

Sistem Prediksi Kualitas Air Yang Dapat Dikonsumsi Dengan Menerapkan Algoritma K-Nearest Neighbor

Hardiana Said¹, Nur Hafifah Matondang², Helena Nurramdhani Irmanda³,

^{1,2,3}S1 Sistem Informasi / Fakultas Ilmu Komputer

Universitas Pembangunan Nasional Veteran Jakarta

Jl. RS. Fatmawati Raya, Pd. Labu, Kec. Cilandak, Kota Depok,

Daerah Khusus Ibukota Jakarta (12450)

¹hardiana@upnvj.ac.id, ²nurhafifahmatondang@upnvj.ac.id, ³helenairmanda@upnvj.ac.id.

Abstrak. Dengan kualitas air diberbagai wilayah yang semakin menurun, hal ini mengkhawatirkan sebab kualitas air yang aman sangat penting bagi kesehatan masyarakat luas. Hal tersebut juga berpengaruh terhadap kebutuhan manusia terkait air minum, kualitas air yang tidak aman yaitu biasanya memiliki sanitasi yang buruk dan memiliki kandungan dengan zat-zat yang melebihi kadar standar dapat menyebabkan penyakit seperti hepatitis A, kolera, diare, dan lainnya. Penelitian ini dilakukan untuk membuat sebuah sistem prediksi kualitas air dan untuk mendukung dilakukannya pemrosesan prediksi yaitu dengan menerapkan algoritma klasifikasi data mining yaitu adalah algoritma *K-Nearest Neighbor*. Data yang diperoleh berasal dari website kaggle tersebut dijadikan sebuah model dan diterapkan kedalam sistem prediksi sehingga sistem dapat memprediksi kualitas air. Hasil pemodelan diukur menggunakan tabel *Confusion Matrix* sehingga dapat dihitung nilai akurasi dari performa model. Setelah dilakukan pengujian, model yang ada memiliki tingkat akurasi tertinggi 85,24% dengan nilai k (tetangga terdekat) = 3. Aplikasi berbasis web yang telah dibangun diharapkan dapat memprediksi kualitas air yang dapat dikonsumsi berdasarkan hasil pemodelan.

Kata Kunci: Kualitas Air, Klasifikasi, Sistem, Prediksi, *K-Nearest Neighbor*.

1 Pendahuluan

Air merupakan sebuah sumber kehidupan bagi setiap makhluk hidup dan menjadi kebutuhan penting bagi setiap manusia, khususnya masyarakat diwilayah dengan tingkat penduduk yang tinggi (perkotaan). Kebutuhan air bersih dari tahun ke tahun mengalami peningkatan namun ketersediaan air bersih malah terbatas, terutama di daerah perkotaan dengan adanya pembangunan yang semakin meluas dan jumlahnya yang meningkat tanpa memperhatikan keseimbangan lingkungan sekitar dan sempitnya daerah resapan yang mana hal tersebut menimbulkan masalah yang cukup serius yaitu ketersediaan sumber air bersih yang berkurang. Pencemaran air merupakan masalah utama yang ada di Indonesia, dan penyebab utama dari pencemaran air di Indonesia sendiri banyak dihasilkan dari limbah domestik dan limbah perusahaan yang kurang baik dalam pengolahan limbah, akibatnya penurunan kualitas air dan ketersediaan air bersih terjadi di setiap wilayah. Dampak yang ditimbulkan dari kualitas air yang buruk dapat menimbulkan penyakit pencernaan yang cukup serius menurut [1] diperkirakan sekitar 800.000 orang meninggal setiap tahun akibat diare akibat konsumsi air kotor atau terkontaminasi, kebersihan tangan dalam konsumsi air, dan masalah kebersihan yang ada pada air minum.

Identifikasi dini yang dapat digunakan sebagai informasi awal diperlukan untuk menentukan kualitas air yang layak dikonsumsi oleh masyarakat luas. Dalam mengidentifikasi masalah awal ini dapat menggunakan teknik *data mining*, dalam penelitian ini metode klasifikasi yang digunakan adalah algoritma klasifikasi yang dapat digunakan untuk memprediksi salah satunya adalah *K-Nearest Neighbor*. Dengan memilih teknik klasifikasi *data mining* dengan menggunakan algoritma *K-Nearest Neighbor*, dikarenakan metode klasifikasi dengan algoritma *K-Nearest Neighbor* ini dianggap lebih tepat dengan perhitungan akurasi tinggi yaitu sebesar sehingga cukup akurat dalam melakukan prediksi, berdasarkan nilai akurasi yang dapat dihasilkan cukup tinggi dengan presentase 85.24% dan perbandingan nilai *recall* dan *precision* yang lebih stabil.

Berdasarkan hal-hal yang telah dijelaskan diatas maka tujuan yang ingin dicapai adalah membangun sebuah sistem prediksi untuk dapat membantu peneliti air lainnya atau masyarakat umum dalam menentukan kualitas air

yang baik dikonsumsi atau tidaknya berdasarkan model data yang didapatkan dari pengolahan model menggunakan algoritma KNN dalam *data mining*.

