

Klasifikasi Kualitas Anggur Putih Menggunakan Metode K-Nearest Neighbor (KNN)

Muhammad Fariz Luthfi (119140199), Muhammad Nadhif Athalla (119140209), Wilando Putrayuda (119140176)

I. PENDAHULUAN

Anggur atau *wine* merupakan minuman beralkohol yang terbuat dari anggur atau buah – buahan yang merupakan hasil fermentasi anaerob (tanpa kehadiran O₂) yang akan menghasilkan minuman yang didalamnya memiliki kadar alkohol sebesar 8 – 15%. Minuman ini dapat berfermentasi tanpa tambahan gula, asam, enzim, air atau nutrisi. Kualitas *wine* dapat ditentukan berdasarkan komposisi bahan baku, proses fermentasi, dan beberapa perubahan yang terjadi secara alami maupun disengaja selama proses fermentasi [1].

Wine banyak dikonsumsi oleh warga masyarakat dan mempunyai banyak jenis dalam klasifikasi minumannya. Untuk jenis *wine* tertentu seperti *wine* putih saja, terdiri dari berbagai jenis dengan kandungan dan kualitasnya masing-masing. Karena *wine* yang memiliki jenis yang sangat banyak, masyarakat yang mengonsumsinya terkadang kesulitan untuk menentukan kualitas *wine* yang layak dikonsumsi atau dikomersialkan karena setiap *wine* memiliki jenis dan kandungan yang berbeda-beda. Oleh karena itu, penelitian ini bertujuan untuk memudahkan proses prediksi kualitas *wine* berdasarkan klasifikasi jenis *wine* dan kadar kandungan dalam minuman tersebut [2].

Penelitian untuk memprediksi kualitas *wine* ini dilakukan dengan mengimplementasikan model *machine learning* sederhana yang mengadaptasi bentuk pendekatan algoritma *supervised learning*. Algoritma *supervised learning* merupakan algoritma *machine learning* yang proses pembelajarannya berada di bawah pengawasan guru atau supervisor. Algoritma ini memerlukan data berlabel untuk membangun sebuah model yang tingkat akurasi bisa ditingkatkan dari waktu ke waktu. Semakin banyak model tersebut mengolah data, maka tingkat akurasi juga akan semakin tinggi.

Untuk menerapkan pendekatan *supervised learning*, penelitian dilakukan dengan menerapkan metode *K-Nearest Neighbour* (KNN) yang biasanya diterapkan untuk mengklasifikasikan data berdasarkan nilai selisih kecil dari jarak tetangga yang terdekat dengan objek [3]. Prinsip umum dari algoritma ini adalah menentukan nilai K pada data latih yang selanjutnya akan diolah menggunakan KNN berdasarkan jaraknya. Nilai mayoritas atau nilai berikutnya

dari KNN akan dijadikan dasar dalam menentukan jenis kelas atau kategori sampel berikutnya [4]. Proses pelatihan data diimplementasikan dengan mengumpulkan dataset *wine* putih terlebih dahulu untuk selanjutnya data tersebut akan melalui tahapan praproses yang kemudian akan dilatih menggunakan metode KNN sehingga data dapat digunakan untuk memberikan prediksi nilai kualitas untuk setiap jenis *wine* putih berdasarkan parameter yang digunakan sesuai dengan data yang telah dilatih sebelumnya.

Untuk detail dari proses penelitian ini, hal pertama yang dilakukan adalah praproses (*preprocessing*) data yang merupakan sebuah tahap awal yang perlu dilakukan. Tujuan dari pra proses dalam data mining adalah untuk mempersiapkan data mentah sebelum dilakukannya proses lain. Beberapa metode praproses data yang dapat dilakukan adalah data *cleaning*, data *integration*, data *reduction*, dan data *transformation*. Selanjutnya, agar dapat menghasilkan bentuk data yang lebih baik, transformasi data dapat dilakukan dengan tujuan untuk melakukan normalisasi data yang berguna untuk menyamakan rentang nilai pada setiap atribut dengan skala tertentu.

