan algoritma *K-Means* maupun pada pembentukan model *Random Forest*.

Contoh perhitungan normalisasi data pertama adalah sebagai berikut.

Contoh Perhitungan Normalisasi Data Pertama

1. Besar pinjaman

$$X_{norm} = \frac{2.400.000 - 1.500.000}{3.000.000 - 1.500.000} = \frac{900.000}{1.500.000} = 0.6$$

2. Sisa Pinjaman

$$X_{norm} = \frac{200.000 - 200.000}{1.000.000 - 200.000} = 0$$

3. Angsuran

$$X_{norm} = \frac{200.000 - 125.000}{250.000 - 125.000} = \frac{75.000}{125.000} = 0.6$$

4. Tenor

$$X_{norm} = \frac{1 - 1}{5 - 1} = 0$$

5. Tunggakan Maksimum

$$X_{norm} = \frac{270.000 - 148.500}{1.953.900 - 148.500} = \frac{121.500}{1.805.400} \approx 0.06$$

Tabel 3. 3 Normalisasi

No.	Besar pinjaman	Sisa Pinjaman	Angsuran	Tenor	Tunggakan Maksimum
1	0.600	0.000	0.600	0.000	0.067
2	0.000	0.063	0.000	0.250	0.067
3	0.000	0.219	0.000	0.500	0.000
4	1.000	1.000	1.000	0.750	0.517
5	0.360	0.813	0.360	1.000	1.000

Pada Tabel 3.3, menunjukkan hasil akhir *preprocessing* data yang telah melalui tahap pembentukan atribut Tunggakan Maksimum, seleksi atribut, dan normalisasi data. Seluruh atribut telah berada pada rentang nilai 0–1 sehingga memiliki kontribusi yang seimbang pada proses pengelompokan menggunakan algoritma *K-Means* dan pembentukan model prediksi menggunakan *Random Forest*.

b. Proses Pengolahan Data

1) Inisialisasi Centroid

Pada tahap awal algoritma K-Means, ditentukan jumlah *cluster* sebanyak $k = 3$ yang merepresentasikan tiga kategori risiko kredit, yaitu berisiko rendah, berisiko sedang, dan berisiko tinggi. *Centroid* awal dipilih secara acak dari data nasabah, yaitu data ke-1, ke-3, dan ke-4.

Tabel 3. 4 Centroid

C	Besar pinjaman	Sisa Pinjaman	Angsuran	Tenor	Tunggakan Maksimum
C1	0.600	0.000	0.600	0.000	0.067
C2	0.000	0.219	0.000	0.500	0.000
C3	1.000	1.000	1.000	0.750	0.517

2) Perhitungan Jarak *Euclidean*

Jarak antara setiap data dengan *centroid* dihitung menggunakan rumus *Euclidean Distance* sebagaimana ditunjukkan pada Persamaan (2.1).

Contoh Perhitungan Jarak Data ke-2 terhadap Semua Centroid

Jarak ke C1 (Risiko Rendah)

$$d = \sqrt{(0.000 - 0.600)^2 + (0.063 - 0.000)^2 + (0.000 - 0.600)^2 + (0.063 - 0.000)^2 + (0.067 - 0.067)^2}$$

$$d = \sqrt{0.36 + 0.0039 + 0.36 + 0.0039 + 0} = \sqrt{0.7278} \approx 0.853$$

Jarak ke C2 (Risiko Sedang)

$$d = \sqrt{(0.000 - 0.000)^2 + (0.063 - 0.219)^2 + (0.000 - 0.000)^2 + (0.063 - 0.500)^2 + (0.067 - 0.000)^2}$$

$$d = \sqrt{0 + 0.0243 + 0 + 0.1909 + 0.0045} = \sqrt{0.2197} \approx 0.469$$

Jarak ke C3 (Risiko Tinggi)

$$d = \sqrt{(0.000 - 1.000)^2 + (0.063 - 1.000)^2 + (0.000 - 1.000)^2 + (0.063 - 0.750)^2 + (0.067 - 0.517)^2}$$

$$d = \sqrt{1 + 0.878 + 1 + 0.472 + 0.203} = \sqrt{3.553} \approx 1.885$$

Berdasarkan hasil perhitungan tersebut, jarak terkecil diperoleh pada *Cluster* C2 sehingga data ke-2 ditempatkan pada *Cluster* C2 yang merepresentasikan kategori Risiko Sedang.

3) Penetapan Semua nasabah ke *Cluster* Terdekat

Setelah proses perhitungan jarak *Euclidean* dilakukan, setiap data nasabah ditempatkan ke dalam *cluster* yang memiliki jarak paling kecil terhadap *centroid*. Hasil pengelompokan awal menunjukkan bahwa data dengan nilai Tunggalan Maksimum yang tinggi cenderung berada pada cluster risiko tinggi, sedangkan data dengan nilai Tunggalan Maksimum yang lebih rendah cenderung berada pada cluster risiko rendah atau sedang. Hasil penetapan *cluster* awal dapat dilihat pada Tabel 3.5 berikut.

Tabel 3. 5 Penetapan Nasabah

Data	Cluster	Kategori Risiko
Data 1	C1	Risiko Rendah
Data 2	C2	Risiko Sedang
Data 3	C2	Risiko Sedang
Data 4	C3	Risiko Tinggi
Data 5	C3	Risiko Tinggi

Pada Tabel 3.5 terlihat bahwa data nasabah berhasil terbagi ke dalam tiga kelompok risiko berdasarkan tingkat kemiripan karakteristik finansialnya. *Cluster* C1 merepresentasikan risiko rendah karena memiliki nilai Tunggalan Maksimum yang relatif rendah. *Cluster* C2 merepresentasikan risiko sedang karena memiliki nilai Tunggalan Maksimum pada tingkat menengah. Sementara itu, *cluster* C3 merepresentasikan risiko tinggi karena memiliki nilai Tunggalan Maksimum tertinggi dibandingkan *cluster* lainnya. Hasil pengelompokan ini menjadi dasar dalam proses pembaruan *centroid* pada iterasi berikutnya hingga diperoleh posisi *centroid* yang stabil dan tidak terjadi perpindahan data antar *cluster*.

4) Pembaruan *Centroid*

90

154

Setelah seluruh data nasabah berhasil ditempatkan ke dalam *cluster* masing-masing, langkah berikutnya adalah melakukan pembaruan *centroid*. Proses ini dilakukan dengan menghitung rata-rata setiap atribut pada data yang berada dalam *cluster* yang sama.

Tujuan pembaruan *centroid* adalah memperoleh titik pusat *cluster* baru yang lebih merepresentasikan karakteristik data pada masing-masing kelompok. Nilai *centroid* baru ini nantinya digunakan kembali pada proses perhitungan jarak di iterasi berikutnya.

91

303

Berdasarkan hasil penetapan cluster sebelumnya, cluster C2 terdiri dari data ke-2 dan data ke-3, sedangkan cluster C3 terdiri dari data ke-4 dan data ke-5. Sementara itu, cluster C1 hanya berisi data ke-1 sehingga nilai centroid C1 tetap sama. Sebagai contoh, perhitungan centroid baru untuk cluster C2 dilakukan menggunakan rata-rata dari data ke-2 dan data ke-3 sebagai berikut.

91

Besar pinjaman

$$\frac{0.000 + 0.000}{2} = 0.000$$

Sisa Pinjaman

$$\frac{0.063 + 0.219}{2} = 0.141$$

Angsuran

$$\frac{0.000 + 0.000}{2} = 0.000$$

Tenor

$$\frac{0.250 + 0.500}{2} = 0.375$$

Tunggakan Maksimum

$$\frac{0.067 + 0.000}{2} = 0.034$$

Selanjutnya, proses pembaruan *centroid* dilakukan pada seluruh *cluster* hingga diperoleh *centroid* baru yang lebih representatif. Proses iterasi akan terus dilakukan sampai tidak terjadi perpindahan data antar *cluster* atau nilai *centroid* telah stabil.

5) Penentuan Label Risiko

Setelah proses iterasi *K-Means* selesai dan nilai *centroid* telah stabil, setiap *cluster* diberikan label risiko berdasarkan karakteristik nilai Tunggakan Maksimum. *Cluster* dengan nilai Tunggakan Maksimum terendah dikategorikan sebagai Risiko Rendah, *cluster* dengan nilai Tunggakan Maksimum menengah dikategorikan sebagai Risiko Sedang, sedangkan *cluster* dengan nilai Tunggakan Maksimum tertinggi dikategorikan sebagai Risiko Tinggi.