2 Tinjauan Pustaka

2.1 Sistem Prediksi

Sistem merupakan kesatuan unsur atau komponen yang saling berhubungan untuk memudahkan perpindahan materi, energi, atau informasi. Sistem ini disebut juga sebagai satu komponen yang saling berhubungan dan memiliki elemen penggerak. Secara umum unsur-unsur yang membentuk suatu sistem dan perlu diketahui adalah tujuan, input, proses, dan output. Ada juga faktor lain seperti batasan, mekanisme kontrol dan umpan balik, lingkungan sistem, dan lainnya [2]. Prediksi adalah proses memprediksi variabel masa depan yang lebih intuitif daripada data historis, tetapi lebih intuitif, data kuantitatif sering digunakan dalam peramalan harga sebagai pelengkap peramalan [3]. Maka dapat disimpulkan sistem prediksi adalah suatu kesatuan unsur atau komponen yang memiliki tujuan untuk dapat memproses variabel tertentu yang digunakan untuk memprediksi variabel baru berdasarkan variabel yang telah diolah atau ada.

2.2 SDLC *Waterfall*

SDLC (Software Development Life Cycle) merupakan proses pembuatan atau siklus pengembangan sistem, dengan tujuan sistem yang dibangun ini dapat berjalan sesuai harapan. Pada SDLC terdapat banyak jenis metode pembangunan sistem yang dapat dimanfaatkan setiap jenis metode memiliki kelebihan dan kekurangannya masing-masing, namun dalam penelitian ini memilih proses pengembangan sistem dengan metode *waterfall*. Metode *waterfall* adalah suatu proses perangkat lunak yang memiliki proses kerja berurutan dan sistematis layaknya air terjun, sehingga metode ini menekankan pada setiap proses, dari proses awal hingga akhir yang mengalir ke satu arah “kebawah” atau per-fase. Menurut [6], tahapan dari metode *waterfall* dapat dilihat secara detail pada penjelasan dibawah ini, berikut:

- a. *Requirement Gathering & Analysis* (Analisis Kebutuhan Sistem), pada tahap ini yaitu melakukan analisis dan mendefinisikan data untuk kebutuhan yang harus dipenuhi untuk program yang nanti akan dibutuhkan dalam proses pembangunan sistem.
- b. *Design* (Desain Sistem), ditahap kedua ini yaitu melakukan perancangan desain sistem yang akan dibuatnya, beberapa alat bantu dari penvisualisasian desain sistem itu sendiri dapat menggunakan DFD, UML Diagram dan lainnya.
- c. *Implementation* (Implementasi Sistem), pada tahap ketiga ini desain-desain yang telah dirancang sebelumnya diubah menjadi kode-kode program.
- d. *Testing* (Pengujian Sistem), ditahap pengujian ini modul-modul yang dibangun sebelumnya akan langsung dilakukan testing sehingga penguji dapat mengetahui apakah sistem yang dibangun sesuai dengan desain dan fungsi utamanya.
- e. *Deployment & Maintenance System*, pada tahap terakhir ini sistem disebar dan dilakukan perawatan untuk mengurangi adanya *error* dan *bug* pada saat digunakan oleh user secara umum.

2.3 *Data Mining*

Dalam *Knowledge Discovery in Database*, *data mining* sering disebut sebagai penemuan pengetahuan. KDD adalah aktivitas yang melibatkan pengumpulan dan penggunaan data historis untuk menemukan keteraturan, pola, atau hubungan dalam kumpulan data yang besar [4]. Aktivitas ini melakukan sebuah proses pencarian pola dengan menggunakan teknik statistik dan matematik dari *record data* yang tersimpan dalam basis data ataupun data secara keseluruhan (*big data*), pola yang ditemukan akan dijadikan sebagai pengambilan keputusannya atau informasi yang mana dapat memberikan manfaat dan dapat juga dijadikan pedoman dalam upaya pemeliharaan dan pelestarian kualitas air untuk dapat dikonsumsi kedepannya.

2.4 K-Nearest Neighbor

K-Nearest Neighbor (K-NN) adalah algoritma *supervised* yang digunakan untuk mengklasifikasikan hasil sampel uji baru berdasarkan sebagian besar kategori K-NN. Tujuan dari algoritma ini adalah untuk mengklasifikasikan objek baru berdasarkan atribut dan contoh pelatihan [5]. Algoritma ini dapat melakukan klasifikasi terhadap objek baru dengan melakukan perhitungan jarak terdekat pada objek tersebut terhadap data-data yang ada. Langkah perhitungan Algoritma *K-Nearest Neighbor* sebagai berikut:

- Menentukan parameter K yaitu jumlah tetangga paling dekat.
- Setelah itu dimulai untuk menghitung kuadrat jarak *Euclid* (*query instance*) terhadap data sampel yang diberikan.
- Kemudian mengurutkan objek yang telah dilakukan perhitungan *Euclid* kedalam kelompok yang memiliki jarak perhitungan terkecil dan bobot yang mendukung.
- Mengumpulkan kategori Y (Klasifikasi *Nearest Neighbor*).

