

PENERAPAN DATA MINING UNTUK KLASIFIKASI KUALITAS AIR DENGAN ALGORITMA SUPPORT VECTOR MACHINE PADA DINAS LINGKUNGAN HIDUP DAN PERTANAHAN PROVINSI SUMSEL

Nanda Fadillah Fattah

Program Studi Sistem Informasi, Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data Palembang
Jl. Rajawali No.14, 9 Ilir, Kec. Ilir Tim. II, Kota Palembang, Sumatera Selatan 30113
E-mail: nan.id@mhs.mdp.ac.id

Abstrak - Kualitas air merupakan parameter kritis yang mempengaruhi keberlanjutan lingkungan hidup dan pertanian. Dalam rangka meningkatkan pemahaman dan pengelolaan kualitas air, penelitian ini mengusulkan penerapan teknik Data Mining dengan menggunakan algoritma Support Vector Machine (SVM). Penelitian ini dilakukan dengan menggunakan data yang diperoleh dari Dinas Lingkungan Hidup dan Pertanian Provinsi Sumatera Selatan, yang mencakup berbagai parameter kualitas air. Algoritma SVM digunakan untuk membangun model klasifikasi yang dapat memprediksi kualitas air berdasarkan parameter-parameter yang diukur. Evaluasi model dilakukan dengan menggunakan data uji, dan hasilnya menunjukkan kemampuan model SVM dalam mengklasifikasikan kualitas air dengan akurasi yang signifikan. Implementasi Data Mining pada data kualitas air di lingkungan ini dapat memberikan kontribusi penting dalam pengambilan keputusan untuk pemantauan dan perbaikan kualitas air di Provinsi Sumatera Selatan.

Kata Kunci: Data Mining, SVM Algoritma, Kualitas Air, Sumsel.

I. PENDAHULUAN

Untuk mempercepat pemrosesan data yang ada di pemerintahan, maka harus dilakukan peningkatan pemahaman di bidang teknologi dan bidang teknologi informasi. Pemerintahan akan mendapatkan dampak positif yang diberikan jika menerapkan teknologi informasi agar tidak mengalami hambatan dan keterlambatan kerja. Sehingga pemerintah bisa mempersingkat waktu dengan lebih efisien dan cepat.

Pendataan dapat membantu pemerintah dalam memahami dan merespons kebutuhan masyarakat dengan lebih baik. Melalui analisis data yang terkumpul, pemerintah dapat mengidentifikasi pola-pola perilaku, preferensi, dan masalah yang dihadapi oleh masyarakat, sehingga dapat merancang kebijakan yang lebih sesuai (Pentland, 2014).

Dinas Lingkungan Hidup dan Pertanian (DLHP) Provinsi Sumatera Selatan adalah dinas pemerintahan yang bergerak di bidang lingkungan hidup daerah yang meliputi kegiatan dalam melakukan pengawasan, pengendalian, dan penertiban terhadap segala sesuatu mengenai lingkungan hidup di Provinsi Sumatera Selatan. DLHP memiliki amanah untuk menjaga kualitas lingkungan hidup demi kehidupan di masa depan. Oleh sebab itu, diperlukan perlindungan dan pengelolaan lingkungan hidup yang sungguh-sungguh dan konsisten oleh semua pihak. Kantor kedinasan ini buka di setiap hari Senin-Jumat pada pukul 08:00 WIB sampai 16:30. Kantor kedinasan ini dikenal sebagai kantor kedinasan lainnya hanya saja berbeda bidangnya saja kantor DLHP ini memiliki beberapa berbagai bidang yang di

wewenangannya, seperti bagian lingkungan, pemetaan lingkungan perizinan dan pemantauan lainnya (dlhp sumselprov, 2017).

Permasalahan yang ada dari penjelasan uraian pada latar belakang yang telah dijelaskan, maka dapat disimpulkan bahwa adanya kesulitan dalam melakukan penginputan data dan melakukan penerapan penjelasan data secara baik dengan begitu, hasil dari output data yang sudah dikelola tersebut akan dijelaskan kepada bidang yang lainnya, mereka harus membutuhkan waktu yang termasuk lama karena mereka hanya menggunakan office saja. Bidang pengendalian pencemaran hanya menggambarkan data dalam bentuk batangan atau grafik masih dengan manual. Sehingga mereka masih membutuhkan waktu yang cukup lama. Dan pihak kedinasan juga masih kesulitan melakukan kesimpulan data yang ada pada data mereka, kesimpulan dibutuhkan agar dinas bisa memperjelas makna dari sampel data mereka sendiri.

Tujuan dari penelitian ini adalah untuk mengorganisir informasi sehingga dapat dengan mudah diidentifikasi, dikelola, dan diambil. yang lebih signifikan dan lebih akurat, dengan begitu pihak DLHP akan lebih tepat dalam mengambil keputusan, perencanaan jangka panjang yang lebih akurat, pengelolaan resiko yang lebih teratur, dan peningkatan efisiensi operasional yang sedang berjalan, selain itu dengan adanya pengelompokan klasifikasi yang lebih baik diharapkan DLHP bisa memberikan inovasi yang lebih baik kepada dinas bidang lainnya.

Manfaat yang diberikan dari penelitian ini adalah memberikan efisiensi waktu yang lebih baik lagi untuk pengembangan data dan pengolahan

data pada Dinas Lingkungan Hidup dan Pertanian. Dengan adanya efisiensi waktu diharapkan aktivitas dinas yang dilakukan lebih baik dan efisien. Peningkatan sistem informasi dari dlhp diharapkan bisa memotivasi bidang lainnya agar tidak ketinggalan penggunaan teknologi yang ada.

II. TINJAUAN PUSTAKA

Data Mining merupakan suatu proses penggalian data atau penyaringan data dengan memanfaatkan kumpulan data yang memiliki ukuran cukup besar melalui serangkaian proses untuk mendapatkan informasi yang berharga dari data tersebut. Data Mining bisa digunakan oleh perusahaan besar untuk menggali data dan mendapatkan informasi yang dapat menunjang serta meningkatkan proses bisnis perusahaan tersebut. Data mining melibatkan penerapan teknik-teknik statistik, matematika, dan kecerdasan buatan untuk menganalisis data besar dan mengidentifikasi pola-pola yang bermanfaat. Prosesnya melibatkan langkah-langkah seperti pemilihan data, *preprocessing* data, pemodelan, evaluasi, dan interpretasi hasil (Suyanto, 2017).

Data mining dapat digunakan dalam berbagai bidang dan industri untuk menghasilkan wawasan yang berguna. Beberapa contoh penggunaannya termasuk dalam pemasaran, keuangan, manufaktur, kesehatan, penelitian ilmiah, dan lain sebagainya. Dengan analisis data yang mendalam, organisasi dapat mengambil keputusan yang lebih baik, mengoptimalkan operasi mereka, mengidentifikasi peluang bisnis, mendeteksi penipuan, dan memprediksi perilaku atau kejadian di masa depan (Naldy & Andri, 2021).

Support Vector Machine adalah suatu teknik yang relatif baru (1995) untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi, yang sangat populer belakangan ini. Support vector machine berada dalam satu kelas dengan Neural Network dalam hal fungsi dan konsider permasalahan yang bisa diselesaikan, keduanya masuk ke dalam kelas *supervised learning*. Baik para ilmuwan maupun para praktisi telah banyak melakukan penelitian dengan menerapkan teknik ini untuk menyelesaikan masalah-masalah nyata dalam kehidupan. Baik dalam masalah *gene expression analysis*, finansial, cuaca hingga bidang kedokteran terbukti dalam banyak implemetasi support vector machine memberi hasil yang lebih baik dari neural network terutama dalam hal solusi yang dicapai (Nurhayati, dan Luthfi, 2015).

Dalam algoritma SVM, data direpresentasikan dalam ruang n-dimensi dimana SVM dapat mengklasifikasi apakah *instance* pelatihan baru termasuk dalam kategori yang sama atau kategori yang berbeda. Kumpulan data yang diubah menjadi format numerik dengan label penyandian menyederhanakan proses klasifikasi dengan SVM

ini. Hal ini akan lebih jelas terlihat ketika membandingkan kategori yang sama atau berbeda yaitu antara Kelas dengan nilai 0 atau 1. Tujuan dari SVM adalah untuk menemukan hyperplane dalam ruang n-dimensi yang dapat mengklasifikasikan titik data. SVM menggunakan kernel garis lurus yang membagi dua kelas dengan persamaan linier yang dapat dilihat pada persamaan (A. Toha, P. dan W. Gata, 2020).

$$w * x - b = 0 \quad (1)$$

Berdasarkan persamaan (1), simbol w adalah parameter hyperplane yang dicari, x untuk data input, dan b adalah bias. Teknik untuk menghasilkan hyperplane yang optimum pada SVM dapat dilakukan dengan menggunakan persamaan 2 dan 3 (A. Toha, P. dan W. Gata, 2020).

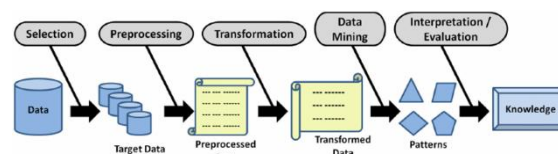
$$\min \frac{1}{2} \|\omega\|^2 \quad (2)$$

$$y_i(wx_i + b) \geq 1, i = 1, \dots, \lambda$$

Persamaan (3) digunakan untuk memaksimalkan nilai $\|\omega\|^2$ dengan fokus pada pembatas $y_i(wx_i+b)$. Jika data keluaran adalah $y_i=+1$, maka pembatas menjadi $(wx_i+b) \geq 1$ dan sebaliknya jika $y_i=-1$, maka pembatas menjadi $(wx_i+b) \leq -1$ (A. Toha, P. dan W. Gata, 2020).