1. *Cluster* C1 (Risiko Rendah)

Cluster C1 merepresentasikan kelompok data dengan nilai Tunggakan Maksimum yang relatif rendah sehingga memiliki tingkat risiko kredit yang rendah. Nasabah pada *cluster* ini cenderung memiliki kemampuan pembayaran yang lebih baik dibandingkan *cluster* lainnya.

2. *Cluster* C2 (Risiko Sedang)

Cluster C2 merepresentasikan kelompok data dengan nilai Tunggakan Maksimum pada tingkat menengah sehingga dikategorikan sebagai kelompok risiko sedang.

3. *Cluster* C3 (Risiko Tinggi)

Cluster C3 merepresentasikan kelompok data dengan nilai Tunggakan Maksimum tertinggi sehingga memiliki

potensi risiko kredit yang lebih besar dibandingkan cluster lainnya.

Label risiko hasil *clustering K-Means* ini selanjutnya digunakan sebagai target kelas pada proses prediksi menggunakan algoritma *Random Forest*.

6) Prediksi Risiko Menggunakan Random Forest

Setelah proses *clustering* menggunakan algoritma *K-Means* selesai dilakukan, tahap berikutnya adalah melakukan prediksi risiko kredit menggunakan algoritma *Random Forest*. Pada tahap ini, label risiko hasil *clustering K-Means* digunakan sebagai target kelas dalam proses pelatihan model *Random Forest*.

Algoritma *Random Forest* bekerja dengan membangun sejumlah pohon keputusan (*decision tree*) menggunakan teknik *bootstrap sampling*, yaitu pengambilan sampel data secara acak dari *dataset pelatihan*. Setiap pohon keputusan dibangun menggunakan subset data dan subset fitur yang berbeda sehingga menghasilkan model yang lebih stabil dan mampu mengurangi risiko *overfitting*.

Pada setiap node pohon keputusan, algoritma akan memilih fitur terbaik berdasarkan nilai *Gini Impurity* untuk menentukan pemisahan data yang paling optimal. Proses ini dilakukan secara berulang hingga terbentuk sejumlah pohon keputusan yang digunakan untuk melakukan prediksi prediksi risiko kredit.

Prediksi Data Uji Menggunakan *Random Forest*

Sebagai ilustrasi proses prediksi, digunakan data ke-2 sebagai data uji dengan hasil normalisasi sebagai berikut:

Besar pinjaman : 0.000

Sisa Pinjaman : 0.063

Angsuran : 0.000

Tenor : 0.250

Tunggakan Maksimum: 0.067

Data tersebut kemudian dimasukkan ke dalam model Random Forest yang telah dilatih menggunakan label risiko hasil K-Means.

Setiap pohon keputusan pada *Random Forest* melakukan prediksi secara independen terhadap data uji. Hasil prediksi dari masing-masing pohon keputusan ditunjukkan pada Tabel 3.6.

Tabel 3. 6 Pohon Prediksi

Pohon	Prediksi
1	Risiko Sedang
2	Risiko Sedang
3	Risiko Rendah
4	Risiko Sedang
5	Risiko Sedang

Berdasarkan hasil prediksi pada Tabel 3.6, diperoleh voting mayoritas sebanyak 4 dari 5 pohon keputusan yang memprediksi Risiko Sedang. Oleh karena itu, hasil akhir prediksi *Random Forest* untuk data uji tersebut adalah Risiko Sedang.

7) Probabilitas Prediksi

Selain menghasilkan label prediksi, algoritma *Random Forest* juga mampu memberikan informasi probabilitas untuk setiap kelas risiko. Nilai probabilitas diperoleh berdasarkan proporsi jumlah pohon keputusan yang memilih masing-masing kelas dibandingkan dengan jumlah keseluruhan pohon pada model Random Forest.

Sebagai contoh, pada data uji diperoleh 4 dari 5 pohon keputusan memprediksi Risiko Sedang sehingga probabilitas prediksi Risiko Sedang dapat dihitung sebagai berikut:

$$\frac{4}{5} \times 100\% = 80\%$$

Berdasarkan hasil tersebut, model Random Forest memberikan probabilitas sebesar 80% terhadap kategori Risiko Sedang.

Informasi probabilitas ini dapat membantu pihak koperasi dalam memahami tingkat keyakinan model terhadap hasil prediksi yang diberikan sehingga proses pengambilan keputusan kredit dapat dilakukan secara lebih objektif dan terukur.

8) Simulasi Random Forest

Secara ringkas, tahapan simulasi *Random Forest* dalam penelitian ini adalah sebagai berikut:

- a) Label risiko hasil *K-Means* digunakan sebagai target kelas pada proses pelatihan *Random Forest*.
- b) Setiap pohon keputusan dibangun menggunakan *bootstrap sampling* dari data nasabah koperasi.
- c) Pada setiap *node* pohon, *subset* fitur dipilih secara acak dan *split terbaik* ditentukan berdasarkan nilai *Gini Impurity*.
- d) Data nasabah uji dimasukkan ke dalam seluruh pohon keputusan, dan prediksi dari setiap pohon dicatat.
- e) Label akhir ditentukan menggunakan mekanisme *majority voting*, sedangkan probabilitas prediksi dihitung berdasarkan proporsi voting masing-masing kelas.

Dengan kombinasi *K-Means* dan *Random Forest*, sistem mampu menentukan risiko kredit secara objektif serta menyediakan indikator probabilitas sebagai dasar pengambilan keputusan.

c. Data Output dan Evaluasi

Setelah proses iterasi *K-Means* selesai dan nilai *centroid* mencapai kondisi stabil, setiap data nasabah memperoleh label risiko sesuai dengan *cluster* yang terbentuk. Label risiko tersebut kemudian digunakan sebagai target kelas pada proses prediksi menggunakan algoritma *Random Forest*.

Hasil akhir pengelompokan data menggunakan algoritma *K-Means* ditunjukkan pada Tabel 3.7.

Tabel 3. 7 Data Output – Label Risiko K-Means

No.	Data	Cluster K-Means	Label Risiko
1	Data 1	C1	Risiko Rendah
2	Data 2	C2	Risiko Sedang
3	Data 3	C2	Risiko Sedang
4	Data 4	C3	Risiko Tinggi
5	Data 5	C3	Risiko Tinggi

Berdasarkan hasil pengelompokan pada Tabel 3.7, cluster C1 merepresentasikan kelompok data dengan tingkat risiko rendah karena memiliki nilai pinjaman, jasa, dan tunggakan maksimum yang relatif kecil. Cluster C2 merepresentasikan risiko sedang karena memiliki kondisi finansial menengah dan tingkat tunggakan maksimum yang masih dalam batas wajar. Sementara itu, cluster C3 merepresentasikan risiko tinggi karena terdiri dari data dengan nilai pinjaman, sisa pinjaman, jasa, dan tunggakan maksimum yang lebih besar dibandingkan cluster lainnya.

1) Hasil Prediksi *Random Forest*

Sebelum proses pelatihan *Random Forest* dilakukan, dataset dibagi menjadi data pelatihan sebesar 80% dan data pengujian sebesar 20% menggunakan metode *stratified sampling* agar proporsi masing-masing kategori risiko tetap terjaga pada kedua subset data. Pembagian ini bertujuan agar proses pelatihan dan evaluasi model dapat dilakukan secara lebih representatif.

Hasil prediksi *Random Forest* untuk seluruh nasabah koperasi ditunjukkan pada Tabel 3.8 Berdasarkan hasil tersebut, dapat dilihat bahwa prediksi *Random Forest* konsisten dengan label risiko yang dihasilkan oleh K-Means. Konsistensi ini menunjukkan bahwa *Random Forest* berhasil mempelajari pola risiko kredit yang terbentuk dari proses clustering sebelumnya.

Tabel 3. 8 Data Output Prediksi Random Forest

No.	Data	Label K-Means	Prediksi RF	Probabilitas RF
1	Data 1	Risiko Rendah	Risiko Rendah	100%
2	Data 2	Risiko Sedang	Risiko Sedang	80%
3	Data 3	Risiko Sedang	Risiko Sedang	100%
4	Data 4	Risiko Tinggi	Risiko Tinggi	100%
5	Data 5	Risiko Tinggi	Risiko Tinggi	100%

Berdasarkan hasil pada Tabel 3.8, prediksi yang dihasilkan oleh Random Forest sesuai dengan label risiko hasil *clustering K-Means*. Hal ini menunjukkan bahwa *Random Forest* mampu mempelajari pola pengelompokan risiko kredit yang terbentuk dari proses *clustering* sebelumnya.