Selanjutnya, proses klasifikasi akan dilakukan untuk mengelompokkan data atau objek baru ke dalam kelas atau label berdasarkan atribut-atribut tertentu. Proses ini bertujuan untuk memprediksi kelas dari suatu objek yang tidak diketahui sebelumnya. Klasifikasi terdiri dari tiga tahap dan dua fase, tiga tahap tersebut adalah pembangunan, penerapan, dan evaluasi model. Sedangkan dua fase tersebut adalah fase *training* dan fase *testing*. Fase *training* adalah fase di mana data digunakan untuk membangun sebuah model sedangkan fase *testing* adalah pengujian model yang telah dibuat dengan data lainnya untuk mengetahui akurasi dari model tersebut.

Berdasarkan permasalahan, pendekatan, dan metode yang telah dipaparkan sebelumnya, maka kami akan melakukan penelitian dengan data yang telah melalui tahapan *preprocessing* dengan tujuan untuk mengklasifikasikan kualitas jenis anggur putih dengan mengimplementasikan algoritma K-NN.

II. STUDI LITERATUR

1. Machine Learning

Machine Learning adalah salah satu cabang dari disiplin ilmu Kecerdasan Buatan (*Artificial Intelligence*) yang membahas mengenai pembangunan sistem yang berdasarkan pada data. Banyak hal yang dipelajari, akan tetapi pada dasarnya ada 4 hal pokok yang dipelajari dalam machine learning yaitu: Pembelajaran Terarah (*Supervised Learning*), Pembelajaran Tak Terarah (*Unsupervised Learning*), Pembelajaran Semi Terarah, (*Semi-supervised Learning*) dan Reinforcement Learning. [5]

2. Data Mining

Data mining adalah proses untuk menemukan korelasi, pola, dan tren baru yang bermakna dengan memilah-milah data dalam jumlah besar yang disimpan di dalam repositori, menggunakan teknologi pengenalan pola serta teknik statistik dan matematika. Data mining adalah proses menemukan pola dan pengetahuan menarik dari data dalam jumlah yang besar. Istilah data mining memiliki beberapa pandangan, seperti knowledge discovery ataupun pattern recognition. Istilah knowledge discovery atau penemuan pengetahuan tepat digunakan karena tujuan utama dari data mining memang untuk mendapatkan pengetahuan yang masih tersembunyi di dalam bongkahan data. Sedangkan istilah untuk pattern recognition atau pengenalan pola tepat untuk digunakan karena guna menemukan pola yang tersembunyi di dalam bongkahan data. [6]

3. Supervised Learning

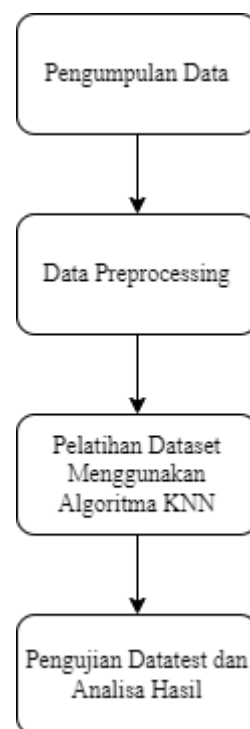
Supervised learning merupakan salah satu teknik machine learning yang menggunakan dataset (data training) yang sudah berlabel (labeled data) untuk melakukan pembelajaran pada mesin, sehingga mesin mampu mengidentifikasi label input dengan menggunakan fitur yang dimiliki untuk selanjutnya melakukan prediksi maupun klasifikasi, sedangkan unsupervised learning adalah teknik dengan menarik kesimpulan berdasarkan dataset yang merupakan input data labeled response. [7]

4. K-Nearest Neighbour (KNN)

Algoritma KNN adalah salah satu algoritma yang sudah populer. KNN ini termasuk ke dalam grup instance-based, learning. Metode KNN merupakan teknik lazy, learning [8]. Ada juga yang berpendapat bahwa algoritma KNN adalah algoritma pembelajaran yang banyak digunakan dalam sistem cyber-fisik sosial (CPSS) untuk menganalisis dan menambang data (*Data Mining*).