Knowledge Discovery In Database (KDD) merupakan metode untuk memperoleh pengetahuan dari *database* yang ada. Dalam *database* terdapat tabel-tabel yang saling berhubungan / berelasi. Hasil pengetahuan yang diperoleh dalam proses tersebut dapat digunakan sebagai basis pengetahuan (*knowledge base*) untuk keperluan pengambilan keputusan.

Istilah KDD dan data mining seringkali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain, dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining. Proses KDD secara garis besar dapat dijelaskan sebagai berikut:



Gambar 1. Task Knowledge Discovery In Database

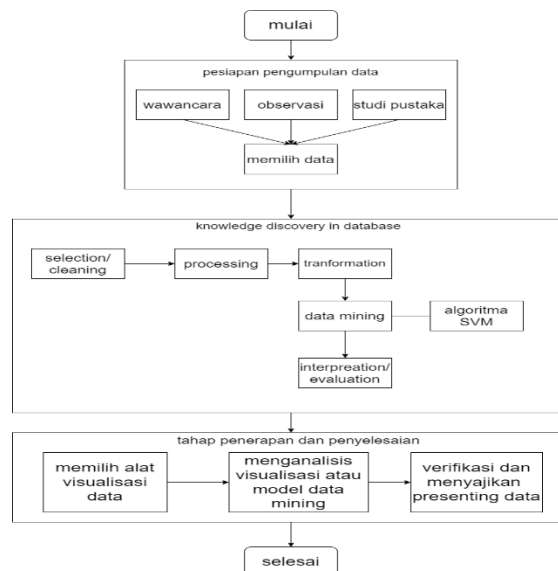
Berikut beberapa penelitian terdahulu yang dilakukan oleh beberapa penulis sebagai berikut:

Tabel 1. Penelitian Terkait

No	Judul Penelitian	Penulis	Kesimpulan
1	Model Prediksi Kualitas Udara dengan <i>Support Vector Machines</i> dengan Optimasi Hyperparameter GridSearch CV	(Ahmad Toha, Purwono, Windu Gata, 2022)	Kontribusi dalam penelitian ini adalah melakukan kembali pembuatan model klasifikasi dengan SVM dengan teknik baru yaitu perbaikan dalam pemrosesan data hingga melakukan tuning hyperparameter. dinas melihat para peneliti sebelumnya hanya mengejar nilai akurasi yang tinggi.
2	Implementasi <i>Data Mining</i> Menggunakan Model SVM Untuk Prediksi Kepuasan Pengunjung Taman Tabebuaya	(Agus Darmawan, Nunu Kustian, Wanti Rahayu, 2018)	Pada penelitian yang dilakukan Penelitian ini adalah penelitian eksperimen yang bertujuan untuk mengukur akurasi algoritma SVM. Data yang digunakan dalam penelitian dari data kuesioner pengunjung taman Tabebuaya.
3	Analisis Prediksi Keterlambatan Pembayaran Listrik Menggunakan Komparasi Metode Klasifikasi Decision Tree dan <i>Support Vector Machine</i>	(Dinda Nabila Batubara, Agus Perdana Windarto, Eka Irawan, 2022)	memprediksi faktor atau melihat faktor utama yang menjadi penyebab keterlambatan pembayaran listrik. Hasil penelitian diharapkan dapat menjadi menjadi bahan.

III. METODE PENELITIAN

Pada proses melakukan penelitian ini terdapat 3 tahapan yaitu tahap perencanaan proyek, tahap persiapan data, dan tahap analisis data, dan penyelesaian proyek. Pada tahap persiapan untuk mengumpulkan data-data yang diinginkan, peneliti melakukan pendekatan dengan mencari permasalahan dengan melakukan pendekatan wawancara dan beberapa observasi.



Gambar 2. Langkah Pada Tahap Penelitian

Setelah mendapatkan masalah yang ada, peneliti akan memastikan dan akan meminta data yang sesuai dengan permasalahan tersebut. Setelah itu data akan di proses dan di analisis terkait permasalahan yang ada. Dengan menggunakan algoritma SVM, peneliti akan mencari penyesuaian data yang signifikan dan relevan. Agar permasalahan bisa diselesaikan. Kemudian masuk pada tahap pengulangan agar peneliti bisa memastikan hasil dari data tersebut apakah sudah sesuai atau tidak, jika tidak pembaruan akan dilakukan agar bisa memastikan lagi penerapan data harus sesuai dengan tujuan dan permasalahan. Jika sudah sesuai maka permasalahan pada DLHP bisa diselesaikan.

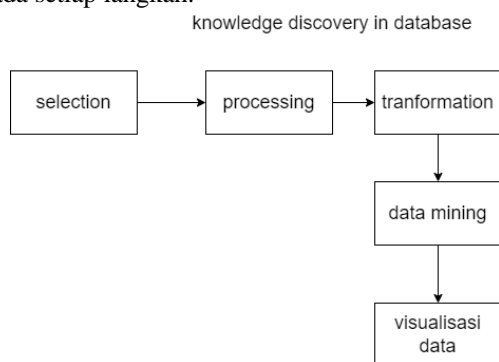
Studi Pustaka adalah proses pengumpulan data-data yang relevan. Pengumpulan data ini dilakukan untuk memperoleh sejumlah informasi atau fakta-fakta yang dibutuhkan pada penelitian dalam upaya untuk mencapai tujuan penelitian. Data tersebut akan mendukung penulis untuk melakukan penelitian dan mendukung penulis dalam menyusun skripsi ini. Studi pustaka yang digunakan penulis bersumber dari e-book yang berjumlah 2 e-book, jurnal yang berjumlah 12 jurnal, dan website yang berjumlah 2 website dan skripsi yang berjumlah 2 skripsi.

Metode wawancara digunakan untuk mengetahui permasalahan dan solusi yang ada pada Dinas Lingkungan Hidup dan Pertanahan. Wawancara dilakukan dengan bantuan para narasumber dari DLHP yang merupakan seorang Kepala Subbagian Umum dan Kepegawaian yaitu ibu Rian Syafni beliau diwawancarai pada tanggal senin 22 mei 2023, dan bagian jabatan fungsional yaitu bapak Ganda Qurniawan beliau diwawancarai pada tanggal jumat 26 mei 2023. Arti narasumber mengacu pada (Ana Nadhya Abrar, 2016) mengartikan narasumber adalah seseorang yang memiliki peran dalam memberikan informasi penting yang didasarkan mempunyai keahlian

tertentu sesuai dengan topik yang dibahas dalam sebuah kegiatan wawancara.

Penulis juga menggunakan metode observasi untuk penelitian ini. Observasi yang digunakan adalah observasi tidak terstruktur dimana observasi tidak dilakukan secara sistematis. Metode pengumpulan data observasi digunakan untuk pengamatan lapangan. Observasi dilakukan dengan cara datang langsung ke Dinas Lingkungan Hidup dan Pertanahan. Data dari melakukan observasi adalah untuk mengetahui keadaan Kedinasan.

KDD adalah proses ekstraksi pengetahuan atau pola berharga dari volume data yang besar. Ini melibatkan beberapa tahapan, dan berbagai metode digunakan pada setiap tahap. Berikut adalah gambaran umum dari proses tipikal Knowledge Discovery in Databases dan metode yang digunakan pada setiap langkah.



Gambar 3. Metode Knowledge Discovery in Database

Di awal tahapan pada *selection*, yang mana *data selection* adalah proses menganalisis data yang relevan dari *database* karena sering ditemukan bahwa tidak semua data dibutuhkan dalam proses data mining. Proses seleksi dilakukan pada data sampel air yang dilakukan oleh para peneliti DLHP. Penelitian dilakukan dengan 72 titik berbeda dan dilakukan setidaknya 4 kali sehingga dalam 1 tahun menghasilkan sekitaran 35.640 data. Yang datanya akan digunakan untuk proses data mining dengan membuat dataset yang memiliki beberapa atribut. Pada tahap ini dilakukan penyeleksian dan penambahan pada data, sebelum penyeleksian data dilakukan, data akan dilengkapi dengan tabel efektif yang mana tabel akan dipisahkan menggunakan microsoft excel 2019 karena data yang dikelola berformat xls/xlsx.

Tahap selanjutnya adalah tahap *preprocessing*. Pada tahap ini dilakukan proses pembersihan data, seperti menghilangkan *noise*, agar tidak ada duplikasi data, memeriksa data yang inkonsisten dan memperbaiki kesalahan pada data seperti kesalahan cetak, sehingga data tersebut dapat diolah dan dilakukan proses data mining. Tahap ini masih menggunakan Microsoft Excel 2019 dengan format *xlsx*.

Transformasi data akan diubah menjadi data yang sesuai untuk dilakukan pengolahan data pada tahap selanjutnya, yaitu tahap data mining. Pada tahap ini hanya terjadi pergantian nama atribut yang akan digunakan.