Nilai probabilitas pada hasil prediksi menunjukkan tingkat keyakinan model terhadap kelas yang dipilih. Sebagai contoh, pada Data 2 diperoleh probabilitas sebesar 80% untuk kategori Risiko Sedang karena mayoritas pohon keputusan memberikan prediksi pada kelas tersebut.

2) Evaluasi Model

Evaluasi model Random Forest dilakukan menggunakan metrik *accuracy, precision, recall, dan F1-score*. Evaluasi ini bertujuan untuk mengukur tingkat kesesuaian hasil prediksi Random Forest terhadap label risiko hasil clustering K-Means.

Berdasarkan hasil pengujian, seluruh data berhasil diprediksikan dengan benar sehingga diperoleh nilai akurasi sebagai berikut:

$$Accuracy = \frac{5}{5} = 1,0 \text{ (100\%)}$$

Hasil evaluasi *precision*, *recall*, dan *F1-score* ditunjukkan pada Tabel 3.9.

Tabel 3. 9 *Precision*, *Recall*, *F1-score*

Kelas	Precision	Recall	F1-score
Risiko Rendah	1.0	1.0	1.0
Risiko Sedang	1.0	1.0	1.0
Risiko Tinggi	1.0	1.0	1.0

Berdasarkan hasil evaluasi pada Tabel 3.9, seluruh kelas memperoleh nilai *precision*, *recall*, dan *F1-score* sebesar 1.0. Hasil ini menunjukkan bahwa model *Random Forest* mampu memprediksikan seluruh data dengan sangat baik pada data simulasi yang digunakan.

Namun, perlu diperhatikan bahwa hasil evaluasi ini masih bersifat *simulatif* karena data pelatihan dan data pengujian berasal dari dataset yang sama. Oleh karena itu, hasil evaluasi lebih menunjukkan kemampuan model dalam mempelajari pola data dibandingkan kemampuan generalisasi terhadap data baru.

Secara keseluruhan, kombinasi algoritma *K-Means* dan *Random Forest* mampu melakukan pengelompokan dan prediksi risiko kredit secara objektif dan konsisten. Selain menghasilkan label risiko, model juga memberikan informasi probabilitas prediksi yang dapat membantu pihak koperasi dalam proses pengambilan keputusan terkait pemberian kredit dan pengawasan terhadap nasabah dengan tingkat risiko tertentu.

BAB IV

HASIL DAN PEMBAHASAN

A. Hasil Penelitian

1. Implementasi Desain Sistem

a. Implementasi Lembar Kerja

1) Pengumpulan Data

Data penelitian diperoleh dari Koperasi Sri Hadi Dharma dalam bentuk file Microsoft Excel yang berisi data historis pinjaman anggota periode 2020 – 2025. *Dataset* yang diperoleh terdiri dari 7.155 data pinjaman yang memuat informasi terkait kondisi pinjaman dan pembayaran anggota. Data tersebut digunakan sebagai dasar dalam proses pembentukan model prediksi risiko kredit menggunakan metode *K-Means* dan *Random Forest*.

Atribut yang digunakan dalam penelitian meliputi Besar Pinjaman, Sisa Pinjaman, Angsuran Pokok, Jasa, Tunggalan, dan Tenor. Selain atribut utama tersebut, dibentuk atribut tambahan berupa Tunggalan Maksimum yang digunakan sebagai indikator risiko pada proses *clustering* dan prediksi.

Data yang diperoleh selanjutnya digunakan sebagai masukan pada tahap *preprocessing* untuk mempersiapkan data sebelum dilakukan proses *clustering* dan prediksi.

2) *Preprocessing* Data

Tahap *preprocessing* data dilakukan untuk mempersiapkan dataset sebelum digunakan pada proses *clustering* dan prediksi. Proses ini bertujuan untuk menghasilkan data yang lebih representatif sehingga dapat meningkatkan kualitas model yang dibangun. Pada penelitian ini, *preprocessing* meliputi pengecekan *missing value*, pembentukan fitur tambahan, seleksi data anggota, serta normalisasi data.

Langkah pertama yang dilakukan adalah pengecekan *missing value* pada seluruh atribut dataset. Pemeriksaan ini bertujuan untuk

memastikan bahwa data yang digunakan tidak memiliki nilai kosong yang dapat memengaruhi hasil analisis. Hasil pemeriksaan menunjukkan bahwa seluruh atribut tidak memiliki *missing value* sehingga seluruh data dapat digunakan pada proses penelitian tanpa memerlukan penanganan data lebih lanjut.

Tabel 4. 1 Hasil Pengecekan *Missing Value*

Atribut	<i>Missing Value</i>
Besar Pinjaman	0
Sisa Pinjaman	0
Angsuran Pokok	0
Jasa	0
Tunggakan	0
Tenor	0

Setelah dilakukan pengecekan *missing value*, dilakukan pembentukan fitur tambahan berupa Tunggakan Maksimum. Fitur ini dibentuk untuk merepresentasikan nilai tunggakan tertinggi yang dimiliki oleh setiap anggota dan digunakan sebagai indikator tambahan dalam proses analisis risiko kredit. Penambahan fitur tersebut bertujuan untuk memperkaya informasi yang digunakan pada tahap *clustering* dan prediksi.

Selanjutnya dilakukan proses seleksi data untuk memastikan bahwa setiap anggota hanya direpresentasikan oleh satu data pinjaman. Dataset awal terdiri dari 7.155 data pinjaman yang masih mengandung beberapa *record* untuk anggota yang sama. Oleh karena itu, dilakukan pemilihan data berdasarkan nilai tenor terkecil sehingga diperoleh satu data yang dianggap paling representatif untuk setiap anggota.

Hasil proses seleksi data ditunjukkan pada Tabel 4.2

Tabel 4. 2 Hasil Seleksi Data

Tahap	Jumlah Data
Data Awal	7.155
Data setelah Seleksi	471

Berdasarkan Tabel 4.2, jumlah data berkurang dari 7.155 data menjadi 471 data anggota. Data hasil seleksi tersebut kemudian digunakan pada proses analisis selanjutnya.

Setelah proses seleksi selesai, dilakukan normalisasi data menggunakan metode *Min-Max Scaling*. Normalisasi dilakukan karena atribut yang digunakan memiliki rentang nilai yang berbeda-beda. Proses ini bertujuan untuk menyamakan skala seluruh atribut sehingga tidak terdapat atribut yang mendominasi proses perhitungan jarak pada algoritma *K-Means*.

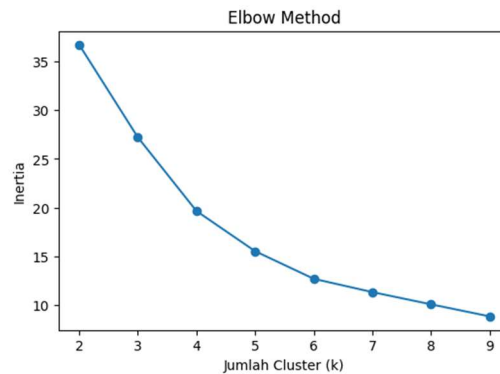
Atribut yang dinormalisasi meliputi Besar Pinjaman, Sisa Pinjaman, Angsuran Pokok, Tenor, dan Tunggakan Maksimum. Hasil normalisasi menghasilkan nilai atribut dalam rentang 0 hingga 1 sehingga data siap digunakan pada tahap *clustering* menggunakan algoritma *K-Means* dan prediksi menggunakan algoritma *Random Forest*.

3) Clustering K-Means

a) Penentuan Jumlah Cluster

Setelah data melalui tahap *preprocessing*, proses *clustering* dilakukan menggunakan algoritma *K-Means*. Langkah pertama adalah menentukan jumlah cluster optimal menggunakan metode *Elbow*. Metode ini dilakukan dengan mengamati nilai *inertia* pada beberapa jumlah cluster. Titik siku (*elbow*) yang terbentuk menunjukkan jumlah cluster yang paling optimal untuk digunakan pada proses pengelompokan data.

Hasil pengujian menggunakan metode *Elbow* ditunjukkan pada Gambar 4.1



Gambar 4.1 Elbow Method

Berdasarkan Gambar 4.1, titik siku mulai terlihat pada nilai $k=3$ sehingga jumlah *cluster* yang digunakan dalam penelitian ini adalah tiga *cluster*. Pemilihan tiga *cluster* juga mempertimbangkan hasil evaluasi *cluster* yang dilakukan menggunakan *Silhouette Score* dan *Davies-Bouldin Index*.

b) Hasil Evaluasi Cluster

Untuk memastikan kualitas *cluster* yang terbentuk, dilakukan evaluasi menggunakan *Silhouette Score* dan *Davies-Bouldin Index*. Hasil evaluasi ditunjukkan pada Tabel 4.3.