III. METODE

Adapun tahap-tahap penelitian yang dilakukan dapat digambarkan dengan diagram berikut :



Gambar 1. Metode Penelitian

1. Pengumpulan Data

Dataset yang digunakan pada penelitian ini terdiri dari 4898 data yang memiliki 12 atribut yaitu fixed acidity, volatile acidity, citric acidity, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol, dan quality. Untuk tahap pelatihan data, penelitian ini menggunakan 90% data tersebut dan sisanya digunakan untuk pengujian penelitian ini. Berikut adalah lampiran dataset yang digunakan :

	fixed acidity	volatile ac	citric acid	residual su	chlorides	free sulfur	total sulfu	density	pH	sulphates	alcohol	quality
1	7	0.027	0.036	20.07	0.045	45	170	1.001	3	0.045	0.08	6
2	06.03	0.003	0.034	01.06	0.049	14	132	0.690278	03.03	0.049	09.05	6
3	08.01	0.028	0.004	06.09	0.005	30	97	6.910417	03.26	0.044	10.01	6
4	07.02	0.023	0.032	08.05	0.058	47	186	6.913889	03.19	0.004	09.09	6
5	07.02	0.023	0.032	08.05	0.058	47	186	6.913889	03.19	0.004	09.09	6
6	08.01	0.028	0.004	06.09	0.005	30	97	6.910417	03.26	0.044	10.01	6
7	06.02	0.032	0.016	7	0.045	30	136	6.909028	03.18	0.047	09.06	6
8	7	0.027	0.036	20.07	0.045	45	170	1.001	3	0.045	0.08	6
9	06.03	0.003	0.034	01.06	0.049	14	132	0.690278	03.03	0.049	09.05	6
10	08.01	0.022	0.043	01.05	0.044	28	129	6.901389	03.22	0.045	11	6
11	08.01	0.027	0.041	01.45	0.033	11	63	6.880556	0.152083	0.056	12	5
12	08.06	0.023	0.004	04.02	0.035	17	109	6.907639	03.14	0.053	09.07	5
13	07.09	0.018	0.037	01.02	0.004	16	75	6.888889	03.18	0.04375	10.08	5
14	06.06	0.016	0.004	01.05	0.044	48	143	6.883333	03.54	0.052	12.04	7
15	08.03	0.042	0.043056	19.25	0.004	41	172	10.002	0.151389	0.046528	09.07	5
16	06.06	0.017	0.038	01.05	0.032	28	112	6.884722	03.25	0.055	11.04	7
17	06.03	0.048	0.004	01.01	0.046	30	99	6.894444	03.24	0.036	09.06	6
18	06.02	0.045833	0.048	01.02	0.029	29	75	6.869444	03.33	0.039	12.08	8
19	07.04	0.034	0.042	01.01	0.033	17	171	6.86806	03.12	0.053	11.03	6
20	06.05	0.031	0.014	07.05	0.044	34	133	6.913194	03.22	0.005	09.05	5
21	06.02	0.045833	0.048	01.02	0.029	29	75	6.869444	03.33	0.039	12.08	8
22	06.04	0.031	0.038	02.09	0.038	19	102	6.883333	03.17	0.035	11	7
23	06.08	0.026	0.042	01.07	0.049	41	122	6.889583	03.47	0.048	10.05	8
24	07.06	0.046528	0.014	01.05	0.074	25	168	6.900694	03.05	0.051	09.03	5
25	06.06	0.027	0.041	9	0.052	16	142	6.910417	03.42	0.047	10	6
26	7	0.025	0.032	9	0.046	56	245	6.913194	03.25	0.005	10.04	6
27	06.09	0.024	0.035	1	0.052	35	146	6.895583	03.45	0.044	10	6