Perhitungan jarak antar objek data baru dengan data lama ini dapat diukur oleh beberapa cara pengukuran salah satunya yaitu dengan memakai rumus *euclidean distance*, dan rumus dari perhitungan *euclidean distance* dapat dilihat sebagai berikut:

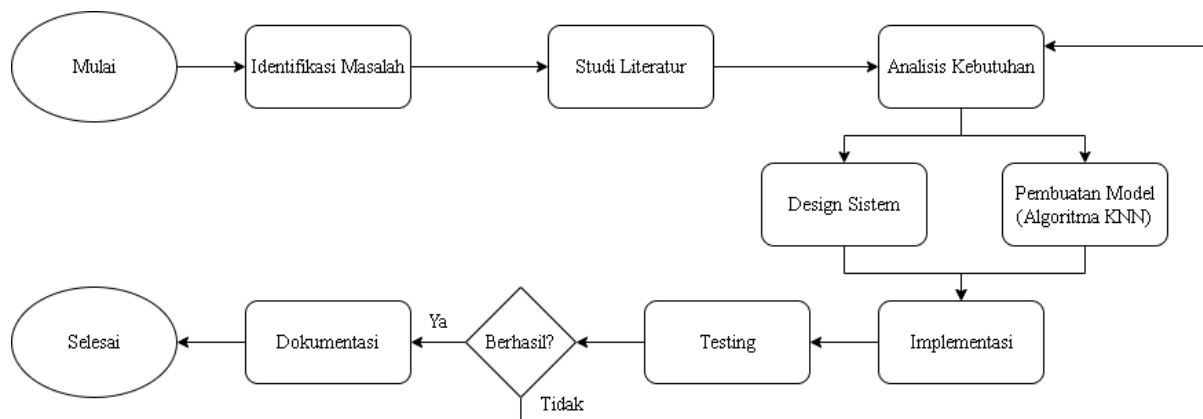
$$\text{Euclidean Distance} = \sqrt{(a_1 - b_1)^2 + \dots + (a_n - b_n)^2} \quad (1)$$

Keterangan:

a = a₁, a₂, a₃, ..., a_n hingga nilai ke n

b = b₁, b₂, b₃, ..., b_n hingga nilai ke n

3 Metode Penelitian



Gambar. 1. Alur Penelitian yang terdiri dari identifikasi masalah, studi literatur, analisis kebutuhan, desain sistem dan pembuatan model, implementasi dan pengujian sistem dan terakhir dokumentasi.

3.1 Identifikasi Masalah, Studi Literatur, Analisis Kebutuhan, dan Desain Sistem

Dalam identifikasi masalah yang mana dilakukan proses pemahaman dari masalah yang akan diteliti yang menjadi alasan dilakukannya penelitian ini, dengan melakukan studi literatur dan studi pustaka serta pengkajian penelitian terkait. Studi literatur yang dilakukan yaitu mencari sebuah referensi teori yang relevan dengan penelitian yang dilakukan yaitu mengenai prediksi kualitas air dengan teknik *data mining* dan memvisualisasikan informasi baru yang didapat dengan merancang sebuah sistem, selanjutnya dalam melakukan analisis kebutuhan yang perlu ada pada sistem dengan menggunakan metode analisis PIECES dan dalam melakukan desain sistem

demikian memperjelas akan seperti apa sistem yang akan dibuat, dengan membuat penggambaran interaksi yang mungkin akan terjadi pada sistem prediksi yang mana akan dibuat dengan menggunakan diagram DFD dan untuk memperjelas desain antarmuka yang akan dibuat akan dibuat sebuah contoh dari sketsa desain berbentuk *low-fidelity* yang dapat menggambarkan desain antarmuka secara simpel.

3.2 Pembuatan Model dan Penerapan Algoritma KNN

Pembuatan model diawali dengan pengumpulan dan pemahaman data yang akan dijadikan sebagai model nantinya, setelah itu data tersebut akan dilakukan proses *preprocessing* untuk menormalkan data-data yang tidak konsisten atau *error*, pembuatan model menggunakan software *data mining* yang cukup populer seperti *jupyter notebook* untuk mengolah data ke dalam bentuk model dengan menerapkan model *KNN* setelah pembuatan model selesai data akan diterapkan untuk dapat dimanfaatkan sistem prediksi agar dapat melakukan prediksi kualitas air yang dapat dikonsumsi atau tidaknya.

3.3 Implementasi Sistem dan Pengujian Sistem

Implementasi pembangunan sistem yang dilakukan yaitu dengan menggunakan *microframework flask*, yang mana framework ini menggunakan bahasa pemrograman *python*, Untuk pembuatan *front-end* nya sendiri dengan memanfaatkan *framework bootstrap* yang mana *framework* ini sudah cukup banyak membantu dan banyak digunakan oleh *developer* dalam hal membangun rancangan antarmuka *website* dengan format Bahasa *HTML*, *CSS*, dan dapat juga menggunakan *JavaScript*, selanjutnya *testing* yang akan dilakukan pertama dengan menggunakan pengujian *Black Box Testing* yang mana test ini merupakan pengujian terhadap kegunaan atau fungsionalitas dari sebuah sistem, yang dimaksudkan untuk menguji fitur spesifik yang terdapat pada sebuah sistem.