Tabel 4.3 Hasil Evaluasi Cluster

Jumlah Cluster	Silhouette Score	Davies-Bouldin Index
2	0,563	0,966
3	0,557	0,901
4	0,466	0,847
5	0,430	0,881

Nilai *Silhouette Score* tertinggi diperoleh pada $k=2$, sedangkan nilai *Davies-Bouldin Index* terbaik diperoleh pada $k=4$. Namun demikian, nilai pada $k=3$ menunjukkan keseimbangan antara kedua metrik tersebut sehingga dipilih sebagai jumlah *cluster* optimal. Selain itu, penggunaan tiga *cluster* dinilai lebih sesuai untuk merepresentasikan tingkat risiko kredit ke dalam kategori risiko rendah, sedang, dan tinggi.

c) Uji Stabilitas *Cluster*

Uji stabilitas *cluster* dilakukan untuk mengetahui tingkat konsistensi hasil pengelompokan yang dihasilkan oleh algoritma *K-Means*. Pengujian ini dilakukan menggunakan metode *Adjusted Rand Index* (ARI) dengan membandingkan hasil *clustering* dari beberapa iterasi yang berbeda. Nilai ARI berada pada rentang -1 hingga 1, dimana nilai yang semakin mendekati 1 menunjukkan bahwa hasil *clustering* memiliki tingkat kesamaan dan konsistensi yang tinggi.

Hasil pengujian stabilitas *cluster* ditunjukkan pada Tabel 4.4.

Tabel 4. 4 Hasil Uji Stabilitas *Cluster* Menggunakan ARI

Perbandingan Iterasi	Nilai ARI
1 vs 10	1,0000
10 vs 20	0,9984
20 vs 30	1,0000
30 vs 42	0,9219

Berdasarkan Tabel 4.4, seluruh nilai ARI berada di atas 0,90 dengan rata-rata sebesar 0,9801. Hasil tersebut menunjukkan bahwa algoritma *K-Means* menghasilkan *cluster* yang sangat stabil meskipun dilakukan pada iterasi yang berbeda. Tingginya nilai ARI mengindikasikan bahwa anggota yang berada pada suatu *cluster* cenderung tetap berada pada *cluster* yang sama pada proses *clustering* berikutnya.

d) Pembentukan Pseudo-label Risiko

Setelah proses *clustering* menghasilkan tiga *cluster*, langkah berikutnya adalah melakukan interpretasi terhadap karakteristik masing-masing *cluster*. Interpretasi dilakukan dengan menganalisis nilai rata-rata atribut pada setiap *cluster*, terutama atribut yang berkaitan dengan risiko kredit seperti sisa pinjaman, tenor, dan tunggakan maksimum. Hasil karakteristik setiap *cluster* ditunjukkan pada Tabel 4.5.

Tabel 4. 5 Karakteristik Cluster

Cluster	Besar Pinjaman	Sisa Pinjaman	Angsuran Pokok	Tenor	Tunggakan Maksimum
0	3.006.254	420.542	260.557	1,67	118.606
1	6.097.421	2.610.526	471.526	6,54	136.725
2	4.261.132	1.067.075	355.094	2,17	2.547.258

Berdasarkan karakteristik pada Tabel 4.5, interpretasi tingkat risiko dilakukan dengan mempertimbangkan nilai rata-rata Tunggakan Maksimum pada setiap *cluster*. Atribut tersebut dipilih karena mampu merepresentasikan tingkat risiko kredit anggota secara lebih jelas dibandingkan atribut lainnya.

Cluster 0 memiliki nilai rata-rata Tunggakan Maksimum sebesar Rp118.606, yang merupakan nilai terendah dibandingkan *cluster* lainnya. Hal ini menunjukkan bahwa anggota pada *cluster* tersebut memiliki tingkat keterlambatan pembayaran yang relatif rendah sehingga dikategorikan sebagai kelompok Risiko Rendah.

Cluster 1 memiliki nilai rata-rata Tunggakan Maksimum sebesar Rp136.725. Nilai tersebut sedikit lebih tinggi dibandingkan *Cluster 0*, sehingga anggota pada *cluster* ini dikategorikan sebagai kelompok Risiko Sedang. Selain memiliki nilai tunggakan yang lebih tinggi, cluster ini juga memiliki rata-rata tenor yang lebih panjang dibandingkan *cluster* lainnya.

Sementara itu, *Cluster 2* memiliki nilai rata-rata Tunggakan Maksimum sebesar Rp2.547.258, yang merupakan nilai tertinggi dibandingkan *cluster* lainnya. Tingginya nilai tunggakan menunjukkan bahwa anggota dalam *cluster* ini memiliki risiko kredit yang lebih besar sehingga dikategorikan sebagai kelompok Risiko Tinggi.

Hasil interpretasi tersebut menghasilkan tiga kategori risiko, yaitu Risiko Rendah, Risiko Sedang, dan Risiko Tinggi. Kategori risiko

yang diperoleh dari proses *clustering* selanjutnya digunakan sebagai pseudo-label pada proses prediksi menggunakan algoritma Random Forest.

4) Random Forest

Setelah proses *clustering K-Means* menghasilkan *pseudo-label* risiko, tahap selanjutnya adalah membangun model prediksi menggunakan algoritma *Random Forest*. Pada penelitian ini, *pseudo-label* berupa kategori Risiko Rendah, Risiko Sedang, dan Risiko Tinggi digunakan sebagai variabel target (label) dalam proses prediksi. Kategori risiko tersebut diperoleh dari hasil interpretasi *cluster* berdasarkan nilai rata-rata Tunggakan Maksimum pada masing-masing *cluster*.

Variabel yang digunakan dalam proses prediksi meliputi Besar Pinjaman, Sisa Pinjaman, Angsuran Pokok, Tenor, dan Tunggakan Maksimum. Sementara itu, variabel target yang digunakan adalah kategori Risiko hasil pembentukan *pseudo-label* dari proses *clustering K-Means*.

Sebelum proses pelatihan model dilakukan, dataset dibagi menjadi data pelatihan (*training data*) sebesar 80% dan data pengujian (*testing data*) sebesar 20% menggunakan metode *stratified sampling*. Metode ini digunakan untuk mempertahankan proporsi masing-masing kategori risiko pada data pelatihan dan data pengujian sehingga distribusi kelas tetap terjaga. Meskipun distribusi data antar kelas tidak seimbang, penggunaan *stratified sampling* memungkinkan setiap kategori risiko tetap terwakili pada kedua *subset* data sehingga proses pelatihan dan evaluasi model dapat dilakukan secara lebih seimbang.

Model Random Forest dibangun menggunakan parameter $n_estimators = 100$ dan $max_depth = 10$. Selanjutnya model dilatih menggunakan data pelatihan untuk mempelajari pola hubungan antara atribut pinjaman dengan kategori risiko yang telah diperoleh dari hasil

1 *clustering* K-Means. Model yang telah terbentuk kemudian digunakan untuk memprediksikan tingkat risiko kredit anggota.

184 Distribusi kategori risiko yang digunakan sebagai target prediksi ditunjukkan pada Tabel 4.6.

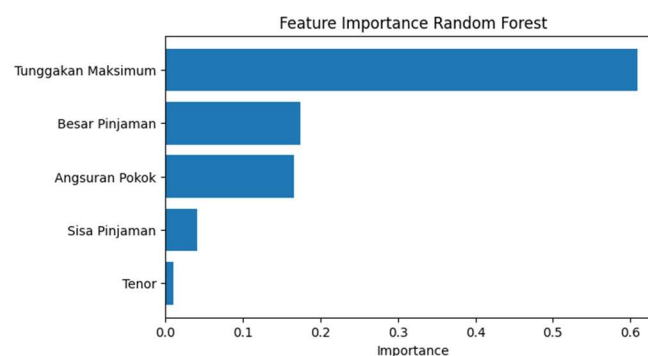
Tabel 4. 6 Distribusi Risiko

Risiko	Jumlah
Rendah	323
Sedang	95
Tinggi	53

264 Berdasarkan Tabel 4.6, mayoritas anggota berada pada kategori Risiko Rendah sebanyak 323 anggota, sedangkan kategori Risiko Sedang dan Risiko Tinggi masing-masing berjumlah 95 dan 53 anggota. Distribusi tersebut menunjukkan bahwa sebagian besar anggota memiliki tingkat risiko kredit yang rendah.