29	7	0.028	0.039	0.087	0.051	32	141	6,917361	0.338	0.053	10.05	6
30	0.0704	0.027	0.048	0.101	0.047	17	132	6,884722	0.319	0.049	11.06	6
31	0.072	0.32	0.36	2	0.033	37	114	6,879167	0.301	0.049006	12.03	7
32	0.085	0.024	0.39	10.04	0.044	20	142	6,926389	0.302	0.053	10	6
33	0.083	0.14	0.34	0.101	0.042	7	47	6,898611	0.347	0.004	10.02	6
34	0.074	0.025	0.36	0.205	0.005	31	100	6,888889	0.319	0.044	10.08	6
35	0.062	0.12	0.34	0.105	0.045	43	117	6,902083	0.342	0.051	9	6
36	0.058	0.27	0.02	0.649306	0.044	22	179	6,918056	0.337	0.037	10.02	5
37	0.073	0.028	0.43	0.149	0.008	21	123	6,878472	0.319	0.042	12.08	5
38	0.065	0.39	0.23	0.504	0.051	25	149	6,898611	0.324	0.035	10	5
39	7	0.33	0.32	0.102	0.053	38	138	6,879167	0.313	0.028	11.02	6
40	0.073	0.024	0.39	0.774306	0.057	45	149	6,94375	0.321	0.036	0.086	5
41	0.073	0.024	0.39	0.774306	0.057	45	149	6,94375	0.321	0.036	0.086	5
42	0.067	0.23	0.39	0.205	0.119444	63	158	6,900694	0.311	0.036	0.094	6
43	0.067	0.024	0.39	0.209	0.120139	63	157	6,900694	0.301	0.034	0.094	6
44	7	0.31	0.26	0.704	0.069	28	160	6,9125	0.313	0.046	0.098	6
45	0.06	0.24	0.27	0.104	0.057	33	152	6,898611	0.322	0.056	0.095	6
46	0.067	0.23	0.26	0.104	0.006	33	154	6,898611	0.324	0.056	0.095	6
47	0.074	0.18	0.31	0.104	0.058	38	167	6,896528	0.316	0.053	10	7
48	0.062	0.45	0.26	0.044	0.063	63	206	6,902083	0.327	0.052	0.098	4
49	0.062	0.46	0.25	0.044	0.066	62	207	6,902083	0.325	0.052	0.098	5
50	7	0.31	0.26	0.704	0.069	28	160	6,9125	0.313	0.046	0.098	6
51	0.069	0.19	0.35	5	0.067	32	150	0,690972	0.336	0.048	0.098	5
52	0.072	0.19	0.31	0.106	0.062	31	173	6,888086	0.335	0.044	11.07	6
53	0.066	0.25	0.29	0.101	0.068	39	124	6,884722	0.34	0.058	11	7
54	0.062	0.16	0.33	0.101	0.057	21	82	6,688194	0.332	0.046	10.09	7
55	0.064	0.18	0.35	1	0.045	39	108	6,882639	0.31	0.035	10.09	6
56	0.068	0.02	0.59	0.009	0.102083	38	132	6,895853	0.305	0.038	0.091	6
57	0.069	0.25	0.35	0.103	0.039	29	191	6,880556	0.313	0.052	11	6
58	0.072	0.21	0.34	11.09	0.043	37	213	6,918056	0.309	0.005	0.096	6
59	6	0.19	0.26	12.04	0.048	50	147	6,925	0.303	0.036	0.089	6
60	0.066	0.08	0.35	0.106	0.044	25	78	6,896528	0.311	0.038	10.02	6
61	0.074	0.02	0.36	0.102	0.038	44	111	6,893056	0.336	0.034	0.099	6
62	0.068	0.22	0.24	0.049	0.092	30	123	6,910417	0.303	0.046	0.086	6