Visualisasi dan interpretasi dari *rule* yang sudah didapat sesuai dengan tujuan yang sudah ditentukan. Dengan melakukan visualisasi data, data-data yang sudah dilakukan pengelolaan dan lainnya dapat mempermudah hasil akhir dari data tersebut, visualisasi data dilakukan dengan menggunakan aplikasi *tableau 2021.4*.

IV. HASIL DAN PEMBAHASAN

Penentuan struktur data yang ada pada sistem yang memenuhi kaidah database relasi yang baik dan benar, dimana suatu relasi berisi data berupa tabel yang saling berhubungan, maka dari itu dilakukan pemrosesan awal terlebih dahulu berupa:

- 1) Perancangan data
- 2) Pembersihan data
- 3) Transformasi data
- 4) Menganalisis data

Pada saat memulai pembersihan data, data dikumpulkan terlebih dahulu menggunakan *microsoft excel*, data digunakan dalam format *xlsx*. *Excel* digunakan untuk menyimpan bagian-bagian data.

Dataset masih belum dapat digunakan dalam model klasifikasi, sehingga diperlukan upaya pengolahan data, seperti normalisasi dan pembersihan data yang tidak berarti untuk membentuk transformasi. Pemrosesan data dapat dimulai dengan menangani *missing data* yaitu mengatasi beberapa data yang bersifat *nullable* atau kosong dengan cara menghapusnya. *Missing data* menggambarkan nilai mana yang hilang dan dapat diamati dalam kumpulan data tersebut.

Lampiran 1. Hasil Perhitungan Status Mutu Air Dengan Metode Storet

PARAMETER	Unit	KATEGORI			SMA	MMA	JMS	Periode	
		SB	S	C				2019	2020
Lokasi Sungai Lengki Gunung Melayu (Subsistem Muara Estu)									
PUSUK									
PM	mg/L	413	1000	413	1000	413	1000	1000	1000
TS	mg/L	144	50	387	1000	1000	1000	1000	1000
KEMAS									
DO	mg/L	100	1000	1000	1000	1000	1000	1000	1000
PH	mg/L	54	1000	1000	1000	1000	1000	1000	1000
ESD	mg/L	140	1000	1000	1000	1000	1000	1000	1000
CSO	mg/L	200	1000	1000	1000	1000	1000	1000	1000
ML	mg/L	100	1000	1000	1000	1000	1000	1000	1000
ML	mg/L	10	1000	1000	1000	1000	1000	1000	1000
TP	mg/L	1000	1000	1000	1000	1000	1000	1000	1000
Ammonium(CN)	mg/L	100	1000	1000	1000	1000	1000	1000	1000
TP	mg/L	100	1000	1000	1000	1000	1000	1000	1000
Phenol	mg/L	1000	1000	1000	1000	1000	1000	1000	1000
Timah Hitam	mg/L	100	1000	1000	1000	1000	1000	1000	1000
BARIS	mg/L	1000	1000	1000	1000	1000	1000	1000	1000
Sulfida(S2)	mg/L	1000	1000	1000	1000	1000	1000	1000	1000
Sulfida(S2)	mg/L	1000	1000	1000	1000	1000	1000	1000	1000
PH	mg/L	1000	1000	1000	1000	1000	1000	1000	1000
Selenium	mg/L	1000	1000	1000	1000	1000	1000	1000	1000

Gambar 4. Data yang Diterima

Dalam konteks data, "parameter" merujuk kepada karakteristik atau sifat tertentu yang digunakan untuk menggambarkan data atau populasi. Parameter adalah ukuran atau statistik yang menggambarkan sesuatu tentang populasi yang ingin diinvestigasi atau dianalisis. Ini berbeda dengan "statistik," yang merujuk kepada ukuran atau metrik yang digunakan untuk menggambarkan data yang telah diambil dari populasi (sampel).

Dalam konteks data, "unit" merujuk kepada entitas atau objek yang sedang diamati, diukur, atau dicatat dalam set data. Unit adalah unit dasar yang membentuk data tersebut, dan bisa berupa individu, objek, observasi, atau entitas lainnya yang menjadi fokus dari pengumpulan data.

dalam konteks data merujuk kepada interval waktu di mana pengambilan sampel atau pengumpulan data dilakukan secara berkala. Ini mengacu pada jadwal atau frekuensi di mana data dikumpulkan atau diukur selama suatu periode waktu tertentu. Periode sampling ini dapat bervariasi tergantung pada jenis data dan tujuan pengumpulan data.

BML dapat memiliki beberapa arti berbeda tergantung pada konteksnya. Namun, salah satu makna umum dari singkatan "BML" adalah "Berat Molekul Rata-rata". Berat molekul rata-rata adalah nilai berat molekul rata-rata suatu molekul atau senyawa, yang dihitung dengan mempertimbangkan komposisi isotopnya. Ini adalah nilai yang sering digunakan dalam kimia dan ilmu-ilmu terkait untuk menghitung jumlah zat kimia dalam berbagai reaksi dan perhitungan.

Dalam konteks data, "arti maksimum" sering kali mengacu pada nilai maksimum atau angka terbesar dalam kumpulan data. Ini adalah titik paling tinggi dalam distribusi data. Istilah "arti min" dalam konteks data mengacu pada nilai minimum atau angka terkecil dalam kumpulan data. Ini adalah titik paling rendah dalam distribusi data. Nilai minimum digunakan untuk mengidentifikasi nilai terkecil dalam satu set data. Rata-rata (average) dalam sebuah kumpulan data adalah ukuran pusat yang menggambarkan "nilai tengah" dari data tersebut. Rata-rata ditemukan dengan menjumlahkan semua nilai dalam kumpulan data dan kemudian membaginya dengan jumlah total nilai.

Pada tahap cleaning data dilakukannya pengecekan data yang kosong, data yang bernilai kosong ada 102 yang memiliki data yang terduplicates. Sedangkan data yang kosong dinas akan cek secara langsung apakah benar ada kolom yang kosong atau tidak. Ada 4 kolom yang kosong yaitu 2 slot data di bagian panol, 2 data pada slot bagian sulfat dan 2 slot data pada bagian crom. Nilai yang kosong akan di konfirmasi lagi ke pihak dinas lalu data akan di update dengan data baru.

Dataset masih belum dapat digunakan dalam model klasifikasi, sehingga diperlukan upaya pengolahan data, seperti normalisasi dan pembersihan data yang tidak berarti untuk membentuk transformasi. Pemrosesan data dapat dimulai dengan menangani missing data yaitu mengatasi beberapa data yang bersifat nullable atau kosong dengan cara menghapusnya. Missing data menggambarkan nilai mana yang hilang dan dapat diamati dalam kumpulan data tersebut. Dataset tersebut memiliki kategori atau bisa disebut dengan class dalam klasifikasi. Pengelompokan setiap fitur yang

merujuk pada suatu kelas masih bersifat string atau teks dan dinas dapat mengubahnya ke dalam bentuk nominal. Sebagai contoh jika kelas adalah baik dinas mengubah menjadi kelas 1 dan seterusnya. Skema yang digunakan dalam penanganan data kategori adalah one hot encoder. Skema ini merupakan cara yang paling sering digunakan yang membandingkan setiap tingkat variabel kategori ke tingkat referensi tetap.

BML	MAX	MIN	AVG	Periode
1000	43.8	12.0	27.9	April
50	36.0	13.4	24.7	Oktober

Gambar 5 Data yang Tidak Digunakan

Selanjutnya data yang tidak akan digunakan yaitu dengan kolom BML, maximum, minimum dan average. Data tersebut hanya nilai max didapat oleh nilai tertinggi antara periode april dan oktober data yang lebih tinggi akan ditampilkan di tabel max. data min adalah data yang ditampilkan nilai terkecil dari periode april dan oktober. Average didapat dari data rata-rata yang dijumlahkan kan dari periode april dan oktober. Sedang kan bml data yang dihasilkan oleh pertanian, disarankan oleh kedinasan tidak menggunakan data tersebut karena didapat tidak diketahui, karena pertanian saja yang tau bagaimana mendapatkan data tersebut.

PARAMETER	UNIT	April	Oktober
FISIKA			
TDS	mg/L	12.0	43.8
TSS	mg/L	36	13.4
KIMIA			
DO	mg/L	6.5	2.51
Ph	Unit	5.77	5.44
BOD	mg/L	1.88	3.49
COD	mg/L	8.28	29.10
NO ₂	mg/L	0.013	0.006
NO ₃	mg/L	0.3	1.8
NH ₃	mg/L	0.042	0.044
klorin bebas (C12)	mg/L	0.1	0.15
TP	mg/L	0.13	0.26
Phenol	mg/L	0.009	0.031
Minyak & Lemak	mg/L	0.4	0.5
MBAS Detergen	ug/L	0.086	0.087
Sulfida (H ₂ S)	mg/L	0.01	0.017
Sianida (Cn)	mg/L	0.011	0.015
Pb	mg/L	0.00198	0.00198
Selenium	mg/L	0.00177	0.00177
Seng	mg/L	0.003	0.008
Mangan	mg/L	0.015	0.015
Air Paksa	mg/L	0.000258	0.000317
Cu	mg/L	0.016	0.016
Cadmium	mg/L	0.000433	0.000433

Gambar 6. Data yang Sebelum Ditransformasi

Transformasi data digunakan untuk mengubah dataset sehingga konten informasi terbaik diambil dan dengan melakukan pengurangan atau perubahan tipe data standar sehingga data siap digunakan untuk di presentasikan ke teknik data mining. Seleksi data bertujuan untuk

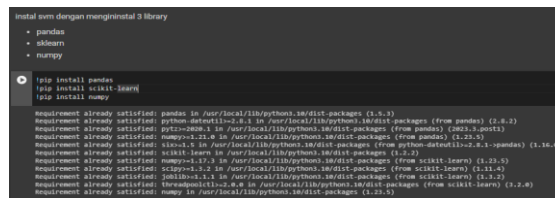
mengidentifikasi variabel-variabel relevan yang digunakan dalam penelitian. Pada data yang di ambil adalah data parameter april dan oktober, karena data periode yang benar benar data yang asli tanpa penghitungan apapun dan data periode lah yang sangat dominan dan mudah dipahami untuk dilakukanny kalsifikasi. Data parameter di ubah menjadi data perwakilan kolom pada data seperti (fisika) TDS, TSS, (kimia) DO, PH dan seterusnya. Sedangkan data april, dan oktober menjadi data periode sedangkan data kandungan air akan menjadi data parameter.