276 Selain membangun model prediksi, dilakukan pula analisis tingkat kepentingan atribut menggunakan *feature importance* pada Random Forest. Hasil analisis ditunjukkan pada Gambar 4.2.

1 Gambar 4. 2 *Feature Importance Random Forest*



Berdasarkan hasil *feature importance*, atribut Tunggalan Maksimum memiliki kontribusi terbesar dalam proses prediksi risiko kredit. Selanjutnya diikuti oleh atribut Besar Pinjaman, Angsuran Pokok, Sisa Pinjaman, dan Tenor. Hasil tersebut menunjukkan bahwa

riwayat tunggakan merupakan faktor yang paling berpengaruh dalam menentukan tingkat risiko kredit anggota.

Dengan demikian, algoritma *Random Forest* mampu memanfaatkan pseudo-label hasil *clustering* K-Means untuk membentuk model prediksi risiko kredit berdasarkan karakteristik pinjaman anggota.

5) Implementasi Sistem

Implementasi sistem dilakukan menggunakan *framework Streamlit* untuk membangun aplikasi berbasis

web yang dapat digunakan dalam proses prediksi risiko kredit anggota koperasi. Sistem yang dibangun mengintegrasikan model prediksi *Random Forest* yang dibentuk menggunakan pseudo-label hasil *clustering* K-Means sehingga pengguna dapat melihat informasi risiko kredit secara visual dan melakukan prediksi terhadap data anggota.

a) Implementasi Dashboard Utama

Halaman *dashboard* merupakan halaman utama yang ditampilkan ketika sistem dijalankan. *Dashboard* berfungsi untuk memberikan ringkasan informasi terkait hasil prediksi risiko kredit anggota koperasi.



Gambar 4. 3 Dashboard Utama

Berdasarkan Gambar 4.3, *dashboard* menampilkan beberapa informasi penting yang terdiri dari total jumlah anggota yang

190 dianalisis, jumlah anggota dengan kategori risiko rendah, risiko
37 sedang, dan risiko tinggi. Informasi tersebut ditampilkan dalam bentuk kartu informasi (*summary card*) sehingga memudahkan pengguna dalam memahami kondisi risiko kredit secara keseluruhan.

164 Pada *dashboard* juga ditampilkan grafik distribusi risiko yang menunjukkan jumlah anggota pada masing-masing kategori risiko. Berdasarkan hasil prediksi, terdapat 323 anggota dengan risiko rendah, 95 anggota dengan risiko sedang, dan 53 anggota dengan risiko tinggi. Visualisasi tersebut membantu pengguna dalam melihat persebaran risiko kredit anggota secara cepat dan intuitif.

Selain itu, *dashboard* menampilkan tabel *preview* data anggota yang berisi informasi identitas anggota beserta kategori risiko yang diperoleh dari hasil prediksi. Tabel ini berfungsi sebagai ringkasan data yang memungkinkan pengguna melakukan pengecekan awal terhadap hasil prediksi yang dihasilkan sistem.

b) Implementasi Menu Prediksi Risiko

77 Menu Prediksi Risiko digunakan untuk melakukan prediksi tingkat risiko kredit berdasarkan data pinjaman yang dimasukkan oleh pengguna. Pada halaman ini, pengguna dapat mengisi beberapa atribut yang digunakan sebagai variabel *prediktor* pada model *Random Forest*, yaitu Besar Pinjaman, Sisa Pinjaman, Angsuran Pokok, Tenor, dan Tunggalan Maksimum.

Gambar 4. 4 Menu Prediksi Risiko

Berdasarkan Gambar 4.4, sistem menyediakan form input yang memungkinkan pengguna memasukkan data pinjaman anggota. Setelah seluruh data diisi, pengguna dapat menekan tombol **Prediksi Risiko** untuk menjalankan proses prediksi menggunakan model *Random Forest* yang telah dilatih sebelumnya. Data yang dimasukkan akan diproses oleh sistem untuk menghasilkan kategori risiko kredit yang sesuai.

Setelah proses prediksi selesai dilakukan, sistem akan menampilkan hasil prediksi beserta informasi pendukung yang dapat digunakan sebagai bahan pertimbangan dalam pengambilan keputusan kredit.

Gambar 4. 5 Hasil Prediksi Risiko

Berdasarkan Gambar 4.5, sistem menampilkan hasil prediksi berupa kategori risiko kredit dan nilai *risk score* yang menunjukkan

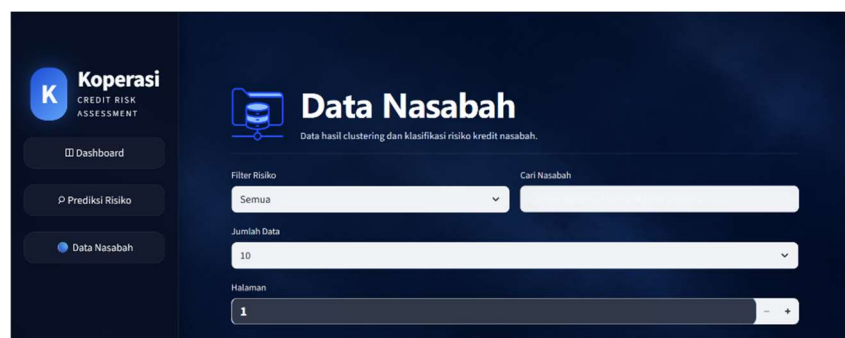
tingkat keyakinan model terhadap hasil prediksi yang diberikan. Pada contoh pengujian yang dilakukan, sistem menghasilkan kategori Risiko Sedang dengan nilai *risk score* sebesar 84,0%.

Selain menampilkan kategori risiko, sistem juga memberikan interpretasi hasil prediksi untuk membantu pengguna memahami tingkat risiko yang dimiliki calon peminjam. Pada kategori Risiko Sedang, sistem memberikan informasi bahwa pengajuan kredit memerlukan evaluasi lebih lanjut karena terdapat beberapa faktor risiko yang perlu diperhatikan. Sistem juga menampilkan rekomendasi bahwa pengajuan masih dapat dipertimbangkan dengan perhatian khusus terhadap kondisi pinjaman anggota.

Fitur prediksi risiko ini memungkinkan pengguna memperoleh hasil prediksi secara cepat dan otomatis tanpa harus melakukan perhitungan manual. Dengan demikian, proses penilaian risiko kredit dapat dilakukan secara lebih efisien dan konsisten sesuai dengan model yang telah dibangun.

c) Implementasi Menu Data Nasabah

Menu Data Nasabah digunakan untuk menampilkan seluruh data anggota yang telah melalui proses *clustering* dan prediksi risiko kredit. Halaman ini berfungsi sebagai media *monitoring* hasil prediksi sehingga pengguna dapat melihat informasi anggota beserta kategori risiko yang dihasilkan oleh sistem.



Gambar 4. 6 Halaman Data Nasabah

Berdasarkan Gambar 4.6, sistem menyediakan beberapa fitur untuk mempermudah pengguna dalam mengakses data anggota. Fitur tersebut meliputi filter risiko, pencarian data anggota, pengaturan jumlah data yang ditampilkan, serta navigasi halaman (*pagination*). Fitur filter risiko memungkinkan pengguna menampilkan anggota berdasarkan kategori Risiko Rendah, Risiko Sedang, atau Risiko Tinggi. Selain itu, tersedia fitur pencarian yang dapat digunakan untuk mencari anggota berdasarkan nama atau ID anggota.

Pengguna juga dapat menentukan jumlah data yang ditampilkan pada setiap halaman sehingga proses penelusuran data menjadi lebih fleksibel. Untuk mempermudah navigasi, sistem menyediakan fitur *pagination* yang memungkinkan pengguna berpindah dari satu halaman ke halaman lainnya.