4 Hasil dan Pembahasan

4.1 Analisis Sistem Yang Berjalan

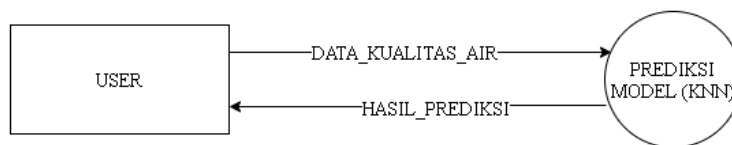
Dalam menganalisis sistem yang sudah berjalan digunakannya metode analisis PIECES untuk mengetahui kebutuhan dalam sistem yang akan dibangun nanti, analisis PIECES ini sendiri merupakan metode yang dapat digunakan mengidentifikasi dan memecahkan permasalahan yang ada pada sistem yang ada atau berjalan, dengan memberikan solusi dari hasil identifikasi masalah utama terhadap sebuah sistem.

Tabel 1. Tabel Analisis Sistem Berjalan (PIECES)

Kriteria	Sistem Yang Berjalan (Manual)	Sistem Usulan
<i>Performance</i>	Proses manual dapat memakan waktu yang cukup lama untuk menghasilkan prediksi yang tepat dan penulisan nilai dan kalimat kandungan juga mungkin bisa terjadi kesalahan penulisan hal ini dapat memperlambat dan membahayakan apabila hasil yang di uji tidak sesuai.	Sistem prediksi secara online melalui website untuk peneliti khusus kualitas air ini dapat melakukan input data dengan cepat, akurat, dan mudah serta menghasilkan hasil prediksi yang cukup akurat dan efisien waktu.
<i>Information</i>	Informasi mengenai bahayanya air dengan yang dapat dikonsumsi atau aman dan tidaknya atau tidak aman tidak mudah ditemukan terkecuali membacanya pada catatan peneliti sebelumnya atau dari buku.	Sistem yang diajukan pada website ini akan memberikan sebuah keterangan informasi mengenai bahayanya kualitas air yang aman atau tidaknya.

<i>Economy</i>	Sebuah alat perhitungan yang cukup mahal untuk melakukan perhitungan yang akurat.	Sistem yang diajukan dengan menggunakan website sistem prediksi ini, perhitungan prediksi dapat dilakukan secara gratis dan hemat biaya.
<i>Control</i>	Dengan semua yang dilakukan secara manual keamanan akan data yang dimiliki atau ditelah didapatkan bisa diketahui oleh peneliti lain dengan hanya melihat catatan manual yang dimiliki atau pada saat proses test dan perhitungan atau keteledoran kehilangan catatan nilai data.	Sistem yang diajukan menyediakan fitur lihat data yang mana nilai yang telah diinput peneliti dapat dicek kembali.
<i>Efficiency</i>	Untuk melakukan sebuah prediksi kualitas air itu sendiri perhitungan prediksinya diperlukan waktu yang lama untuk dapat mendapatkan hasilnya dan mungkin demi meyakinkan prediksi benar dengan hasil yang berubah-ubah mungkin saya terjadi.	Sistem yang diajukan menyajikan penginputan yang cukup cepat dan efisien waktu dalam hal perhitungan prediksi dan menampilkan hasilnya
<i>Services</i>	Pelayanan terhadap penyajian informasi mengenai kualitas air yang baik atau tidaknya, mungkin cukup sulit untuk didapatkan dalam laboratorium.	Sistem yang diajukan menyediakan keterangan informasi itu secara mudah dan gratis untuk dilihat dan dipahami.

4.2 Rancangan Sistem Yang Diusulkan



Gambar. 2. Rancangan sistem prediksi kualitas air dengan aliran data dari *user* yang menginput data kualitas air baru dan model KNN akan memproses dan memprediksi hasil kualitas air yang di *input*.

Berdasarkan gambar diatas, alat bantu perancangan sistem sederhana yang dibangun yaitu menggunakan *Data Flow Diagram*, DFD sendiri merupakan gambaran aliran data yang mengalir pada sebuah sistem informasi yang sedang berjalan [7]. Pada rancangan DFD ini memiliki *external* entitas yaitu bernamakan *user* yang mana pemakai sistem ini bisa digunakan oleh seorang admin, peneliti air lainnya, dan pemakai/masyarakat secara umum dan terdapat juga proses prediksi dengan menggunakan model KNN yang telah dibangun melalui proses data mining dan implementasikan. Proses aliran data sendiri terdapat dua arah aliran yaitu dari *user* ke proses bermakna *user* dapat menginput data kualitas air baru yang didapatkannya, aliran data dari proses ke *user* bermakna hasil data yang dihasilkan dari proses *input* sebelumnya berupa prediksi kualitas air tersebut baik dikonsumsi atau tidaknya.