PARAM	UNITS	PARAM	UNITS	PARAM	UNITS	PARAM	UNITS	PARAM	UNITS	PARAM	UNITS	PARAM	UNITS
Cadmium	mg/L	Barium	mg/L	Fe	mg/L	Arsen	mg/L	Crom	mg/L	Cobalt	mg/L	Sulfat	mg/L
0.009		0.266	0.34	0.0012	0.015	0.068	10	40	340	4			
0.00133		0.0364	0.705	0.0026	0.017	0.068	24	45	700	2			
0.000298		0.242	0.238	0.0012	0.015	0.068	10	45	430	3			
0.000298		0.242	0.234	0.0012	0.022	0.068	15	45	490	3			

Gambar 7. Tampil Data yang Sudah ditransformasi

Dinas Lingkungan Hidup dan Pertanian akan melakukan Implementasi Data Mining dengan algoritma SVM melibatkan beberapa langkah kunci. Pertama, data relevan dikumpulkan dan fitur-fitur yang signifikan untuk tugas klasifikasi atau regresi diidentifikasi. Setelah itu, data dibagi menjadi set pelatihan dan set pengujian. *Preprocessing* data dilakukan untuk memastikan kualitas dan integritas data. Kernel SVM dan parameter seperti C dan gamma dipilih dan disesuaikan sesuai dengan karakteristik data. Model SVM dilatih dengan menggunakan set pelatihan, dan kinerjanya dievaluasi dengan menggunakan set pengujian. Jika diperlukan, parameter dan kernel dapat disesuaikan ulang untuk meningkatkan kinerja model. Akhirnya, model yang dinyatakan baik dapat diimplementasikan pada data baru untuk membuat prediksi. Dokumentasi yang baik dan pemeliharaan model juga merupakan langkah penting dalam siklus implementasi SVM untuk Data Mining.

Google Colab (Colaboratory) adalah platform berbasis cloud yang disediakan oleh Google yang memungkinkan menjalankan kode Python dalam lingkungan kolaboratif. Ini dibangun di atas Jupyter Notebooks dan memberikan akses gratis ke sumber daya GPU dan TPU. Google Colab sangat berguna untuk tugas pembelajaran mesin, karena memanfaatkan daya komputasi dari server Google tanpa memerlukan perangkat keras mahal. Dukungan untuk perpustakaan pembelajaran mesin seperti TensorFlow, PyTorch, dan scikit-learn, antara lain, membuat Google Colab menjadi pilihan populer untuk pengembangan dan pelaksanaan model pembelajaran mesin. Dukungan untuk perpustakaan sains data populer dan kemampuan untuk berbagi notebook dengan mudah juga berkontribusi pada popularitasnya.

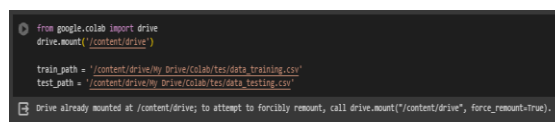


Gambar 8. Tampilan Utama Import Library SVM

Untuk menginstal pustaka pendukung dan melakukan instalasi SVM di Google Colab, pertama-tama pastikan menggunakan runtime Python 3. Kemudian, jalankan sel berikut untuk menginstal pustaka scikit-learn yang mencakup SVM. Setelah itu, menggunakan SVM dengan mengimpor pustaka tersebut dan menggunakannya untuk membuat dan melatih model SVM. Pastikan untuk memasukkan data dan menyusun langkah-langkah yang diperlukan sesuai dengan kebutuhan aplikasi atau proyek Import Data. Import data yang dilakukan bertujuan agar dinas bisa memanggil data yang akan dinas kelola, pengimportan data dilakukan dengan mengubah terlebih dahulu data yang berformat XLXS menjadi CSV karena SVM lebih dominan ke data berformat CSV.



Gambar 9. Tampilan Pemilihan Import Data



Gambar 10. Tampilan Utama Setelah Mengimpor Data Lalu Memanggil Data

Cara melakukan import data dengan melakukan upload data dengan mengklik logo pada bagian tersebut untuk menambahkan data dari manajemen file, lalu pilih data yang ingin di-import dan upload. Sebelum meng-import data ke dalam google colabs akan melakukan upload data ke dalam google drive terlebih dahulu, dengan memisahkan data training dan data testing, data training data yang tidak ada nilai skor nya sedangkan data testing ada nilai data skornya. Pemanggilan data yang dinas lakukan agar data bisa diproses dan ditampilkan.

Pemeriksaan data dengan mengecek baris dan kolomnya agar dinas bisa mengetahui banyaknya data yang akan dinas gunakan. Data akan dinas cek tipenya apa saja dan berapa banyak object, int atau pun float.

```

proses membaca file yang sudah disediakan sebelumnya

import pandas as pd

# Tentukan path file
train_path = '/content/drive/My Drive/colab/tes/data_training.csv'
test_path = '/content/drive/My Drive/colab/tes/data_testing.csv'

# Pandas membaca data dalam format CSV
data_train = pd.read_csv(train_path)
data_test = pd.read_csv(test_path)

# Menampilkan atribut yang ada di dataset
print(data_train.columns.tolist())

```

Gambar 11. Tampilan Membaca Kolom Data

Baris ini meng-*import* library pandas dan memberikan alias sebagai pd. Pandas adalah library Python yang *powerfull* untuk manipulasi dan analisis data, selanjutnya mendefinisikan jalur *file* untuk *dataset* pelatihan dan pengujian. Jalur-jalur ini menunjuk ke *file* CSV yang berada di direktori yang telah ditentukan. Baris-baris ini menggunakan fungsi *read_csv* dari pandas untuk membaca file CSV yang ditentukan oleh jalur file (*train_path* dan *test_path*) ke dalam dua objek DataFrame pandas terpisah (*data_train* dan *data_test*). DataFrame pandas adalah struktur data dua dimensi yang berlabel. Dan selanjutnya mencetak daftar nama kolom dalam dataset pelatihan. Atribut *columns* dari DataFrame berisi label kolom. Metode *tolist()* digunakan untuk mengonversi daftar label kolom ini menjadi daftar Python, dan kemudian *print* menampilkan daftar tersebut. Secara keseluruhan, kode ini membaca dua file CSV (dataset pelatihan dan pengujian) menggunakan pandas, dan kemudian mencetak nama-nama kolom dari dataset pelatihan. Ini adalah praktik umum untuk memeriksa nama kolom guna memahami struktur *dataset*.

```

fitur yang digunakan
label_train = data_train['label'].to_numpy()

fitur yang akan di proses
fitur_train = data_train[['total_sulfat_sulfat (TDS)', 'total_sulfat_sulfat (TDS)', 'ph', 'sulfida_hidrogen (PH)', 'biologi_daya_bening(BOD)',
                        'kepadatan_bahan_keras (COD)', 'mg', 'ph', 'mg', 'sulfida_besul (CS2)', 'mg', 'metana', 'metana (L)', 'metana (L)', 'metana (L)',
                        'sulfida (H2S)', 'sulfida (H2S)', 'ph', 'selena', 'mg', 'nitrat', 'mg', 'nitrat', 'mg', 'nitrat', 'mg', 'nitrat', 'mg', 'nitrat', 'mg', 'nitrat', 'mg',
                        'nitrat', 'nitrat', 'sulfida (H2S)']]

print(fitur_train)

```

Gambar 12. Tampilan Ekstraksi Label Dan Matriks Fitur

Output yang ditampilkan oleh *print(fitur_train)* adalah array NumPy yang berisi nilai-nilai parameter kualitas air yang dipilih untuk setiap baris dalam dataset. Setiap baris sesuai dengan sampel, dan setiap kolom sesuai dengan parameter kualitas air tertentu. Dari potongan kode yang berikan, tampaknya memiliki 28 kolom (fitur) dalam *dataset*. Nilai-nilai dalam array mewakili pengukuran atau tingkat parameter-parameter tersebut untuk setiap sampel. Nilai-nilai ditampilkan dalam notasi ilmiah (misalnya, 3.90e+01 mewakili 39 dalam notasi ilmiah). Matriks fitur ini (*fitur_train*) dapat digunakan untuk melatih model *machine learning* dimana tujuannya kemungkinan besar adalah memprediksi variabel skor berdasarkan nilai-nilai parameter kualitas air.

```

[[3.90e+01 3.48e+01 6.20e+00 ... 6.80e-02 1.00e+01 4.00e+01]
 [6.00e+01 3.12e+01 8.42e+00 ... 6.80e-02 2.40e+01 4.50e+01]
 [3.30e+01 3.77e+01 4.50e+01 ... 6.80e-02 1.00e+01 4.50e+01]
 ...
 [3.70e+02 1.17e+01 6.00e+00 ... 6.80e-02 1.00e+01 4.10e+01]
 [2.80e+01 3.68e+01 6.33e+00 ... 6.80e-02 1.70e+01 4.50e+01]
 [5.73e+01 6.04e+00 3.87e+00 ... 6.80e-02 8.00e+00 4.50e+01]]

