

Komparasi Metode K-Nearest Neighbor (Knn) Dengan Support Vector Machine (Svm) Terhadap Tingkat Akurasi Klasifikasi Kualitas Air

Jauhari Maulani*¹, Mayang Sari²

^{1,2}Fakultas Teknologi Informasi, Universitas Islam Kalimantan MAB Banjarmasin

E-mail: *¹ jauharimaulani@gmail.com, ² mayangsari.si@gmail.com

Abstrak

Air adalah senyawa kimia yang menjadi bahan dasar kehidupan dan berlimpah di dalam makhluk hidup, semua organisme yang hidup mengandung air sekitar 55-78%. Air dalam organisme hidup berperan penting sebagai pelarut agar dapat terjadi reaksi kimia dalam sel hidup yang dikenal dengan istilah metabolisme. Peranan air yang begitu besar di dalam tubuh, maka keseimbangan air di dalam tubuh perlu dijaga. Kualitas air adalah suatu ukuran kondisi air dilihat dari karakteristik fisik, kimiawi, dan biologisnya. Kualitas air juga menunjukkan ukuran kondisi air relatif terhadap kebutuhan biota air dan manusia. Algoritma K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM) merupakan salah satu metode yang terdapat dalam klasifikasi. Salah satu cara untuk mendeteksi klasifikasi kualitas air yang bisa dikonsumsi dalam machine learning yaitu menggunakan dataset sebagai data latih agar dapat dilakukan pengujian performa dengan metode klasifikasi yang tepat. Metode yang digunakan dalam penelitian ini yaitu melakukan komparasi antara 2 buah algoritma, dimana salah satu metode akan mendapatkan nilai yang paling akurat. Berdasarkan hasil pengujian performansi algoritma KNN dengan algoritma SVM, bahwa menunjukkan nilai akurasi SVM yaitu sekitar 69%, sedangkan hasil akurasi KNN bernilai 66%, sehingga hasil akurasi SVM lebih baik daripada KNN. Anda dapat menjalankan hasil prediksi biologis untuk melihat apakah kualitas air yang dapat Anda konsumsi sesuai standar.

Kata Kunci—3-5 Kualitas Air, Klasifikasi, K-Nearest Neighbor, Akurasi, KNN, SVM

1. PENDAHULUAN

Air adalah komponen dasar kehidupan dan senyawa yang melimpah dalam organisme hidup, dengan semua organisme hidup mengandung sekitar 55-78% air [1]. Air dalam organisme hidup berperan penting sebagai pelarut agar dapat terjadi reaksi kimia dalam sel hidup yang dikenal dengan istilah metabolisme [2]. Peranan air yang begitu besar di dalam tubuh, maka keseimbangan air di dalam tubuh perlu dijaga. Kualitas air adalah suatu ukuran kondisi air dilihat dari karakteristik fisik, kimiawi, dan biologisnya [3]. Kualitas air juga menunjukkan ukuran kondisi air relatif terhadap kebutuhan biota air dan manusia. Algoritma K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM) merupakan salah satu metode yang terdapat dalam klasifikasi. Salah satu cara untuk mendeteksi klasifikasi kualitas air yang bisa dikonsumsi dalam machine learning yaitu menggunakan dataset sebagai data latih agar dapat dilakukan pengujian performa dengan metode klasifikasi yang tepat [4], [5], [6]

Referensi yang berkaitan dengan penelitian ini seperti berikut menjelaskan bahwa Kualitas air yang diuji diklasifikasikan menjadi layak dan tidak layak. Berdasarkan hasil pengujian diperoleh akurasi dengan KNN lebih baik yaitu 88% dibandingkan dengan KMeans yaitu 76% [1]. Selanjutnya penelitian lain menjelaskan bahwa Hasil akurasi sebesar 92% dengan nilai $K = 5$ [7].