4.3 Perhitungan Manual *K-Nearest Neighbor*

Untuk mengetahui cara kerja algoritma *K-Nearest Neighbor* maka dilakukan perhitungan manual terhadap data baru, dalam membuktikan apakah model algoritma *K-Nearest Neighbor* dapat memprediksi data baru ini. Perhitungan manual yang akan dilakukan akan menggunakan data berjumlah 10 data latih yang sudah dilakukan

pembagian data sebelumnya, pengambilan 10 data latih ditujukan untuk menghitung jarak dari data/objek baru yang akan diprediksi, untuk lebih detail 10 data latih yang akan diambil sebagai berikut:

Tabel Horizontal	1	2	3	4	5	6	7	8	9	10
Nama	A									
aluminium	0,6100	3,4700	1,8200	0,0700	0,1700	0,0500	2,1300	1,0200	2,9100	0,7800
ammonia	2,4100	15,8400	6,8100	21,8400	13,5500	29,6800	25,1100	23,9000	19,5900	22,2000
arsenic	0,0300	0,0200	0,0100	0,0300	0,0300	0,0300	0,7600	0,7200	0,0400	0,0010
barium	0,5900	0,0600	0,8500	3,4500	3,2800	2,8600	2,6000	2,9000	1,4800	2,2500
cadmium	0,0020	0,0010	0,0060	0,0080	0,0010	0,0060	0,1200	0,1100	0,0080	0,0030
chloramine	1,9400	5,2900	2,5500	6,1000	2,3800	4,5400	1,3000	2,3200	6,6200	4,7400
chromium	0,7700	0,4700	0,2500	0,2300	0,1700	0,7200	0,5200	0,1000	0,0600	0,1500
copper	1,5400	1,0800	1,0900	0,7600	0,2500	0,1400	0,4000	1,7600	1,6900	1,3300
flouride	0,6200	1,4300	1,3500	1,1600	1,2500	1,4200	0,5900	0,7400	0,9400	0,6300
bacteria	0,2300	0,8900	0,1600	0,7900	0,7000	1,0000	0,0000	0,0000	0,0000	0,0000
viruses	0,0010	0,8900	0,0020	0,7900	0,7000	1,0000	0,6100	0,0020	0,0000	0,0000
lead	0,0170	0,0800	0,0310	0,0490	0,1290	0,0820	0,0400	0,0480	0,1980	0,1320
nitrites	1,9900	1,9100	16,9900	16,8800	15,7200	11,3000	18,6600	11,6100	10,6700	11,0400
nitrites	1,0800	1,2000	1,7000	1,4000	1,8100	1,4600	1,8800	2,0400	1,6200	1,8500
mercury	0,0070	0,0080	0,0070	0,0060	0,0030	0,0060	0,0100	0,0040	0,0080	0,0090
perchlorate	11,1600	0,1800	44,7600	59,0300	10,8800	33,0900	51,7800	15,8700	50,5500	30,9100
radium	0,9800	6,8900	1,1500	4,2900	5,4100	4,9600	0,0500	0,1000	7,9200	6,6800
selenium	0,0100	0,0600	0,0800	0,0700	0,0300	0,0000	0,0300	0,0700	0,0800	0,1000
silver	0,4700	0,1200	0,2600	0,0800	0,1100	0,1600	0,0500	0,0900	0,0900	0,1100
uranium	0,0300	0,0800	0,0800	0,0000	0,0700	0,0500	0,0200	0,0100	0,0300	0,0500
kelas										
Is_safe	1	1	1	0	0	0	1	1	0	0

Gambar. 3. 10 Data latih pilihan dari data kualitas air yang didapatkan.

Lalu akan dilakukan prediksi ke data baru yang belum diketahui kelasnya, penentuan kelas berdasarkan jarak terdekat terhadap data latih diatas dan bobot dari kelas yang sama dengan untuk detail datanya sebagai berikut:

Tabel Horizontal	1
Nama	B
aluminium	2,17
ammonia	34,4
arsenic	0,05
barium	0,1
cadmium	0,002
chloramine	4,01
chromium	0,71
copper	1,4
flouride	1,6
bacteria	0
viruses	1
lead	0,25
nitrites	11
nitrites	1,55
mercury	0,007
perchlorate	57,5
radium	5,5
selenium	0,6
silver	0,1
uranium	0,4
kelas	
Is_safe	?

Gambar. 4. Data Uji atau data baru yang dijadikan percobaan prediksi dengan nilai kandungan mendekati kualitas air yang tidak dapat dikonsumsi (*notsafe*).

Untuk perhitungan ini menggunakan langkah-langkah dari algoritma *K-Nearest Neighbor* dengan menggunakan metode perhitungan jarak *euclidean distance*, yaitu sebagai berikut:

- Untuk Langkah pertama yaitu menentukan jumlah tetangga atau parameter nilai k , dalam perhitungan manual ini penulis menggunakan nilai $k = 3$ berdasarkan nilai performa yang terbaik yang telah dilakukan evaluasi model dari algoritma *K-Nearest Neighbor*.
- Langkah kedua ini, yaitu melakukan perhitungan jarak dengan metode *euclidean distance*, perhitungan ini dilakukan dengan memasukkan data latih (a) dan data baru (b) kedalam rumus (1) pada bab 2, untuk contoh perhitungan bisa dilihat dibawah sebagai berikut:

$$d1 = \sqrt{a_1 - b_1 + \dots + (a_n - b_n)^2}$$

$$d1 = \sqrt{(0,61 - 2,17)^2 + (2,41 - 34,4)^2 + (0,03 - 0,05)^2 + (0,59 - 0,1)^2 + (0,002 - 0,002)^2 + (1,94 - 4,01)^2 + (0,77 - 0,71)^2 + (1,54 - 1,4)^2 + (0,62 - 1,6)^2 + (0,23 - 0)^2 + (0,001 - 1)^2 + (0,017 - 0,25)^2 + (1,99 - 11)^2 + (1,08 - 1,55)^2 + (0,007 - 0,007)^2 + (11,16 - 57,5)^2 + (0,98 - 5,5)^2 + (0,01 - 0,6)^2 + (0,47 - 0,1)^2 + (0,03 - 0,4)^2}$$