```

Gambar 13. Tampilan Nilai dalam Array Mewakili Pengukuran Parameter

```

selanjutnya data training dibagi menjadi 2, yaitu:
• training set : membuat model
• validation set/test set : menguji akurasi model

library sklearn digunakan untuk membagi data secara acak menjadi training set dan validation set

from sklearn.model_selection import train_test_split

# Membagi data menjadi data latihan dan data uji
x_train, x_test, y_train, y_test = train_test_split(fitur_train, label_train, random_state=0)

```

Gambar 14. Tampilan Coding Matriks Fitur untuk Data Latih Dan Data Uji

Mengimpor fungsi *train_test_split* dari pustaka Scikit-learn. Fungsi ini digunakan untuk membagi dataset menjadi dua bagian: data latihan (untuk melatih model) dan data uji (untuk menguji performa model). Fungsi *train_test_split* digunakan di sini untuk memisahkan dataset menjadi empat bagian:

- 1) *fitur_train*: Matriks fitur yang telah dijelaskan sebelumnya.
- 2) *label_train*: Array yang berisi nilai target atau label yang telah diekstrak sebelumnya dari kolom 'skor'.
- 3) *random_state=0*: Parameter opsional yang digunakan untuk menjaga konsistensi hasil pemisahan, artinya jika menjalankan kode ini beberapa kali dengan nilai *random_state* yang sama, akan mendapatkan pembagian yang sama.
- Hasil pemisahan dataset:
 - 1) *x_train*: Matriks fitur untuk data latihan.
 - 2) *x_test*: Matriks fitur untuk data uji.
 - 3) *y_train*: Array label untuk data latihan.
 - 4) *y_test*: Array label untuk data uji.

Dengan data ini, dapat melatih model machine learning menggunakan *x_train* dan *y_train*, dan kemudian menguji performa model tersebut menggunakan *x_test* dan *y_test*. Pemisahan ini penting untuk mengevaluasi sejauh mana model dapat melakukan prediksi dengan baik pada data yang belum pernah dilihatnya sebelumnya.

```

training set kemudian digunakan untuk membuat model, menggunakan metode linear svm

from sklearn.model_selection import train_test_split
from sklearn.svm import LinearSVC
from sklearn.metrics import confusion_matrix, accuracy_score

# Membagi data menjadi data latihan dan data uji
x_train, x_test, y_train, y_test = train_test_split(fitur_train, label_train, random_state=0)

# Inisialisasi dan melatih model SVM
svm = LinearSVC()
y_pred = svm.predict(x_test)

# Melakukan prediksi pada data uji
y_pred = svm.predict(x_test)

# Membuat matriks kebingungan
conf_matrix = confusion_matrix(y_test, y_pred)
print(conf_matrix)

# Menghitung akurasi
accuracy = accuracy_score(y_test, y_pred)
print('accuracy: ', accuracy)

Confusion Matrix:
[[ 0  0  0  0]
 [ 0  3  0  0]
 [ 0  0  14  0]
 [ 0  0  0  14]]
Accuracy: 0.9762202380952381

```

Gambar 15. Akurasi Model SVM

Kode tersebut mengimplementasikan sebuah model SVM dengan kernel linier untuk melakukan klasifikasi pada *dataset* yang berisi parameter-parameter kualitas air. Pertama, *dataset* dibagi

menjadi data latih dan data uji menggunakan fungsi `train_test_split` dari Scikit-learn. Selanjutnya, model SVM diinisialisasi dengan `LinearSVC` dan dilatih menggunakan data latih. Setelah melatih model, prediksi dilakukan pada data uji, dan matriks kebingungan (`confusion matrix`) dihitung menggunakan `confusion_matrix`. Matriks kebingungan memberikan gambaran tentang performa model dalam memprediksi kelas-kelas target. Selain itu, akurasi model dihitung dengan `accuracy_score` untuk memberikan ukuran keseluruhan seberapa baik model dapat memprediksi dengan benar kelas-kelas target pada data uji. Hasil matriks kebingungan dan akurasi kemudian dicetak untuk evaluasi performa model tersebut. Hasil running model SVM pada dataset yang digunakan menunjukkan matriks kebingungan (`confusion matrix`) dan akurasi tertentu. Matriks kebingungan menggambarkan sejauh mana model dapat memprediksi setiap kelas target. Dalam matriks kebingungan ini, nilai diagonal dari kiri atas ke kanan bawah merepresentasikan jumlah prediksi yang benar untuk setiap kelas, sementara elemen di luar diagonal merepresentasikan kesalahan prediksi. Dari matriks kebingungan yang dihasilkan, terlihat bahwa model berhasil dengan baik dalam memprediksi kelas tertentu, seperti kelas ketiga dengan nilai 9 pada True Positive (prediksi benar). Namun, terdapat beberapa kesalahan prediksi, seperti pada kelas kedua yang memiliki False Negative (pengklasifikasian salah) sebanyak 1. Akurasi model, yang mengukur proporsi prediksi yang benar secara keseluruhan, diberikan nilai sekitar 67.6%. Dalam mengevaluasi model, perlu dipertimbangkan pula matriks kebingungan untuk mendapatkan pemahaman yang lebih mendalam tentang performa model pada setiap kelas target. Jika model memiliki kinerja yang tidak memuaskan pada kelas tertentu, mungkin diperlukan penyesuaian atau pemilihan model yang lebih sesuai untuk meningkatkan hasil prediksi pada dataset tersebut.

```

model yang berbentuk kamus dan di uji ke validation set (tes set)
filter_text = data_test["Teks Berisikan Kata (101)", "Teks Berisikan Kata (102)", "Teks Berisikan Kata (103)", "Teks Berisikan Kata (104)", "Teks Berisikan Kata (105)", "Teks Berisikan Kata (106)", "Teks Berisikan Kata (107)", "Teks Berisikan Kata (108)", "Teks Berisikan Kata (109)", "Teks Berisikan Kata (110)", "Teks Berisikan Kata (111)", "Teks Berisikan Kata (112)", "Teks Berisikan Kata (113)", "Teks Berisikan Kata (114)", "Teks Berisikan Kata (115)", "Teks Berisikan Kata (116)", "Teks Berisikan Kata (117)", "Teks Berisikan Kata (118)", "Teks Berisikan Kata (119)", "Teks Berisikan Kata (120)", "Teks Berisikan Kata (121)", "Teks Berisikan Kata (122)", "Teks Berisikan Kata (123)", "Teks Berisikan Kata (124)", "Teks Berisikan Kata (125)", "Teks Berisikan Kata (126)", "Teks Berisikan Kata (127)", "Teks Berisikan Kata (128)", "Teks Berisikan Kata (129)", "Teks Berisikan Kata (130)", "Teks Berisikan Kata (131)", "Teks Berisikan Kata (132)", "Teks Berisikan Kata (133)", "Teks Berisikan Kata (134)", "Teks Berisikan Kata (135)", "Teks Berisikan Kata (136)", "Teks Berisikan Kata (137)", "Teks Berisikan Kata (138)", "Teks Berisikan Kata (139)", "Teks Berisikan Kata (140)", "Teks Berisikan Kata (141)", "Teks Berisikan Kata (142)", "Teks Berisikan Kata (143)", "Teks Berisikan Kata (144)", "Teks Berisikan Kata (145)", "Teks Berisikan Kata (146)", "Teks Berisikan Kata (147)", "Teks Berisikan Kata (148)", "Teks Berisikan Kata (149)", "Teks Berisikan Kata (150)", "Teks Berisikan Kata (151)", "Teks Berisikan Kata (152)", "Teks Berisikan Kata (153)", "Teks Berisikan Kata (154)", "Teks Berisikan Kata (155)", "Teks Berisikan Kata (156)", "Teks Berisikan Kata (157)", "Teks Berisikan Kata (158)", "Teks Berisikan Kata (159)", "Teks Berisikan Kata (160)", "Teks Berisikan Kata (161)", "Teks Berisikan Kata (162)", "Teks Berisikan Kata (163)", "Teks Berisikan Kata (164)", "Teks Berisikan Kata (165)", "Teks Berisikan Kata (166)", "Teks Berisikan Kata (167)", "Teks Berisikan Kata (168)", "Teks Berisikan Kata (169)", "Teks Berisikan Kata (170)", "Teks Berisikan Kata (171)", "Teks Berisikan Kata (172)", "Teks Berisikan Kata (173)", "Teks Berisikan Kata (174)", "Teks Berisikan Kata (175)", "Teks Berisikan Kata (176)", "Teks Berisikan Kata (177)", "Teks Berisikan Kata (178)", "Teks Berisikan Kata (179)", "Teks Berisikan Kata (180)", "Teks Berisikan Kata (181)", "Teks Berisikan Kata (182)", "Teks Berisikan Kata (183)", "Teks Berisikan Kata (184)", "Teks Berisikan Kata (185)", "Teks Berisikan Kata (186)", "Teks Berisikan Kata (187)", "Teks Berisikan Kata (188)", "Teks Berisikan Kata (189)", "Teks Berisikan Kata (190)", "Teks Berisikan Kata (191)", "Teks Berisikan Kata (192)", "Teks Berisikan Kata (193)", "Teks Berisikan Kata (194)", "Teks Berisikan Kata (195)", "Teks Berisikan Kata (196)", "Teks Berisikan Kata (197)", "Teks Berisikan Kata (198)", "Teks Berisikan Kata (199)", "Teks Berisikan Kata (200)"]
y_pred = svm.predict(filter_text)
print(y_pred)