63	6	0.19	0.26	12.04	0.048	50	147	6,925	0.303	0.036	0.089	6
64	7	0.47	0.07	0.101	0.035	17	151	6,688194	0.302	0.034	10.05	5
65	0.066	0.38	0.15	0.046	0.044	25	78	6,896528	0.311	0.038	10.02	6
66	0.072	0.24	0.27	0.104	0.038	31	122	6,89375	0.315	0.046	10.03	6
67	0.062	0.35	0.03	0.102	0.064	29	120	6,898611	0.322	0.054	0.091	5
68	0.064	0.26	0.24	0.604	0.004	27	124	6,877083	0.322	0.049	12.06	7
69	0.067	0.25	0.13	0.102	0.041	81	174	6,888889	0.314	0.042	0.098	5
70	0.067	0.23	0.31	0.201	0.046	30	96	6,893056	0.333	0.044444	10.07	8
71	0.074	0.24	0.29	10.01	0.005	21	105	6,918056	0.313	0.035	0.095	6
72	0.062	0.27	0.43	0.708	0.056	48	244	6,913889	0.301	0.051	9	6
73	0.068	0.03	0.23	0.046	0.061	50.05	200	6,915278	0.332	0.006	0.095	5
74	6	0.27	0.28	0.048	0.063	31	201	6,919444	0.172917	0.049306	10	5
75	0.086	0.23	0.46	1	0.054	9	72	6,903472	0.149306	0.049	0.091	6
76	0.067	0.23	0.31	0.201	0.046	30	96	6,893056	0.333	0.044444	10.07	8
77	0.074	0.24	0.29	10.01	0.005	21	105	6,918056	0.313	0.035	0.095	6
78	0.071	0.18	0.36	0.104	0.043	31	87	6,873611	0.326	0.037	12.07	7
79	7	0.32	0.34	0.103	0.042	20	69	6,883333	0.331	0.045139	12	7
80	0.074	0.18	0.03	0.088	0.064	26	103	6,917361	0.148611	0.056	0.093	5
81	0.067	0.54	0.28	0.504	0.006	21	105	6,909028	0.327	0.037	9	5
82	0.068	0.22	0.31	0.104	0.053	34	114	6,895139	0.339	0.053472	10.06	6
83	0.071	0.02	0.34	16	0.005	51	166	6,934028	0.321	0.006	0.092	6
84	0.071	0.34	0.02	0.601	0.063	47	164	6,906944	0.317	0.042	10	5
85	0.073	0.22	0.03	0.082	0.047	42	207	6,920833	0.333	0.046	0.095	6
86	0.071	0.043	0.042361	11.08	0.045	54	155	6,926389	0.311	0.045	0.087	5
87	0.071	0.044	0.043056	11.08	0.044	52	152	6,927083	0.312	0.046	0.087	6
88	0.072	0.39	0.04375	11	0.044	55	156	6,926389	0.309	0.044	0.087	6
89	0.068	0.25	0.31	13.03	0.005	69	202	6,925	0.322	0.048	0.097	6
90	0.071	0.043	0.042361	11.08	0.045	54	155	6,926389	0.311	0.045	0.087	5
91	0.071	0.044	0.043056	11.08	0.044	52	152	6,927083	0.312	0.046	0.087	6
92	0.072	0.39	0.04375	11	0.044	55	156	6,926389	0.309	0.044	0.087	6
93	0.061	0.27	0.43	0.705	0.049	65	243	6,914583	0.312	0.047	9	5
94	0.069	0.24	0.33	0.107	0.035	47	136	0,06875	0.326	0.004	12.06	7
95	0.069	0.21	0.33	0.108	0.034	48	136	6,874306	0.325	0.041	12.06	7