$$d1 = \sqrt{2,4336 + 1023,3601 + 0,0004 + 0,2401 + 0 + 4,2849 + 0,0036 + 0,0196 + 0,9604 + 0,0529 + 0,998001 + 0,054289 + 81,1801 + 0,2209 + 0 + 2147,3956 + 20,4304 + 0,3481 + 0,1396 + 0,1396}$$

$$d1 = \sqrt{3282,26219}$$

$$d1 = 57,29098$$

.... Hingga Perhitungan ke d10

- Setelah didapatkan jarak pada setiap data kolom data latih terhadap data baru, maka akan dilakukan pengurutan dari jarak dengan nilai terkecil ke nilai terbesar.
- Terakhir menentukan kelas dari tetangga terdekat berdasarkan nilai k yang telah ditentukan sebelumnya, dengan nilai k yang cukup besar yaitu 3 maka perlu memperhatikan juga *weighted vote* atau pemilihan bobot terbanyak untuk mendukung suatu kelas yang sama. Untuk detail penentuan dapat dilihat dibawah, sebagai berikut:

Jarak	k = 3	is_safe
14,76956	✓	0
14,97823	✓	1
16,88862	✓	0
25,18023	-	0
29,44636	-	0
31,35279	-	1
43,45422	-	1
51,47850	-	0
57,29098	-	1
60,98759	-	1

Gambar. 5. Hasil perhitungan manual algoritma KNN dalam memprediksi data baru terhadap data kualitas air yang dipakai dengan $k=3$.

Maka dapat disimpulkan berdasarkan hasil diatas, data baru atau objek baru yang diuji memiliki kelas $is_safe = 0(notsafe)$ dikarenakan dari ke 3 tetangga terdekat yaitu memiliki kelas $is_safe = 0(notsafe)$ dan bobot yang lebih banyak untuk kelas $is_safe = 0(notsafe)$.

4.4 Penerapan *K-Nearest Neighbor* Menggunakan *Jupyter Notebook*

Berikut adalah proses penerapan algoritma *K-Nearest Neighbor* dalam *jupyter notebook*, *jupyter notebook* ini sendiri merupakan alat bantu sendiri yaitu untuk membantu dalam membuat sebuah narasi komputasi yang akan menjelaskan makna dari setiap data yang diolah dan memberikan pengetahuan lebih mengenai data, dengan alat bantu ini dengan menggunakan perintah dari bahasa *python*:

```
from sklearn.neighbors import KNeighborsClassifier #Melakukan import algoritma KNN
from sklearn.metrics import accuracy_score #Import Perhitungan Akurasi dari sklearn
from sklearn.metrics import confusion_matrix #Import Confusion Matrix dari sklearn
from sklearn.metrics import classification_report #Import Laporan Klasifikasi dari sklearn

KNN = KNeighborsClassifier(n_neighbors=3)
model = KNN.fit(X_res, y_res)
```

Gambar. 6. Penerapan algoritma KNN dengan menggunakan *tools* data mining *jupyter notebook*.

Berdasarkan gambar 3, nilai yang digunakan dengan $k = 3$. Nilai parameter pada kode program *n_neighbors* ini dapat diubah sesuai dengan jumlah parameter yang akan digunakan atau diuji coba. Performa yang didapatkan dengan menggunakan nilai $k = 3$ mendapatkan performa akurasi sebesar 85,52% dan nilai ini sudah merupakan nilai paling optimal dibandingkan dengan nilai k lainnya setelah dilakukannya proses *data preprocessing*. Nilai performa model yang telah diterapkannya model KNN dapat dilihat pada Gambar 4 dibawah:

```
print(classification_report(y_test, hasil_prediksi))
```

	precision	recall	f1-score	support
0	0.85	0.88	0.86	97
1	0.86	0.83	0.84	86
accuracy			0.85	183
macro avg	0.85	0.85	0.85	183
weighted avg	0.85	0.85	0.85	183