```

Gambar 16. Test Pada Data Test

Hasil prediksi kelas untuk data uji diberikan dalam array yang dicetak. Setiap angka dalam array mewakili kelas yang diprediksi untuk satu sampel dalam urutan yang sesuai dengan data uji. Sebagai contoh, hasil prediksi menunjukkan bahwa untuk beberapa sampel, kelas penilaian skor adalah 1, 2, 3, 4, 5. Penting untuk menganalisis hasil ini dengan merujuk pada interpretasi masing-masing kelas. Jika hasil prediksi memperlihatkan kelas yang signifikan, dapat diperoleh wawasan lebih lanjut mengenai karakteristik dan kondisi tertentu yang diprediksi

oleh model. Selanjutnya, hasil ini dapat dibandingkan dengan label sebenarnya (jika tersedia) untuk mengevaluasi sejauh mana model dapat memprediksi dengan akurat pada data uji yang belum pernah dilihat sebelumnya.

```

from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, classification_report

# Mengasumsikan x_train, y_train, x_test, y_test telah didefinisikan dan berisi data latih dan uji
# Menginisialisasi model SVM dengan kernel linear
clf = SVC(kernel='linear')

# Melatih model
clf.fit(x_train, y_train)

# Melakukan prediksi pada data uji
y_pred = clf.predict(x_test)

# Menghitung akurasi
clf_acc = accuracy_score(y_test, y_pred)

# Menampilkan laporan klasifikasi dan akurasi
print(classification_report(y_test, y_pred))
print('Akurasi SVM: {:.2f}%'.format(clf_acc * 100))

```

	precision	recall	F1-score	support
1	1.00	1.00	1.00	3
2	0.97	1.00	0.99	38
3	0.82	0.83	0.82	76
4	0.80	0.79	0.79	63
5	1.00	1.00	1.00	3
accuracy			0.85	183
macro avg	0.92	0.92	0.92	183
weighted avg	0.85	0.85	0.85	183

Akurasi SVM: 85.25%

Gambar 17. Hasil Akurasi Model SVM untuk Data Baru

Kode yang disediakan menggunakan model SVM dengan kernel linear untuk melakukan klasifikasi data. Pertama-tama, pustaka-pustaka yang diperlukan diimpor, termasuk `SVC` dari `scikit-learn` untuk model SVM dan beberapa metrik evaluasi seperti `accuracy_score` dan `classification_report`. Selanjutnya, model SVM dengan kernel linear diinisialisasi dan dilatih menggunakan data latih (`x_train` dan `y_train`). Setelah itu, model tersebut digunakan untuk melakukan prediksi pada data uji (`x_test`), dan akurasinya dihitung menggunakan fungsi `accuracy_score`. Hasil evaluasi model ditampilkan dalam bentuk laporan klasifikasi yang mencakup presisi, recall, dan F1-score untuk setiap kelas (dalam kasus ini, kelas 1 hingga 5). Selain itu, dilaporkan juga akurasi secara keseluruhan, serta rata-rata makro dan tertimbang dari metrik-metrik tersebut. Dari hasil running kode, terlihat bahwa akurasi model SVM pada data uji mencapai 85,25%, yang berarti model berhasil memprediksi label kelas dengan benar untuk sebagian besar instans uji. Laporan klasifikasi memberikan wawasan lebih lanjut tentang kinerja model untuk setiap kelas, dengan nilai presisi, recall, dan F1-score yang mengukur sejauh mana model dapat mengenali dan memprediksi masing-masing kelas. Keseluruhan, hasil ini memberikan gambaran tentang seberapa baik model SVM dengan kernel linear dapat menangani data klasifikasi yang diberikan.


```
import pandas as pd

# Input data baru
new_data = {'Total Dissolved Solids (TDS)': [2],
            'Total Suspended Solids (TSS)': [4],
            'DO': [12],
            'pH': [10],
            'pH': [10],
            'Biological Oxygen Demand(BOD)': [156],
            'Chemical Oxygen Demand (COD)': [18],
            'NH3': [27],
            'NO2': [9],
            'NO3': [9],
            'Klorin bebas (Cl2)': [2],
            'TP': [19],
            'Mneme': [127],
            'Mneme': [127],
            'Nitrat & Nitrit': [48],
            'Mneme': [127],
            'Sulfida (H2S)': [1],
            'Sulfida (S2)': [3],
            'PH': [4],
            'Sulfida': [8],
            'Seng': [18],
            'Mangan': [67],
            'Air Raksa': [79],
            'Cu': [14],
            'Cadmium': [13],
            'Barium': [7],
            'Fe': [34],
            'Arsen': [5],
            'Crom': [11],
            'Cobalt': [4],
            'Sulfat': [16],
            'Coliforme': [18]
            }

# Buat DataFrame dari data baru
new_data_df = pd.DataFrame(new_data)

# Tampilkan data baru
print(new_data_df)

# Output:
# Data baru:
# Total Dissolved Solids (TDS) Total Suspended Solids (TSS) DO \
# pH
# pH
# Biological Oxygen Demand(BOD) \
# COD
# Chemical Oxygen Demand (COD) NH3 NO2 NO3 Klorin bebas (Cl2) ... \
# TP
# Mneme
# Nitrat & Nitrit
# Sulfida (H2S)
# Sulfida (S2)
# PH
# Sulfida
# Seng
# Mangan
# Air Raksa
# Cu
# Cadmium
# Barium
# Fe
# Arsen
# Crom
# Cobalt
# Sulfat
# Coliforme
# [1 rows x 18 columns]
```

Gambar 18. Menginput Data Baru untuk di Prediksi

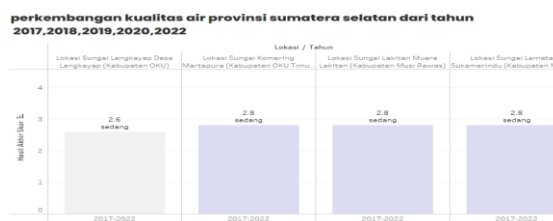
Data baru diberikan dalam bentuk *dictionary*, dimana setiap kunci adalah nama kolom dan nilai-nilai yang sesuai adalah data untuk setiap kolom. DataFrame (*new_data_df*) dibuat menggunakan pustaka *pandas* dengan menggunakan data baru yang telah diberikan. Hasil dari kode ini adalah DataFrame *new_data_df* yang berisi data baru yang telah diberikan. Setiap baris dalam DataFrame mewakili satu sampel data, dan setiap kolom sesuai dengan parameter kualitas air yang telah dijelaskan sebelumnya. Data baru ini dapat digunakan untuk membuat prediksi menggunakan model SVM yang telah dilatih sebelumnya atau untuk evaluasi keberlanjutan model pada data baru. Pada baris pertama menghitung total nilai dari setiap baris dalam DataFrame *new_data_df* dengan menjumlahkan nilai pada setiap kolom. *axis=1* menunjukkan bahwa penjumlahan dilakukan per baris. *Values [0]* digunakan untuk mengambil nilai total dari hasil perhitungan. Berdasarkan total nilai yang telah dihitung, status kualitas air ditentukan menggunakan kriteria tertentu. Dalam hal ini, status dibagi menjadi beberapa kategori berdasarkan rentang total nilai. Hasil running dari kode ini menunjukkan bahwa total nilai dari data baru adalah 624. Berdasarkan kriteria yang ditentukan, status kualitas air diinterpretasikan sebagai "Tidak Aman". Status ini ditentukan berdasarkan rentang total nilai tertentu, di mana nilai 624 berada dalam rentang yang menunjukkan tingkat kualitas air yang dianggap tidak aman. Proses ini memberikan informasi interpretatif yang dapat membantu dalam pengambilan keputusan terkait dengan kondisi kualitas air berdasarkan parameter - parameter tertentu.

Pada tahap ini dilakukan pembuatan *dashboard* Dinas Lingkungan Hidup Provinsi SumSel secara singkat dari data yang telah diproses. Data yang sudah ditransformasi dan dibersihkan dilakukan dengan teliti, akan dilakukan visualisasi data dengan aplikasi *tableau*. Berikut hasil akhir skor yang

didapat oleh provinsi sumsel dari terendah hingga tertinggi dari tahun 2017 hingga 2022. Sebelum menjelaskan data terendah dan tertinggi akan melihat tabel kategori agar bisa memutuskan kategori yang mana akan ditetapkan ke dalam setiap sungai yang ada.