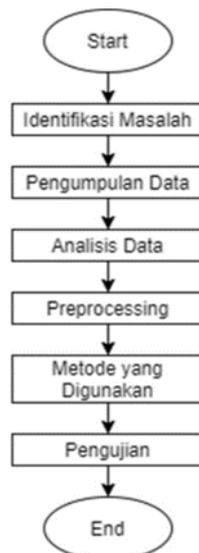
Studi ini mengklasifikasikan kualitas air yang mana dapat dikonsumsi atau tidak, data yang dikumpulkan bersifat universal (diperoleh dari berbagai entitas) yang diperoleh di situs web Kaggle, dan lebih jauh lagi, perbedaan antara penelitian ini dan penelitian lain ada dalam data dan sifat kontennya. kumpulan data berkualitas yang disimpan di permukaan dalam format kumpulan data ini [4]

Dalam penelitian ini menggunakan metode Klasifikasi K-Nearest Neighbor yang di komparasi dengan metode SVM dengan beberapa fitur yang ada agar klasifikasi yang mana bisa lebih akurat dan spesifik [8]

Algoritma K-Nearest Neighbor dan SVM digunakan atau dipilih untuk mengolah data awal menjadi model yang nantinya dapat diprediksi dan untuk menghasilkan skor akhir untuk algoritma menggunakan tabel confusion matrix yang mendasari pemilihan berbasis algoritma ini [9].

2. METODE PENELITIAN

Untuk mengklasifikasi kualitas air ada beberapa pendekatan penelitian menggunakan metode kuantitatif, lalu pada tahap metode penelitian yang akan dilakukan adalah dengan pengumpulan data dengan fitur terkait, pengolahan data awal (pre-processing data), lalu melakukan penerapan dengan algoritma KNN dan SVM pada dataset, dan gambaran penelitian yang dilakukan seperti gambar 1 dibawah ini [10]



Gambar 1. Tahap Metode Penelitian [11]

Karakteristik dataset terdiri dari 31 variabel. 30 fitur yang dijadikan variabel X atau disebut variabel dependen dan 1 variabel yang dijadikan variabel Y atau variabel independen, dan data sebanyak 3276.

Pengolahan Data dimulai dari identifikasi dataset yaitu memiliki tipe data integer dan data float. Setelah melakukan analisa data, tahap selanjutnya melakukan preproccessing, kemudian melakukan split dataset yaitu pembagian data training dan data testing, seperti yang terlihat pada gambar berikut [12]:



Gambar 2. Pengolahan Data [12]

3. HASIL DAN PEMBAHASAN

3.1. Tahap Pengolahan Data

Pengumpulan dataset dengan cara pengunduhan pada API Command kaggle (kaggle datasets download -d adityakadiwal/water-potability).

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3276 entries, 0 to 3275
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   ph                     2785 non-null   float64
1   Hardness               3276 non-null   float64
2   Solids                 3276 non-null   float64
3   Chloramines           3276 non-null   float64
4   Sulfate                2495 non-null   float64
5   Conductivity          3276 non-null   float64
6   Organic_carbon        3276 non-null   float64
7   Trihalomethanes       3114 non-null   float64
8   Turbidity              3276 non-null   float64
9   Potability             3276 non-null   int64
dtypes: float64(9), int64(1)
memory usage: 256.1 KB
    