Gambar. 7. Hasil Akurasi Performa dari nilai $k=3$

4.5 Pembuatan Model

Setelah diterapkannya algoritma KNN maka akan dibuat sebuah model algoritma KNN berdasarkan pengujian dan penerapan algoritma berdasarkan hasil performa pada Gambar 4, yaitu proses pembuatan model dapat dilihat pada gambar dibawah sebagai berikut:

```
import joblib
joblib.dump(KNN, "knn-kualitas-air-default.pkl")

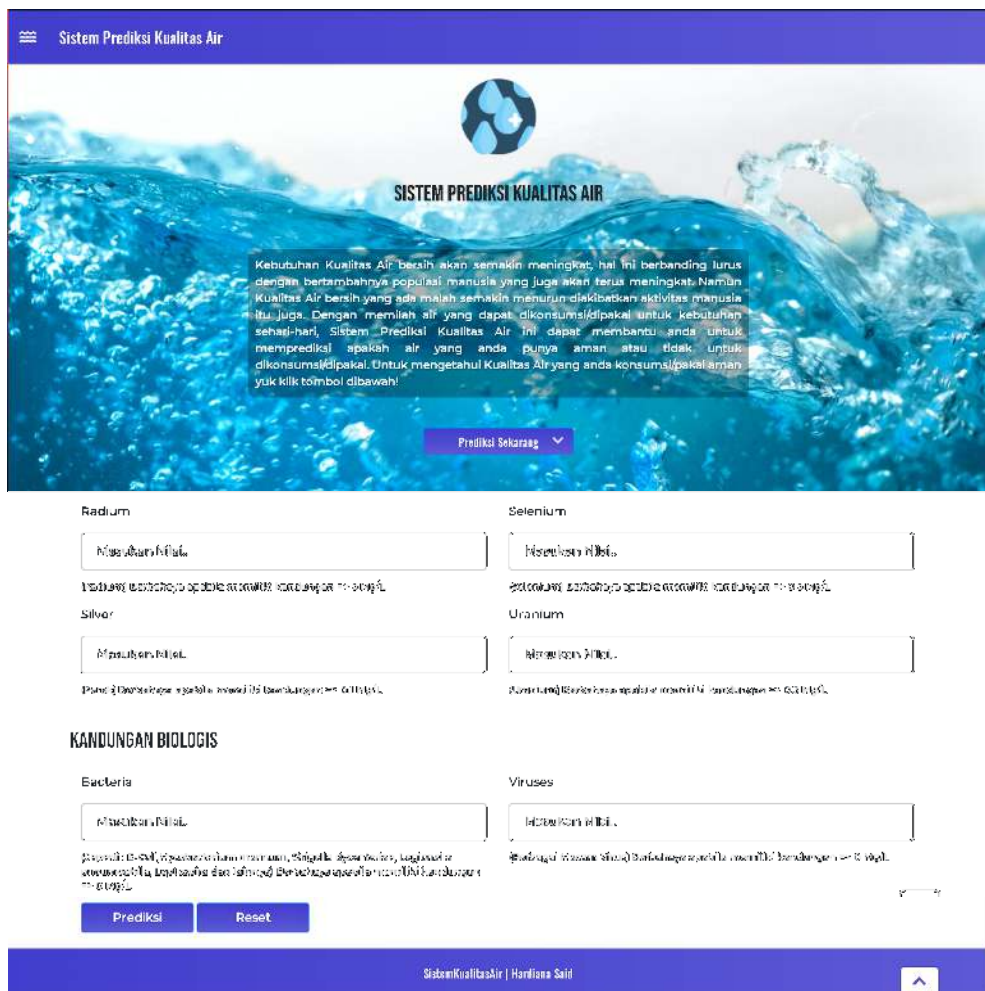
['knn-kualitas-air-default.pkl']
```


Gambar. 8. Pembuatan model algoritma KNN dengan menggunakan *library joblib* untuk dapat diterapkan kedalam sistem prediksi yang telah dibuat.

4.6 Implementasi Sistem

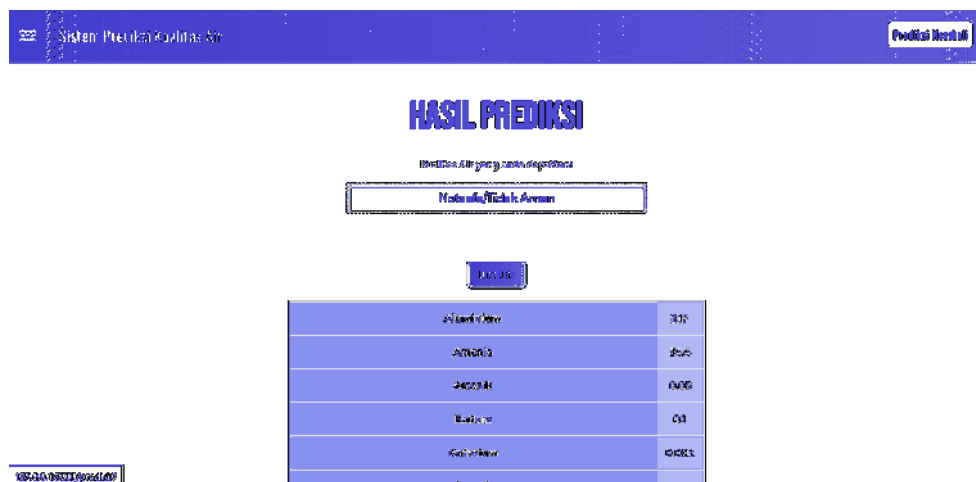
Desain antarmuka *website* ini dibangun dengan menggunakan perangkat lunak bantuan seperti, untuk *front-end design* menggunakan *bootstrap* (*HTML, CSS, dan beberapa tombol menggunakan JavaScript*) untuk membuat sistem prediksi sederhana ini lebih efektif dan efisien waktu.

4.6.1 Halaman Beranda



Gambar. 9. Page awal dari halaman beranda ini berisikan informasi singkat mengenai kegunaan dari sistem prediksi ini melalui narasi yang diberikan, lalu user dapat menekan tombol prediksi sekarang setelah membaca informasi di atasnya untuk mempercepat redirect page ke page input data untuk memprediksi, dan pada page kedua dari halaman beranda ini berisikan user dapat melakukan proses prediksi data kualitas air yang didapatkannya berdasarkan kandungan kimia dan kandungan biologis yang diperlukan.