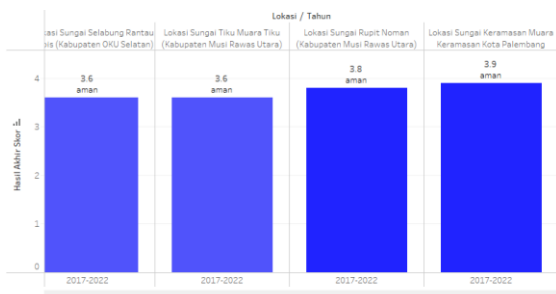
Tabel 2. Standarisasi Kualitas Air

Nilai skor akhir	Status
0.5 sampai 1.5	Sangat tidak aman
1.6 sampai 2.5	Tidak aman
2.6 sampai 3.5	Sedang
3.6 sampai 4.5	Aman
4.6 sampai 5.5	Sangat aman



Gambar 19. Sungai Dengan Nilai Terendah dari Tahun 2017 Sampai 2022

Daerah yang memiliki nilai empat terendah dimiliki oleh (Sungai Lengkyap Desa Lengkyap Kabupaten Oku) dengan nilai 2.6 saja nilai ini perbatasan dengan kategori tidak aman jika nilai yang didapat 2.5 akan mendapatkan status tidak aman. selanjutnya sungai terndah berikutnya dimiliki oleh (Sungai Komering Martapura Kabupaten Oku Timur) dengan nilai 2.8 nilai tersebut diiringi sama dengan 2 sungai lainnya yaitu Sungai Lakitan Muara Lakitan Kabupaten Musi Rawas dan Sungai Lematang Sukamerindu Kabupaten Muara Enim. Dengan penilaian tersebut diharapkan masyarakat bisa menjaga kualitas air mereka agar mereka bisa menggunakan air tersebut dengan baik. Kualitas air dengan kategori tidak aman dianggap berbahaya untuk dikonsumsi dan tidak bisa digunakan oleh masyarakat sekitar, setidaknya harus dikelola terlebih dahulu oleh PDAM sekitar, dan itupun harus diumumkan oleh PDAM sekitar apakah air itu layak minum atau tidak. Sedangkan dengan kategori sangat tidak aman, air tersebut tidak boleh disentuh atau digunakan untuk apapun, karena kandungan yang ada sangat berbahaya bagi kulit manusia seperti gatal, kulit memerah dan sebagainya. Dan jika kualitas air dengan kategori sedang, masyarakat bisa mengonsumsi air tersebut jika dikelola oleh PDAM terlebih dahulu. Walaupun sudah dikelola oleh PDAM diharuskan masyarakat mengelola air tersebut agar masyarakat tidak mengalami diare dan sebagainya.



Gambar 20. Air Dengan Nilai Tertinggi Dari Tahun 2017 Sampai 2022

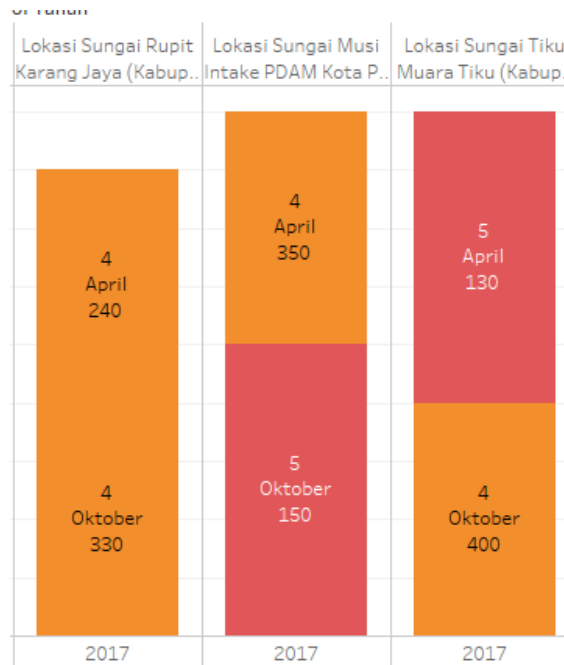
Pada gambar 20 menjelaskan kualitas air dengan nilai tertinggi dengan dominan kategori aman yaitu pada peringkat pertama yaitu di sungai Keramasan Muara Keramasan Kota Palembang, dengan nilai 3.9 nilai tersebut adalah nilai terbaik dengan 5 tahun terakhir. Selanjutnya dengan nilai 3.8 dengan Sungai Rupit Noman Kabupaten Musi Rawas Utara. Sungai ini mendapatkan kategori aman dan selanjutnya disusul oleh kedua sungai dengan nilai yang sama yaitu 3.6, dan sungai tersebut adalah Sungai Selabung Rantau Nipis Kabupaten Oku Selatan dan Sungai Tiku Muara Tiku Kabupaten Musi Rawas Utara, dengan kategori aman. Dengan nilai tersebut masyarakat harusnya bisa mengonsumsi air dari sungai tersebut dengan mengelola terlebih dahulu beda dengan kategori sangat aman, jika data dengan status sangat aman maka masyarakat bisa langsung meminum air tersebut. Jika dengan kategori aman saja, masyarakat harus mengelola air tersebut terlebih dahulu agar bisa dikonsumsi.

Sebelum menjelaskan data tertinggi dan terendah dari tahun 2017 sampai 2022 maka perlu dipahami terlebih dahulu kategori yang ada agar mudah dipahami, pada nilai total semakin tinggi kandungan yang ada pada air maka semakin rendah *point* yang didapat. Berikut semua kategori yang ada pada tabel berikut:

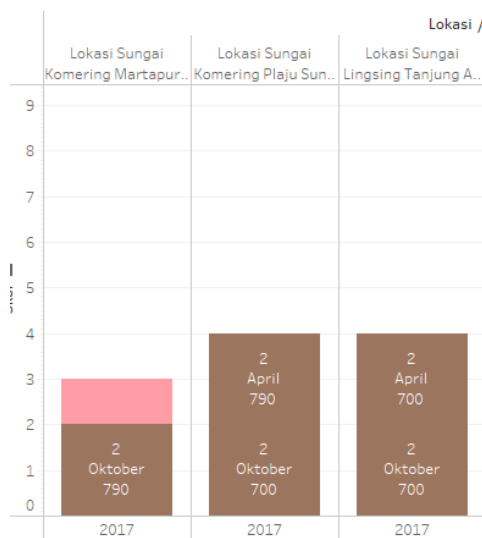
Tabel 3. kesetaraan skor dan total

Total	Skor
1-200	5 (sangat aman)
201-400	4 (aman)
401 – 600	3 (sedang)
601 – 800	2 (tidak aman)
801 - 1000	1 (sangat tidak aman)

Sampel dilakukan 2 kali pada bulan April dan bulan Oktober karena di bulan tersebut cuaca akan berubah bulan April dengan kategori cuaca hujan ke kemarau dan Oktober kemarau ke hujan, melihat tahun 2017 dengan perubahan cuaca dan pasang surut air serta kesadaran masyarakat. Melihat data tahun 2017 dengan data tertinggi sebagai berikut.



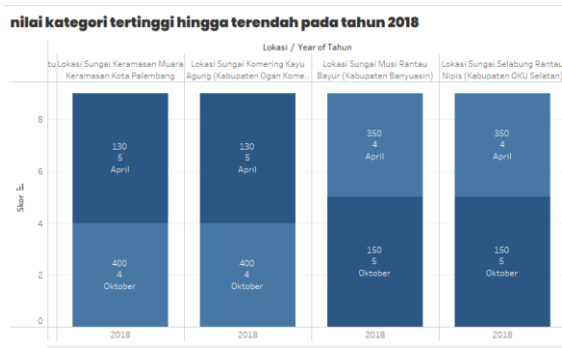
Gambar 21. Air dengan Nilai Tertinggi pada Tahun 2017



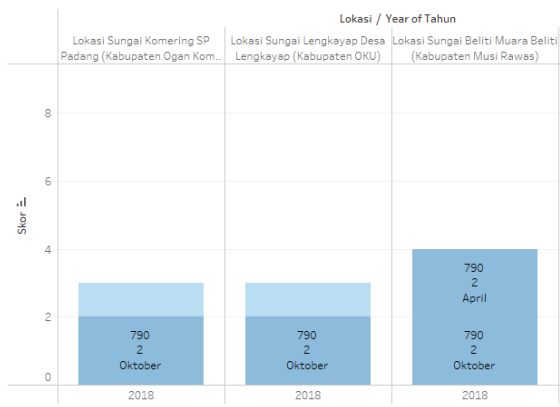
Gambar 22. Air dengan Nilai Terendah pada Tahun 2017

Penjelasan pada data tahun 2017 dengan nilai tertinggi menjelaskan Sungai Tiku Muara Tiku Kabupaten Rawas Utara mendapatkan nilai 5 pada bulan April dan nilai 4 pada bulan Oktober, kategori ini mendapatkan nilai sangat baik *point* tersebut menjadi nilai tertinggi pada tahun 2017, dengan kategori sangat aman. Sedangkan lokasi sungai intake PDAM kota Palembang mendapatkan nilai dengan 4 pada bulan April dan 5 pada bulan Oktober, kategori data ini mendapatkan status sangat aman. Sedangkan data Sungai Rupit Karang Jaya Kabupaten Musi Rawas Utara, mendapatkan nilai 4 dari bulan April dan 4 dari bulan Oktober. Status data ini mendapatkan kategori aman. Selanjutnya pada tahun 2017 dengan nilai terendah pada tahun

2017 yaitu pada lokasi Sungai Komerling Martapura Kabupaten Oku Timur, dengan nilai 3 yaitu 1 nilai dengan total 940 dari bulan April dan 2 dari bulan Oktober dengan total kandungan air 790 dengan status tidak aman. Rendahnya kesadaran masyarakat untuk menjaga lingkungan dan membuang sampah pada tempatnya sehingga masyarakat sekitar membuang limbah rumah tangga mereka ke sungai dengan begitu dampaknya masyarakat tidak akan bisa menikmati lingkungan air disekitar.