```

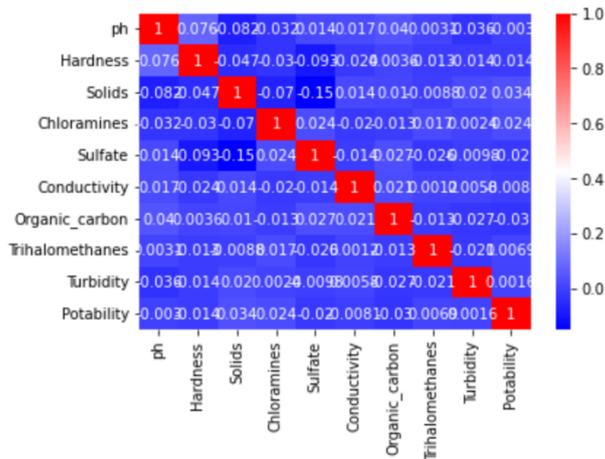
Gambar 3. Dataset

Pengolahan Data dimulai dari identifikasi dataset yaitu memiliki tipe data integer dan data float. Setelah melakukan analisa data, tahap selanjutnya melakukan preprocessing, kemudian melakukan split dataset yaitu pembagian data training dan data testing, seperti yang terlihat pada gambar berikut :

	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic_carbon	Trihalomethanes	Turbidity	Potability
0	True	False	False	False	False	False	False	False	False	False
1	False	False	False	False	True	False	False	False	False	False
2	False	False	False	False	True	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False
...
3271	False	False	False	False	False	False	False	False	False	False
3272	False	False	False	False	True	False	False	True	False	False
3273	False	False	False	False	True	False	False	False	False	False
3274	False	False	False	False	True	False	False	False	False	False
3275	False	False	False	False	True	False	False	False	False	False

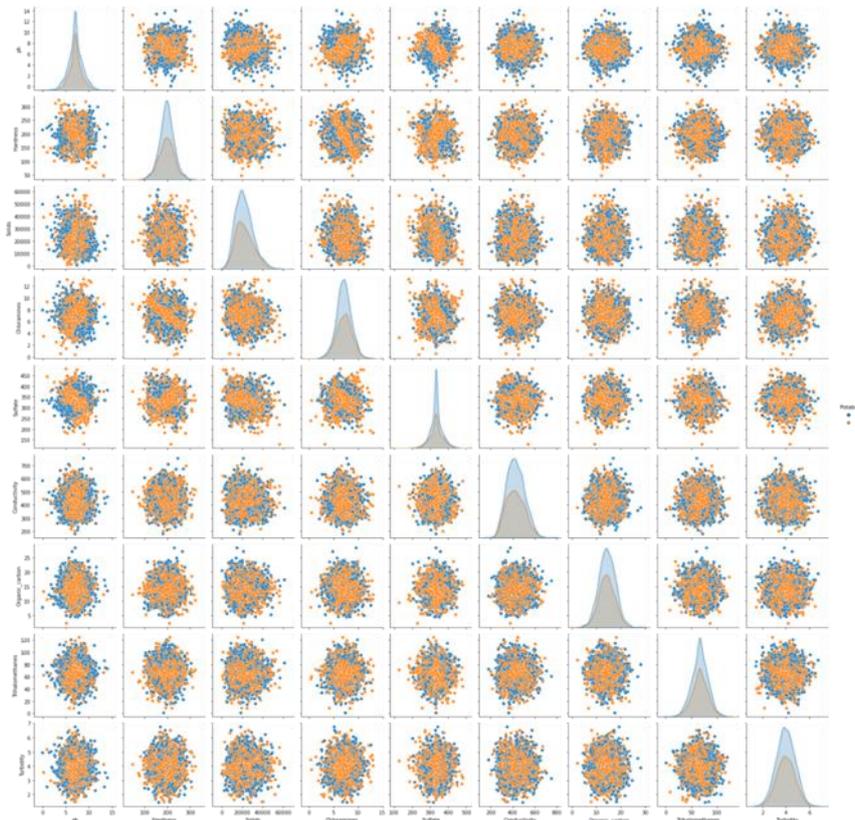
3276 rows x 10 columns

Gambar 4. Pengolahan Data



Gambar 5. heatmap korelasi antar fitur

Dari data visualisasi gambar 4 dapat dilihat bahwa semakin pekat warnanya maka semakin besar korelasi antar fitur tersebut dan semakin mempengaruhi akurasi analisa kualitas air.



Gambar 6. pairplot kategori data

Pada gambar 5 menjelaskan visualisasi berbentuk pairplot untuk mengkategorikan data, kuning melambangkan kategori air yang sehat, dan biru melambangkan kategori air yang tidak sehat. Dari gambar tersebut terlihat bahwa kualitas air tidak sehat lebih dominan dibandingkan dengan kualitas air yang tidak sehat.

3.2. Tahap Pengujian training model menggunakan training data yang sudah di split dan menentukan nilai akurasi

	precision	recall	f1-score	support
0	0.68	0.93	0.79	412
1	0.70	0.27	0.39	244
accuracy			0.69	656
macro avg	0.69	0.60	0.59	656
weighted avg	0.69	0.69	0.64	656

Gambar 7. Hasil akurasi SVM