96	0.075	0.17	0.32	0.107	0.004	51	148	6,886111	0.321	0.044	11.05	7
97	0.071	0.26	0.29	12.04	0.044	62	240	6,922917	0.304	0.042	0.092	6
98	6	0.34	0.045833	15.09	0.046	26	164	6,929861	0.314	0.005	0.088	6
99	0.086	0.184028	0.36	0.102	0.034	15	80	6,884028	0.149306	0.036	11.04	7
100	0.098	0.36	0.46	10.05	0.038	4	83	6,913889	0.145139	0.003	10.01	4
101	6	0.34	0.045833	15.09	0.046	26	164	6,929861	0.314	0.005	0.088	6
102	0.074	0.25	0.37	13.05	0.006	52	192	6,927083	3	0.044	0.091	5
103	0.071	0.12	0.32	0.096	0.054	64	162	6,918056	0.304	0.041	0.094	5
104	6	0.21	0.24	12.01	0.005	55	164	0,692361	0.334	0.039	0.094	5
105	0.075	0.211806	0.04	18.09	0.059	44	170	1	0.152083	0.046	9	5
106	0.074	0.25	0.37	13.05	0.006	52	192	6,927083	3	0.044	0.091	5
107	0.073	0.13	0.32	14.04	0.051	34	109	6,926389	0.302	0.035	0.092	6
108	0.071	0.12	0.32	0.096	0.054	64	162	6,918056	0.304	0.041	0.094	5
109	0.071	0.23	0.35	16.05	0.004	60	171	0,69375	0.316	0.059	0.091	6
110	0.071	0.23	0.35	16.05	0.004	60	171	0,69375	0.316	0.059	0.091	6
111	0.069	0.33	0.28	0.103	0.051	37	187	6,89375	0.327	0.006	10.03	5
112	0.065	0.17	0.54	0.085	0.082	64	163	6,915972	0.145139	0.039	0.088	6
113	0.072	0.27	0.46	0.802083	0.052	45	255	1	0.304	0.052	0.089	5
114	0.072	0.31	0.05	13.03	0.056	68	195	6,931944	0.301	0.047	0.092	5
115	0.067	0.41	0.34	0.902	0.049	29	150	6,922222	0.322	0.051	0.091	5
116	0.067	0.41	0.34	0.902	0.049	29	150	6,922222	0.322	0.051	0.091	5
117	0.055	0.336806	0	0.105	0.065	8	103	0,690278	0.16875	0.004	0.097	4
118	6	0.31	0.24	0.033	0.041	25	143	6,884722	0.331	0.044	11.03	6
119	7	0.14	0.04	0.107	0.035	16	85	6,882639	0.319	0.042	11.08	6
120	0.072	0.31	0.05	13.03	0.056	68	195	6,931944	0.301	0.047	0.092	5
121	0.073	0.32	0.48	13.03	0.006	57	196	6,931944	0.304	0.005	0.092	5
122	0.059	0.36	0.04	0.507	0.046	21	87	6,898611	0.322	0.051	10.02	5
123	0.078	0.24	0.32	12.02	0.054	42	138	6,933333	0.301	0.054	0.088	5
124	0.074	0.16	0.31	0.309028	0.059	31	131	6,911111	0.329	0.034	0.097	5
125	0.069	0.19	0.28	5	0.058	14	146	6,911111	0.329	0.036	0.091	6
126	0.064	0.13	0.47	0.106	0.092	40	158	6,894444	0.321	0.036	0.098	6
127	0.067	0.19	0.36	0.101	0.026	63	143	6,883333	0.327	0.048	11	6
128	0.074	0.39	0.23	7	0.033	29	126	0				