ID Anggota	Nama Anggota	Besar Pinjaman	Sisa Pinjaman	Angsuran Pokok	Risiko
10	Soepandi/ Hentri	5,400,000	1,800,000	450,000	Sedang
26	Pitran Triyono	5,040,000	840,000	420,000	Rendah
46	Soni	9,000,000	750,000	750,000	Rendah
59	Juari	4,020,000	335,000	335,000	Rendah
60	Didik Subandi	3,000,000	750,000	250,000	Rendah
69	Karto	2,520,000	630,000	210,000	Rendah

Menampilkan 1-10 dari 471 data

Halaman 1 dari 48

Gambar 4. 7 Tabel Data Nasabah

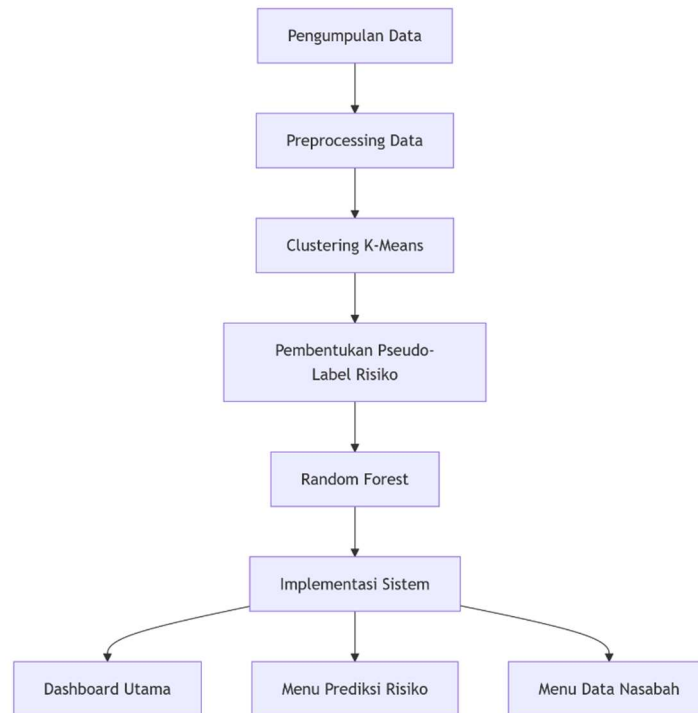
Berdasarkan Gambar 4.7 sistem menampilkan informasi anggota dalam bentuk tabel yang terdiri dari beberapa atribut, yaitu ID Anggota, Nama Anggota, Besar Pinjaman, Sisa Pinjaman, Angsuran Pokok, dan kategori Risiko. Kategori risiko ditampilkan menggunakan indikator warna yang berbeda untuk memudahkan identifikasi tingkat risiko anggota. Risiko Rendah ditampilkan dengan warna hijau, Risiko Sedang dengan warna kuning, dan Risiko Tinggi dengan warna merah.

23 Tabel data nasabah memungkinkan pengguna melakukan pengamatan terhadap hasil prediksi secara lebih rinci. Melalui tampilan ini, pengguna dapat mengetahui kondisi pinjaman masing-masing anggota beserta tingkat risiko kredit yang telah dihasilkan oleh sistem. Dengan demikian, proses pemantauan dan evaluasi risiko kredit dapat dilakukan secara lebih cepat dan terstruktur.

23 Menu Data Nasabah juga terintegrasi dengan hasil prediksi yang diperoleh dari model *Random Forest* sehingga setiap data anggota yang ditampilkan telah memiliki kategori risiko yang dapat digunakan sebagai bahan pertimbangan dalam proses pengambilan keputusan kredit di koperasi.

b. Keterkaitan Antar Lembar Kerja

177 Keterkaitan antar lembar kerja menggambarkan hubungan antar proses yang dilakukan dalam penelitian mulai dari pengumpulan data hingga implementasi sistem. Setiap proses saling terhubung dan menghasilkan keluaran yang digunakan sebagai masukan pada proses berikutnya.



Gambar 4. 8 Keterkaitan Antar Lembar Kerja

Berdasarkan Gambar 4.8, proses penelitian diawali dengan pengumpulan data pinjaman anggota Koperasi Sri Hadi Dharma. Data yang diperoleh kemudian melalui tahap *preprocessing* yang meliputi pengecekan *missing value*, pembentukan fitur Tunggalan Maksimum, seleksi data anggota, dan normalisasi data.

Data hasil *preprocessing* selanjutnya digunakan pada proses *clustering* menggunakan algoritma *K-Means* untuk mengelompokkan anggota berdasarkan karakteristik pinjaman yang dimiliki. Hasil *clustering* kemudian diinterpretasikan menjadi tiga kategori risiko, yaitu Risiko Rendah, Risiko Sedang, dan Risiko Tinggi sehingga terbentuk pseudo-label risiko.

Pseudo-label yang dihasilkan dari proses *clustering* digunakan sebagai variabel target pada tahap prediksi menggunakan algoritma *Random Forest*. Model *Random Forest* yang telah dilatih kemudian diimplementasikan ke dalam sistem berbasis web menggunakan *framework Streamlit*.

Pada tahap implementasi sistem, model yang telah dibangun digunakan untuk mendukung berbagai fitur yang tersedia pada aplikasi, yaitu dashboard utama, menu prediksi risiko, dan menu data nasabah. Dengan demikian, seluruh proses mulai dari pengumpulan data hingga penyajian hasil prediksi saling terhubung dalam satu alur sistem yang terintegrasi.

2. Pengujian Fungsional

Pengujian fungsional dilakukan untuk memastikan bahwa setiap fitur utama pada sistem prediksi risiko kredit anggota koperasi dapat berjalan sesuai dengan kebutuhan dan perancangan sistem yang telah dibuat. Pengujian dilakukan menggunakan metode *Black Box Testing*, yaitu dengan menguji fungsi sistem berdasarkan masukan dan keluaran yang dihasilkan tanpa memperhatikan proses internal program. Setiap fitur diuji untuk mengetahui kesesuaian antara hasil yang diharapkan dengan hasil aktual yang diperoleh dari sistem.

Tabel 4. 7 Hasil Pengujian Fungsional

No	Fitur Sistem	Skenario Pengujian	Hasil yang Diharapkan	Hasil Aktual	Status
1	Dashboard Utama	Membuka halaman dashboard	Informasi total nasabah, distribusi risiko, dan preview data nasabah ditampilkan	Informasi berhasil ditampilkan sesuai data sistem	Berhasil
2	Prediksi Risiko	Memasukkan data pinjaman dan menekan tombol Prediksi Risiko	Sistem menampilkan kategori risiko, nilai risk score, dan interpretasi hasil prediksi	Kategori risiko, nilai risk score, dan interpretasi hasil berhasil ditampilkan	Berhasil
3	Data Nasabah	Membuka menu Data Nasabah	Data anggota ditampilkan dalam bentuk tabel	Data anggota berhasil ditampilkan	Berhasil

4	Filter Risiko	Memilih kategori risiko tertentu	Data anggota ditampilkan sesuai kategori risiko yang dipilih	Data berhasil difilter sesuai kategori risiko	Berhasil
5	Pencarian Data	Memasukkan nama atau ID anggota pada kolom pencarian	Data anggota yang dicari ditampilkan	Data berhasil ditemukan	Berhasil
6	Jumlah Data	Memilih jumlah data yang ditampilkan pada tabel	Jumlah data pada tabel berubah sesuai pilihan	Jumlah data berhasil diperbarui	Berhasil
7	Navigasi Halaman (Pagination)	Mengubah halaman data menggunakan tombol navigasi	Data pada halaman yang dipilih ditampilkan	Data berhasil ditampilkan sesuai halaman yang dipilih	Berhasil

Berdasarkan hasil pengujian fungsional pada Tabel 4.7, seluruh fitur utama pada sistem prediksi risiko kredit anggota koperasi telah berjalan dengan baik sesuai dengan fungsi yang dirancang. Halaman dashboard mampu menampilkan informasi mengenai jumlah anggota dan distribusi risiko kredit secara visual. Selain itu, menu prediksi risiko dapat melakukan prediksi berdasarkan data pinjaman yang dimasukkan pengguna serta menampilkan kategori risiko, nilai *risk score*, dan interpretasi hasil prediksi.

Pada menu Data Nasabah, sistem mampu menampilkan data anggota dalam bentuk tabel serta mendukung fitur pencarian data, filter kategori risiko, pengaturan jumlah data yang ditampilkan, dan navigasi halaman. Seluruh skenario pengujian memperoleh status berhasil sehingga dapat disimpulkan bahwa sistem yang dibangun telah memenuhi kebutuhan fungsional dan siap digunakan untuk membantu proses prediksi risiko kredit anggota pada Koperasi Sri Hadi Dharma.

3. Pengujian Non Fungsional

Pengujian non-fungsional dilakukan untuk mengevaluasi performa model *Random Forest* dalam melakukan prediksi risiko kredit anggota koperasi.

Pengujian ini bertujuan untuk mengetahui kemampuan model dalam melakukan prediksi risiko kredit serta mengetahui tingkat penerimaan pengguna terhadap sistem yang dikembangkan.

a. Evaluasi Model *Random Forest*

Evaluasi model dilakukan menggunakan data pengujian sebesar 20% dari keseluruhan dataset. Pembagian data dilakukan menggunakan metode *stratified sampling* sehingga proporsi setiap kategori risiko tetap terjaga pada data pelatihan dan data pengujian. Evaluasi model dilakukan menggunakan metrik *Accuracy*, *Precision*, *Recall*, *F1-Score*, serta *Confusion Matrix*.