```

: from sklearn.metrics import classification_report
  y_pred=knn1.predict(xtes_scaled)
  print(classification_report(ytes,y_pred))

```

	precision	recall	f1-score	support
0	0.65	0.95	0.77	394
1	0.73	0.22	0.34	262
accuracy			0.66	656
macro avg	0.69	0.58	0.55	656
weighted avg	0.68	0.66	0.59	656

Gambar 7. Hasil akurasi KNN metric manhatan

Pada gambar 6 dan gambar 7 terlihat hasil komparasi yang berbeda, pada gambar 6 menunjukkan nilai akurasi SVM yaitu sekitar 69%, sedangkan hasil akurasi KNN pada gambar 7 bernilai 66%, sehingga hasil akurasi SVM lebih baik daripada KNN.

4. KESIMPULAN

Berdasarkan hasil perhitungan performansi algoritma KNN dengan algoritma SVM, Dari hasil running di atas, menunjukkan nilai akurasi SVM yaitu sekitar 69%, sedangkan hasil akurasi KNN bernilai 66%, sehingga hasil akurasi SVM lebih baik daripada KNN. Anda dapat menjalankan hasil prediksi biologis untuk melihat apakah kualitas air yang dapat Anda konsumsi sesuai standar. Hal ini sangat baik untuk prediksi dimana nantinya kesalahan prediksi baru meminimalkan data.

DAFTAR PUSTAKA

- [1] N. Nurmahaludin, G. C.-S. N. Riset, and undefined 2019, "Klasifikasi Kualitas Air PDAM Menggunakan Algoritma KNN Dan K-Means," repository.poliban.ac.id.
- [2] S. Rahayu, P. Sumber Daya Air Jl IrH Juanda No, and P. H. Sumber Daya Air Jl Ir Juanda No, "PENELITIAN KUALITAS AIR BENGAWAN SOLO PADA SAAT MUSIM KEMARAU."
- [3] M. Nuzapril, S. Susilo, J. P.-J. Segara, and undefined 2019, "Sebaran produktivitas primer kaitannya dengan kondisi kualitas air di perairan Karimun Jawa," ejournal-balitbang.kkp.go.id.
- [4] I. G. Vidiastanta, N. Hidayat, and R. K. Dewi, "Komparasi Metode K-Nearest Neighbors (K-NN) Dengan Support Vector Machine (SVM) Untuk Klasifikasi Status Kualitas Air," 2020.

- [5] M. Ruswandi, D. I. Mulyana, and A. Awaludin, "Optimasi Klasifikasi Kematangan Buah Alpukat Menggunakan KNN dan Fitur Statistik," *Smart Comp Jurnalnya Orang Pint. Komput.*, vol. 11, no. 2, pp. 210–219, Apr. 2022, doi: 10.30591/SMARTCOMP.V11I2.3531.
- [6] A. C. Barus, T. M. Panggabean, D. Pakpahan, and S. G. D. Sirait, "Verifikasi Kualitas Gambar Dengan Algoritma Support Vector Machine (SVM) Untuk Studi Kasus Ulos Batak Toba," *Smart Comp Jurnalnya Orang Pint. Komput.*, vol. 11, no. 3, pp. 473–483, Jul. 2022, doi: 10.30591/SMARTCOMP.V11I3.3900.
- [7] C. Muhammad, R. Maulana, M. Hannats, and H. Ichsan, "Purwarupa Perahu untuk Monitoring dan Klasifikasi Kualitas Air Bendungan dengan Metode K-Nearest Neighbor (KNN)," vol. 4, no. 2, pp. 651–659, 2020.
- [8] S. Aulia, S. Hadiyoso, and D. Nur Ramadan, "Analisis Perbandingan KNN dengan SVM untuk Klasifikasi Penyakit Diabetes Retinopati berdasarkan Citra Eksudat dan Mikroaneurisma."
- [9] K. Aditya Pratama, W. Priyo Atmaja, and V. Lusiana, "Klasifikasi Tingkat Kematangan Buah Kersen Menggunakan Citra HSI Dengan Metode K-Nearest Neighbor (KNN)," *Smart Comp Jurnalnya Orang Pint. Komput.*, vol. 11, no. 1, pp. 105–108, Jan. 2022, doi: 10.30591/smartcomp.v11i1.3171.
- [10] Y. Nurdiansyah, "Informal : informatics journal.," vol. 2, no. 2, pp. 114–122, Jul. 2017.
- [11] H. Situs et al., "Model Prediksi Kualitas Udara dengan Support Vector Machines dengan Optimasi Hyperparameter GridSearch CV," *Bul. Ilm. Sarj. Tek. Elektro*, vol. 4, no. 1, pp. 12–21, 2022, doi: 10.12928/biste.v4i1.6079.
- [12] A. Ali, M. Alrubei, L. F. M. Hassan, M. Al-Ja'afari, and S. Abdulwahed, "Diabetes classification based on KNN," *IJUM Eng. J.*, vol. 21, no. 1, pp. 175–181, 2020, doi: 10.31436/iiumej.v21i1.1206.