Algoritma K-Nearest Neighbor (K-NN) adalah sebuah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasi sebelumnya. Termasuk dalam supervised learning, dimana hasil query instance yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam K-NN.

1. Menentukan Parameter K (Jumlah tetangga paling dekat).
2. Menghitung jarak antara data yang akan dievaluasi dengan semua pelatihan.
3. Mengurutkan jarak yang terbentuk.
4. Menentukan jarak terdekat sampai urutan K.
5. Memasangkan kelas yang bersesuaian.
6. Mencari jumlah kelas dari tetangga yang terdekat dan menetapkan kelas tersebut sebagai kelas data yang akan dievaluasi

$$d(P, Q) = \sqrt{\sum_{i=1}^n (P_i - Q_i)^2}$$

$d(P,Q)$: Jarak Euclidean

p : Inputan ke-1 dari data training

IV. HASIL DAN PEMBAHASAN

Proses klasifikasi diawali dari tahapan preprocessing dengan metode transformasi data yang bertujuan untuk mengurangi

```

# Search Top Correlation Pairs
correlated_pairs = wine.cor()
top_correlated_pairs = correlated_pairs.unstack().sort_values(kind="paircount")
pd.DataFrame(top_correlated_pairs[(top_correlated_pairs > 0.5) & (top_correlated_pairs < 1)]) # (top_correlated_pairs < 1) & (top_correlated_pairs > 0.5), columns = [
Correlation
Coefficient
freedomofdislike totalbufrdislike 0.819501
totalbufrdislike freedomofdislike 0.819501

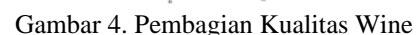
# Dropping Non Used Attribute
wine.drop(columns = ["freedomofdislike", "totalbufrdislike"], inplace = True)

# Show The First 10 Value Of Datasets after Non Used Attribute Dropped
print(wine.shape)
wine.head(10)

# wine, 10)
freedomofdislike volatiliticity citricacid residual sugars chlorides density pH sulphates alcohol quality
0 0.5 0.27 0.36 25.7 0.046 1.010 3.00 0.46 8.8 6.0
1 0.3 0.30 0.33 1.8 0.048 0.930 0.45 9.5 6
2 0.1 0.28 0.40 6.6 0.060 0.960 0.36 0.44 10.1 6

```

Dataset yang digunakan pada penelitian ini memungkinkan pembagian atribut kualitas wine menjadi dua jenis, atribut kualitas ini yang kemudian akan digunakan sebagai parameter dari kualitas wine tersebut, apakah wine dalam dataset ini berkualitas tinggi ataupun rendah. Visualisasi data menunjukkan bahwa wine putih yang berkualitas tinggi ada berjumlah 3258 dan sebanyak lainnya 1640 berkualitas rendah.



```
[ ] # Splitting dataset into features & target variable
X = np.array(aline_data)
y = np.array(aline_label)

# Training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3)

[ ] print(X_train)
print(y_train)
```

```
[ ] [[0.9927  3.39   0.45]
      [0.9917  3.28   0.46]
      [0.9914  3.67   0.47]
      ...
      [0.9916  3.27   0.56]
      [0.9908  3.39   0.55]
      [0.9906  3.23   0.48]
      ...
      [0.9901  3.26   0.55]]
```

Langkah selanjutnya adalah menentukan kualitas data test yang merupakan 10% bagian dataset yang telah didefinisikan sebelumnya. Penentuan kualitas ini dilakukan menggunakan metode KNN dengan nilai $K=3$. Hasil kualitas dari data test wine putih dan penerapan metode KNN yang digunakan adalah sebagai berikut.

Gambar 6. Hasil Penelitian

```

KNN Classification Algorithm
class KNN:
    def __init__(self, k=1):
        self.k = k

    def train(self, X_train, y_train):
        self.X_train = X_train
        self.y_train = y_train

    def predict(self, X_test):
        y_pred = []
        for x in X_test:
            distances = []
            for i in range(len(self.X_train)):
                distance = self.distance(x, self.X_train[i])
                distances.append(distance)
            sorted_distances = sorted(distances)
            k_nearest_indices = sorted(range(len(distances)), key=lambda i: sorted_distances[i])
            k_nearest_labels = [self.y_train[i] for i in k_nearest_indices]
            most_common_label = Counter(k_nearest_labels).most_common(1)[0][0]
            y_pred.append(most_common_label)
        return y_pred

    def distance(self, x1, x2):
        return np.sqrt(np.sum((x1 - x2)**2))

```

Gambar 7. Metode KNN

Proses klasifikasi ini selanjutnya akan dinilai dan menunjukkan tingkat akurasi sebagai berikut.

```

Accuracy Test
accuracy = 0
for i in range(len(X_test)):
    print("Accuracy level: ", accuracy)
    accuracy += 1 if y_pred[i] == y_test[i] else 0
Accuracy level: 0.6400000000000001

```

Gambar 8. Akurasi Penelitian

V. KESIMPULAN DAN SARAN

A. Kesimpulan dari penelitian klasifikasi “Klasifikasi Kualitas Anggur Putih Menggunakan Metode K-Nearest Neighbor (KNN)” ini adalah sebagai berikut :