Hasil evaluasi model *Random Forest* ditunjukkan pada Tabel 4.8.

Tabel 4. 8 Hasil Evaluasi Model *Random Forest*

Metrik	Nilai
Accuracy	98,95%
Precision Macro	99%
Recall Macro	97%
F1-Score Macro	98%

Berdasarkan Tabel 4.8, model *Random Forest* memperoleh nilai *accuracy* sebesar 98,95%, yang menunjukkan bahwa model mampu melakukan prediksi risiko kredit dengan tingkat ketepatan yang sangat tinggi. Nilai *precision macro* sebesar 99% menunjukkan bahwa hasil prediksi yang dihasilkan memiliki tingkat ketepatan yang sangat baik. Sementara itu, nilai *recall macro* sebesar 97% menunjukkan bahwa model mampu mengenali sebagian besar data pada setiap kategori risiko. Nilai *F1-score macro* sebesar 98% menunjukkan keseimbangan yang sangat baik antara *precision* dan *recall*.

Selanjutnya, hasil prediksi model divisualisasikan menggunakan *confusion matrix* sebagaimana ditunjukkan pada Gambar 4.9.



Gambar 4. 9 Confusion Matrix Random Forest

Berdasarkan Gambar 4.9, sebagian besar data pada setiap kategori risiko berhasil diprediksi dengan benar oleh model *Random Forest*. Dari 65 data pada kategori Risiko Rendah, seluruh data berhasil diprediksi secara tepat. Pada kategori Risiko Sedang terdapat 19 data dan seluruhnya berhasil diprediksi dengan benar. Sementara itu, pada kategori Risiko Tinggi terdapat 11 data, dimana 10 data berhasil diprediksi dengan benar dan 1 data diprediksi sebagai Risiko Rendah.

Hasil tersebut menunjukkan bahwa tingkat kesalahan prediksi model relatif rendah. Kesalahan prediksi hanya terjadi pada satu data kategori Risiko Tinggi yang diprediksi sebagai Risiko Rendah. Secara keseluruhan, *confusion matrix* menunjukkan bahwa model *Random Forest* memiliki kemampuan yang sangat baik dalam membedakan kategori risiko kredit anggota koperasi.

3) Analisis Hasil Evaluasi

Untuk mengetahui performa model pada masing-masing kategori risiko, dilakukan evaluasi menggunakan *classification report* yang hasilnya ditunjukkan pada Tabel 4.9.

Tabel 4. 9 *Classification Report Random Forest*

Kategori Risiko	Precision	Recall	F1-Score	Support
Rendah	0,98	1,00	0,99	65
Sedang	1,00	1,00	1,00	19
Tinggi	1,00	0,91	0,95	11

Berdasarkan Tabel 4.9, kategori Risiko Rendah memperoleh nilai *precision* sebesar 0,98, *recall* sebesar 1,00, dan *F1-score* sebesar 0,99. Hal ini menunjukkan bahwa model mampu mengenali anggota dengan risiko rendah dengan tingkat ketepatan yang sangat baik. Pada kategori Risiko Sedang diperoleh nilai *precision*, *recall*, dan *F1-score* masing-masing sebesar 1,00, yang menunjukkan bahwa seluruh data pada kategori tersebut berhasil diprediksi dengan benar.

Sementara itu, kategori Risiko Tinggi memperoleh nilai *precision* sebesar 1,00, *recall* sebesar 0,91, dan *F1-score* sebesar 0,95. Nilai *recall* yang lebih rendah dibandingkan kategori lainnya menunjukkan bahwa masih terdapat satu data Risiko Tinggi yang diprediksi sebagai Risiko Rendah. Meskipun demikian, secara keseluruhan hasil *classification report* menunjukkan bahwa model *Random Forest* mampu mengenali seluruh kategori risiko dengan performa yang sangat baik dan memiliki tingkat kesalahan prediksi yang relatif rendah.

4) Analisis Kinerja Model

Berdasarkan keseluruhan hasil evaluasi, model *Random Forest* yang digunakan dalam penelitian ini memiliki performa yang sangat baik dalam melakukan prediksi risiko kredit anggota koperasi. Tingginya nilai *accuracy*, *precision*, *recall*, dan *F1-score* menunjukkan bahwa model mampu mempelajari pola data dengan baik sehingga menghasilkan prediksi yang konsisten terhadap pseudo-label yang diperoleh dari proses *clustering K-Means*.

72 Performa model yang baik dipengaruhi oleh beberapa faktor, antara lain kualitas data historis pinjaman yang digunakan dalam penelitian, proses *preprocessing* yang meliputi pembentukan fitur Tunggalan Maksimum, seleksi data anggota, dan normalisasi data, pembentukan pseudo-label menggunakan algoritma *K-Means*, serta penggunaan metode *stratified sampling* yang menjaga proporsi masing-masing kategori risiko pada data pelatihan dan data pengujian.

8 Selain itu, penggunaan algoritma *Random Forest* dengan parameter $n_estimators = 100$ dan $max_depth = 10$ mampu menghasilkan model yang memiliki kemampuan generalisasi yang baik terhadap data yang digunakan. Dengan demikian, model *Random Forest* yang dibangun mampu melakukan prediksi risiko kredit dengan tingkat ketepatan yang sangat tinggi dan dapat digunakan sebagai dasar dalam mendukung proses pengambilan keputusan pada Koperasi Sri Hadi Dharma.

B. Pembahasan

273 Penelitian ini mengembangkan sistem prediksi risiko kredit anggota Koperasi Sri Hadi Dharma menggunakan pendekatan *hybrid* yang menggabungkan algoritma *K-Means* dan *Random Forest*. Pendekatan tersebut dipilih karena data historis yang dimiliki koperasi belum memiliki label risiko yang jelas, sehingga diperlukan metode yang mampu membentuk kategori risiko secara otomatis sebelum dilakukan proses prediksi. Sistem yang dikembangkan berhasil diimplementasikan dalam bentuk aplikasi berbasis web menggunakan *framework Streamlit* sehingga dapat membantu proses penilaian risiko kredit secara lebih terstruktur dan berbasis data.

63 Pada tahap *clustering*, algoritma *K-Means* berhasil mengelompokkan data anggota menjadi tiga kelompok risiko. Pemilihan jumlah *cluster* dilakukan berdasarkan hasil evaluasi menggunakan metode *Elbow*, *Silhouette Score*, dan *Davies-Bouldin Index*. Meskipun nilai terbaik tidak diperoleh pada jumlah

cluster yang sama, penggunaan tiga *cluster* dipilih karena lebih mampu merepresentasikan kondisi nyata tingkat risiko kredit anggota koperasi ke dalam kategori Risiko Rendah, Risiko Sedang, dan Risiko Tinggi. Selain itu, nilai *Adjusted Rand Index (ARI)* yang tinggi menunjukkan bahwa hasil *clustering* memiliki tingkat konsistensi yang baik sehingga dapat digunakan sebagai dasar pembentukan *pseudo-label* risiko.

Hasil pembentukan *pseudo-label* menunjukkan bahwa atribut Tunggalan Maksimum memiliki peranan penting dalam membedakan tingkat risiko kredit anggota. Anggota dengan nilai Tunggalan Maksimum yang rendah cenderung berada pada kelompok risiko rendah, sedangkan anggota dengan nilai Tunggalan Maksimum yang tinggi berada pada kelompok risiko tinggi. Kondisi tersebut sejalan dengan keadaan nyata pada koperasi, dimana riwayat keterlambatan pembayaran merupakan salah satu indikator utama dalam menilai kemampuan anggota dalam memenuhi kewajibannya. Dengan demikian, pembentukan *pseudo-label* menggunakan *K-Means* mampu menghasilkan kategori risiko yang logis dan sesuai dengan karakteristik data historis yang dimiliki koperasi.

Pada tahap prediksi, algoritma *Random Forest* mampu memanfaatkan *pseudo-label* hasil *clustering* sebagai target pembelajaran. Hasil pengujian menunjukkan bahwa model memiliki performa yang sangat baik dalam mempelajari pola data. Tingginya performa model dipengaruhi oleh proses *preprocessing* yang dilakukan, pembentukan atribut Tunggalan Maksimum melalui *feature engineering*, serta penggunaan metode *stratified sampling* yang mampu menjaga proporsi setiap kategori risiko pada data pelatihan dan data pengujian. Hal tersebut menunjukkan bahwa kombinasi *K-Means* dan *Random Forest* mampu menghasilkan model prediksi yang stabil meskipun data awal tidak memiliki label risiko aktual.