4.6.2 Halaman Prediksi



Gambar. 10. Pada halaman ini sistem akan menampilkan hasil prediksi dari inputan data yang dilakukan *user* dari halaman beranda (page inputan prediksi), halaman ini juga menampilkan data inputan dalam bentuk tabel dengan menekan tombol terlebih dahulu, dan menampilkan informasi tambahan mengenai kualitas air.

4.7 Pengujian Sistem

Pada pengujian pertama ini yaitu menggunakan *Black Box Testing* yang mana pengujian dilakukan untuk mengetahui apakah sistem prediksi yang telah dibuat dapat bekerja dengan baik, pengujian yang akan dilakukan dari segi fungsionalitas pada sistem yang mana sistem harus memberikan respon yang benar dari setiap fungsi yang ada pada sistem seperti contoh fungsi tombol.

Tabel 2. Pengujian *BlackBox Testing*

Modul	Input	Proses	Ouput	Status
Halaman Beranda	Mengakses halaman URL	Mengarah ke sistem	Menampilkan halaman awal	Berhasil
	Klik tombol Prediksi Sekarang	Mengarah ke page ke 2 untuk melakukan input data	Menampilkan page prediksi.	Berhasil
	Klik tombol Prediksi	Mengarah ke halaman prediksi	Menampilkan halaman prediksi yang berisikan hasil prediksi	Berhasil

	Klik tombol Reset	Mengarah ke penghapusan data input yang telah di masukan	Menampilkan kolom data inputan prediksi yang terhapus	Berhasil
Halaman Prediksi	Klik tombol Prediksi Kembali	Mengarah ke halaman beranda (page-prediksi)	Menampilkan halaman beranda	Berhasil
	Klik tombol Lihat Data	Mengarah ke pembukaan tabel hasil inputan sebelumnya atau <i>pop-up</i> tabel.	Menampilkan tabel data inputan sebelumnya.	Berhasil
Kedua Halaman	Klik tombol Back-to-top	Mengarah ke halaman bagian atas website	Menampilkan halaman awal pada setiap halaman	Berhasil

5 Penutup

Berdasarkan penelitian yang telah dilakukan dalam membangun sebuah sistem prediksi kualitas air, dapat ditarik kesimpulan yang mana model algoritma *K-Nearest Neighbor* mampu dalam memprediksi kualitas air dengan sangat baik berdasarkan nilai akurasi yang didapatkan, yaitu sebesar 85,24% dan sistem yang telah dibuat dapat memprediksi data baru dengan baik, hal ini dibuktikan dengan samanya hasil prediksi dari kedua percobaan mulai dari perhitungan manual dan percobaan pengujian sistem yang menghasilkan prediksi yang sama. Bagi penelitian selanjutnya untuk sistem prediksi dapat dikembangkan lagi untuk dapat memprediksi data secara dinamis dan mengembangkan fitur baru yang mungkin dapat membantu peneliti lain atau masyarakat yang memakai sistem prediksi kualitas air ini dalam mendapatkan informasi prediksi dan rekomendasi yang lebih luas dan cobalah menggunakan *framework* perancangan website lainnya yang mungkin memiliki fitur dan proses pengimplementasian sistem yang lebih baik.

6 Daftar Pustaka

- [1] Saraswati, Rieke. (2020). Ini 6 Penyakit Akibat Pencemaran Air yang Perlu Diwaspadai. Diakses pada: <https://idcloudhost.com/apa-itu-data-mining-definisi-fungsi-metode-dan-penerapannya/>
- [2] Hasmawati, dkk. (2017). Aplikasi Prediksi Penjualan Barang Menggunakan Metode *K-Nearest Neighbor* (*KNN*) (Studi Kasus Tumaka Mart). *semanTIK*, Vol. 3, No.2.
- [3] Armono, Rio Adi, dkk. (2018). Implementasi Algoritma *K-Nearest Neighbor* Untuk Identifikasi Kualitas Air (Studi Kasus: Pdam Kota Surakarta). *Jurnal TIKomSIN*, Vol. 6, No. 2.
- [4] Marisa, Fitri. (2013). *Educational Data Mining* (Konsep Dan Penerapan). *Jurnal Teknologi Informasi*, Vol. 4, No.2.
- [5] Dsn, Cahya. (2018). *Algoritma K-Nearest Neighbor (k-NN)*. Diakses pada: <https://cahyadsn.phpindonesia.id/extra/knn.php>
- [6] Anwardi, dkk. (2020). Analisis PIECES Dan Pengaruh Perancangan Website Fikri Karya Gemilang Terhadap Sistem Promosi Menggunakan Model Waterfall. *Jurnal ReKayasa Sistem dan Industri*, Vol.7, No.1.
- [7] Adani, Muhammad Robith. (2021). Ketahui Apa Itu Data Flow Diagram (DFD) Beserta Jenis Dan Fungsinya. Diakses pada: <https://www.sekawanmedia.co.id/blog/pengertian-dfd/>