1. Penelitian ini bertujuan untuk mengklasifikasikan kualitas wine putih berdasarkan atribut yang dimilikinya
2. Klasifikasi wine putih pada penelitian ini dilakukan dengan mengimplementasikan pendekatan supervised learning dengan metode K-Nearest Neighbor (KNN)
3. Penelitian ini mengimplementasikan proses transformasi berupa penghilangan atribut yang tidak diperlukan sebelum data diklasifikasikan
4. Penelitian ini membagi dataset menjadi dua bagian, 90% digunakan sebagai data latih dan 10% sisanya digunakan sebagai data test dari metode KNN
5. Proses klasifikasi yang dilakukan menunjukkan keberhasilan dengan tingkat akurasi sebesar 64%

B. Saran dari penelitian “Klasifikasi Kualitas Anggur Putih Menggunakan Metode K-Nearest Neighbor (KNN)” ini adalah untuk mengembangkan penelitian dengan algoritma KNN yang lebih optimal sehingga dapat menghasilkan tingkat akurasi penelitian yang lebih baik.

VI. DAFTAR PUSTAKA

- [1] D. A. Nasution, H. H. Khotimah, and N. Chamidah, “Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN,” *Comput. Eng. Sci. Syst. J.*, vol. 4, no. 1, p. 78, 2019, doi: 10.24114/cess.v4i1.11458.
- [2] K. H. Tandrio, K. Aprilia, P. Nuzry, Y. P. Ardi, H. Prasetyo, and A. Wine, “Classification of Wine Types Based on Composition Using Backpropagation And Particle Swarm Optimization,” *J. Ilm. Teknol. dan Inf.*, vol. 7, no. 1, pp. 16–21, 2018.
- [3] A. A. Fauzi, F. Utaminigrum, and F. Ramdani, “Road surface classification based on LBP and GLCM features using KNN classifier,” *Bull. Electr. Eng. Informatics*, vol. 9, no. 4, pp. 1446–1453, 2020, doi: 10.11591/eei.v9i4.2348.
- [4] M. Muhathir, S. Theofil Tri Saputra, and A.-K. Al-Khowarizmi, “Analysis K-Nearest Neighbors (KNN) in Identifying Tuberculosis Disease (Tb) By Utilizing Hog Feature Extraction,” *Al’adziya Int. Comput. Sci. Inf. Technol. J.*, vol. 1, no. 1, pp. 33–38, 2020.
- [5] A. U. Zailani, A. Perdananto, Nurjaya, and Sholihin., “PENGENALAN SEJAK DINI SISWA SMP TENTANG MACHINE LEARNING UNTUK KLASIFIKASI GAMBAR DALAM MENGHADAPI KOMMAS : Jurnal Pengabdian Kepada Masyarakat,” *KOMMAS J. Pengabd. Kpd. Masy.*, vol. 1, no. 1, pp. 7–15, 2020.
- [6] Z. Nabila, A. Rahman Isnain, and Z. Abidin, “Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means,” *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>.
- [7] E. Retnoningsih and R. Pramudita, “Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python,” *Bina Insa. Ict J.*, vol. 7, no. 2, p. 156, 2020, doi: 10.51211/biict.v7i2.1422.
- [8] D. Cahyanti, A. Rahmayani, and S. A. Husniar, “Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara,” *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 39–43, 2020, doi: 10.33096/ijodas.v1i2.13.
- [9] H. Ma’rifah, A. P. Wibawa, and M. I. Akbar, “Klasifikasi Artikel Ilmiah Dengan Berbagai Skenario Preprocessing,” *Sains, Apl. Komputasi dan Teknol. Inf.*, vol. 2, no. 2, p. 70, 2020, doi: 10.30872/jsakti.v2i2.2681.
- [10] A. Rahmat Dian Nugraha, K. Auliasari, and Y. Agus Pranoto, “IMPLEMENTASI METODE K-NEAREST NEIGHBOR (KNN) UNTUK SELEKSI CALON KARYAWAN BARU (Studi Kasus : BFI Finance Surabaya),” *JATI (Jurnal Mhs.*

Tek. Inform., vol. 4, no. 2, pp. 14–20, 2020, doi:
10.36040/jati.v4i2.2656.