Berdasarkan hasil *confusion matrix*, sebagian besar data berhasil diprediksi dengan benar. Kesalahan prediksi hanya terjadi pada satu data kategori Risiko Tinggi yang diprediksi sebagai Risiko Rendah. Hal ini menunjukkan bahwa batas antar kategori risiko tidak selalu dapat dipisahkan

secara sempurna karena terdapat karakteristik anggota yang memiliki kemiripan antar kelompok. Meskipun demikian, tingkat kesalahan yang relatif kecil menunjukkan bahwa model memiliki kemampuan yang baik dalam membedakan tingkat risiko kredit anggota koperasi.

Distribusi data pada penelitian ini masih belum seimbang, dimana kategori Risiko Rendah memiliki jumlah data yang lebih banyak dibandingkan kategori Risiko Sedang dan Risiko Tinggi. Kondisi tersebut berpotensi menyebabkan model lebih banyak mempelajari pola pada kategori mayoritas. Namun demikian, penggunaan metode *stratified sampling* membantu mempertahankan proporsi masing-masing kategori pada data pelatihan dan data pengujian sehingga performa model tetap menunjukkan hasil yang baik. Pada penelitian selanjutnya, metode penanganan data tidak seimbang seperti SMOTE dapat dipertimbangkan untuk meningkatkan kemampuan model dalam mengenali seluruh kategori risiko secara lebih merata.

Hasil analisis *feature importance* menunjukkan bahwa atribut Tunggakan Maksimum merupakan variabel yang memiliki kontribusi paling besar dalam proses prediksi risiko kredit. Temuan ini menunjukkan bahwa riwayat tunggakan menjadi faktor yang paling berpengaruh dalam menentukan tingkat risiko kredit anggota. Semakin besar nilai tunggakan maksimum yang dimiliki anggota, maka semakin tinggi pula risiko kredit yang dapat terjadi. Hasil tersebut juga mendukung proses pembentukan pseudo-label yang sebelumnya menggunakan karakteristik tunggakan sebagai dasar interpretasi tingkat risiko.

Berdasarkan hasil pengujian fungsional menggunakan metode *Black Box Testing*, seluruh fitur pada sistem berhasil berjalan sesuai dengan fungsi yang telah dirancang. Dashboard mampu menampilkan informasi distribusi risiko anggota, menu Prediksi Risiko dapat digunakan untuk melakukan prediksi secara cepat, sedangkan menu Data Nasabah memudahkan pengguna dalam melakukan pencarian dan pengelolaan data anggota. Integrasi antara model prediksi dengan antarmuka berbasis web menjadikan sistem lebih mudah digunakan dan mendukung proses pengambilan keputusan secara lebih efisien.

Jika dibandingkan dengan kondisi nyata di Koperasi Sri Hadi Dharma, proses penilaian risiko kredit sebelumnya masih dilakukan secara manual berdasarkan pengalaman dan pertimbangan petugas. Pendekatan tersebut berpotensi menimbulkan subjektivitas dan membutuhkan waktu yang lebih lama. Sistem yang dikembangkan dalam penelitian ini menawarkan alternatif berupa proses penilaian risiko yang lebih objektif dan terstruktur berdasarkan data historis anggota. Dengan demikian, proses pengambilan keputusan kredit dapat dilakukan secara lebih cepat dan konsisten.

Dibandingkan dengan penelitian sebelumnya yang umumnya menggunakan metode *clustering* dan klasifikasi secara terpisah, penelitian ini mengintegrasikan algoritma *K-Means* dan *Random Forest* dalam satu alur sistem. Pendekatan *hybrid* tersebut memungkinkan pembentukan *pseudo-label* dari data yang belum memiliki label risiko sehingga model prediksi tetap dapat dibangun. Selain itu, sistem yang dikembangkan tidak hanya menghasilkan kategori risiko, tetapi juga menyediakan visualisasi distribusi risiko, analisis *feature importance*, serta antarmuka berbasis web yang memudahkan pengguna dalam melakukan *monitoring* data anggota.

Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan. *Pseudo-label* yang digunakan berasal dari hasil *clustering* sehingga belum merepresentasikan label risiko aktual yang sebenarnya. Selain itu, sistem yang dibangun hanya dapat digunakan pada anggota yang telah memiliki riwayat pinjaman sehingga belum dapat diterapkan pada nasabah baru. Penelitian ini juga belum mempertimbangkan faktor lain seperti pendapatan, pekerjaan, dan kondisi ekonomi anggota yang kemungkinan dapat mempengaruhi tingkat risiko kredit. Oleh karena itu, penelitian selanjutnya dapat mempertimbangkan penggunaan label risiko aktual, penambahan variabel yang lebih beragam, serta penerapan metode penanganan data tidak seimbang untuk meningkatkan performa model yang dihasilkan.

Secara keseluruhan, sistem prediksi risiko kredit yang dikembangkan mampu memberikan alternatif penilaian risiko yang lebih objektif, terstruktur, dan berbasis data. Pendekatan *hybrid K-Means* dan *Random Forest*

2

memungkinkan pemanfaatan data historis yang belum memiliki label risiko sehingga dapat mendukung proses pengambilan keputusan kredit pada Koperasi Sri Hadi Dharma secara lebih efektif.

BAB V

Penutup

A. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat diambil beberapa kesimpulan sebagai berikut:

1. Algoritma *K-Means* berhasil digunakan untuk mengelompokkan data anggota Koperasi Sri Hadi Dharma menjadi tiga kelompok risiko, yaitu Risiko Rendah, Risiko Sedang, dan Risiko Tinggi. Pembentukan kelompok dilakukan berdasarkan karakteristik data pinjaman anggota, dimana interpretasi tingkat risiko dilakukan menggunakan nilai rata-rata Tunggal Maksimum pada setiap *cluster*. Hasil evaluasi *clustering* menggunakan *Silhouette Score*, *Davies-Bouldin Index*, dan *Adjusted Rand Index (ARI)* menunjukkan bahwa penggunaan tiga cluster mampu menghasilkan kelompok yang stabil dan representatif untuk membentuk pseudo-label risiko.
2. Algoritma *Random Forest* berhasil dimanfaatkan untuk membangun model prediksi risiko kredit berdasarkan pseudo-label yang dihasilkan dari proses *clustering K-Means*. Model yang dibangun memperoleh nilai *accuracy* sebesar 98,95%, *precision macro* sebesar 99%, *recall macro* sebesar 97%, dan *F1-score macro* sebesar 98%. Selain menghasilkan kategori Risiko Rendah, Risiko Sedang, dan Risiko Tinggi, sistem juga menghasilkan nilai *risk score* yang digunakan sebagai dasar dalam menilai tingkat risiko masing-masing anggota. Sistem prediksi yang dikembangkan dalam bentuk aplikasi berbasis web menggunakan *Streamlit* mampu menampilkan informasi risiko secara visual dan mendukung proses pengambilan keputusan kredit pada Koperasi Sri Hadi Dharma secara lebih cepat, objektif, dan terstruktur.

B. Saran

Berdasarkan hasil penelitian yang telah dilakukan, beberapa saran yang dapat diberikan untuk pengembangan penelitian selanjutnya adalah sebagai berikut:

1. Penelitian selanjutnya dapat menggunakan label risiko aktual dari pihak koperasi sehingga hasil prediksi yang diperoleh dapat dibandingkan secara langsung dengan kondisi risiko yang sebenarnya.
2. Penambahan variabel lain seperti pendapatan, pekerjaan, lama keanggotaan, dan kondisi ekonomi anggota dapat dilakukan untuk meningkatkan kemampuan model dalam memprediksi risiko kredit.
3. Penelitian selanjutnya dapat menerapkan metode penanganan data tidak seimbang, seperti *SMOTE* atau teknik *oversampling* lainnya, untuk meningkatkan kemampuan model dalam mengenali kategori risiko yang memiliki jumlah data lebih sedikit.
4. Sistem yang dikembangkan dapat disempurnakan dengan menambahkan kemampuan prediksi untuk nasabah baru yang belum memiliki riwayat pinjaman serta menambahkan fitur pengelolaan data secara langsung pada aplikasi.
5. Penelitian selanjutnya dapat membandingkan performa *Random Forest* dengan algoritma lain, seperti *XGBoost*, *LightGBM*, atau *Support Vector Machine (SVM)*, untuk memperoleh model prediksi risiko kredit dengan performa yang lebih optimal.