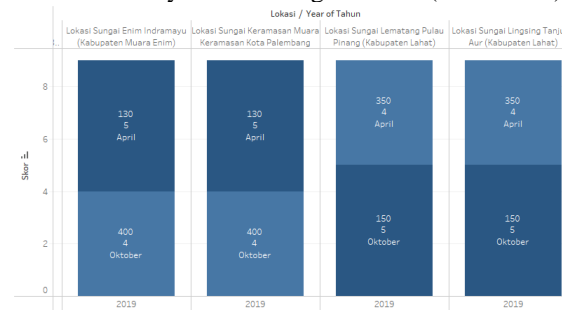
Gambar 23. Air dengan Nilai Tertinggi pada Tahun 2018



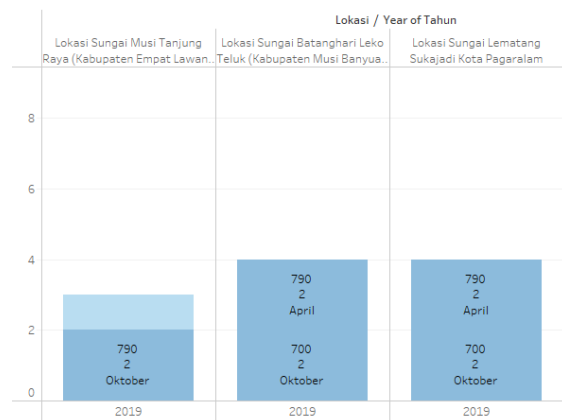
Gambar 24. Air dengan Nilai Terendah pada Tahun 2018

Tahun 2018 dengan nilai tertinggi ada di isi dengan Sungai Selabung Rantau Nipis, Sungai Musi Rantau Bayur, Sungai Komerling Kayu Agung, dan Sungai Keramasan Muara Keramasan Kota Palembang. Dengan nilai keseluruhan 9. Untuk daerah Sungai Selabung Rantau Nipis dan Rantau Bayur dimulai dengan bulan April yang memiliki nilai 4 (aman) dan Oktober mendapatkan nilai 5 (sangat aman), sebaliknya untuk Sungai Komerling Kayu Agung dan Muara Keramasan Kota Palembang mendapatkan nilai 5 pada bulan April dan 4 di bulan Oktober. Dengan adanya perkembangan dari tahun 2017 ke 2018 meningkat, diharapkan bisa memotivasi penduduk lainnya agar kesadaran kebersihan alam yang ada. Untuk tahun 2018 dengan kategori terendah yaitu dimiliki oleh

Sungai Komerling Sp Padang dan Sungai Lengkayap Desa Lengkayap. Dengan perolehan nilai 1 (sangat tidak aman) pada bulan April, dengan nilai 940 untuk kedua desa tersebut, sedangkan pada bulan Oktober kedua desa tersebut mendapatkan nilai 790 dengan nilai 2 (tidak aman). Sedangkan pada peringkat berikutnya dengan Sungai Beliti Kab. Musi Rawas, dengan mendapatkan nilai April dan Oktober sama yaitu 790 dengan nilai 2 (tidak aman).



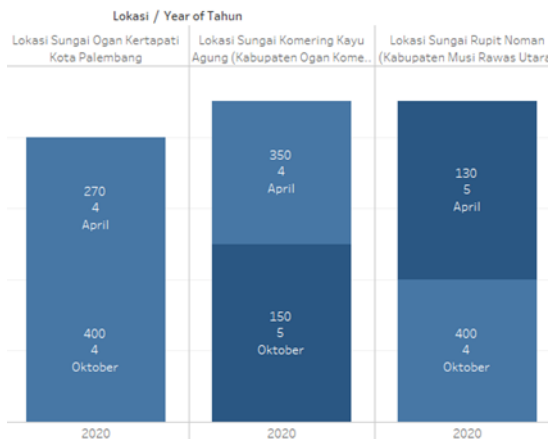
Gambar 25. Air dengan Nilai Tertinggi pada Tahun 2019



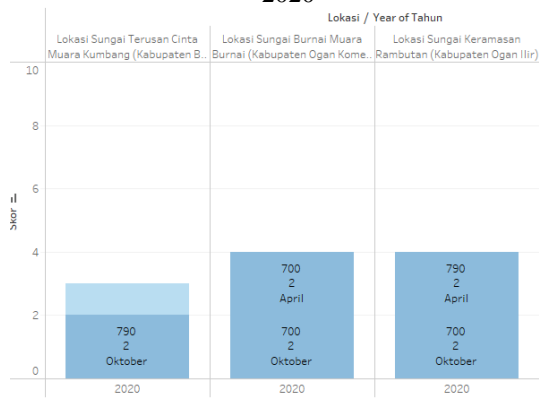
Gambar 26. Air dengan Nilai Terendah pada Tahun 2019

Tahun 2019 dengan nilai tertinggi adalah empat daerah nilai tertinggi dengan nilai yang sama yaitu Sungai Lingsing Tanjung Aur, Lematang Pulau Pinang, Sungai Keramasan Muara Keramasan Kota Palembang, Sungai Enim Indramayu. Pada Sungai Lematang Pulau Pinang dan Sungai Lingsing Tanjung Aur mendapatkan nilai 4 pada bulan April dan pada bulan Oktober mendapatkan peningkatan data menjadi 5 (sangat aman) dengan begitu pada tahun 2019 kedua sungai tersebut mengami peningkatan secara signifikan. Sedangkan Sungai Keramasan Muara Keramasan dan Sungai Enim Indramayu mendapatkan nilai yang sama tetapi mengalami penurunan di bulan Oktober di bulan April kedua sungai tersebut mendapatkan nilai 5 tetapi selanjutnya ketika pergantian musim kemarau ke hujan mengalami penurunan pada bulan Oktober yang mendapatkan nilai 4. Tahun 2019 dengan nilai

terendah dimiliki oleh Sungai Musi Tanjung Raya dengan total pada bulan April 940 dan bulan Oktober 790, dengan nilai skor 1 bulan April dan 2 di bulan Oktober, selanjutnya pada daerah Sungai Batanghari Leko Teluk dan Lematang Sukajadi Kota Pagar Alam masing masing mendapatkan total di bulan April 790 dengan skor 2 sedangkan di bulan Oktober mendapatkan total 700 dengan nilai skor 2, dengan total 4 di tahun 2019.



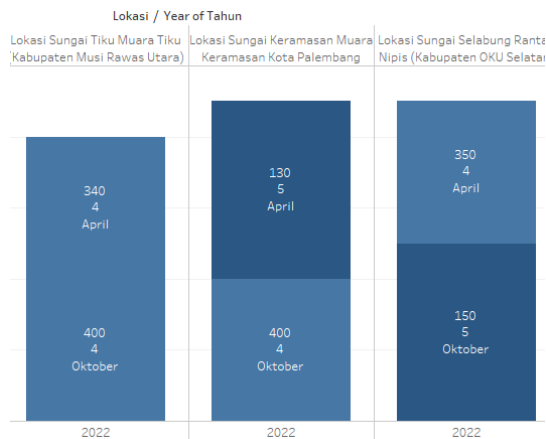
Gambar 27. Air dengan Nilai Tertinggi pada Tahun 2020



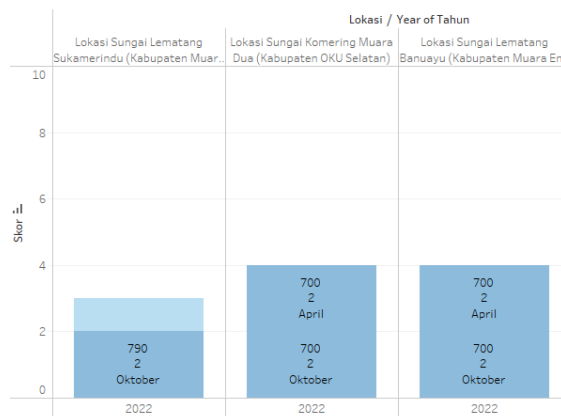
Gambar 28. Air dengan Nilai Terendah pada Tahun 2020

Selanjutnya tahun 2020 dengan kategori tertinggi yaitu dengan urutan teratas Sungai Rupit Noman dan Sungai Komerling Kayu Agung, Sungai Rupit Noman mendapatkan nilai 9 dengan bulan April mendapatkan total 130 dan skor 5 pada bulan April dan pada bulan Oktober 2020 mendapatkan 400 dengan skor 4, sedangkan Sungai Komerling Kayu Agung pada bulan April 2020 mendapatkan total 350, dan 150 pada bulan Oktober. Selanjutnya data peringkat ketiga diduduki sungai 270 pada bulan April dan mendapatkan 400 pada bulan Oktober, dengan mendapatkan nilai 4 pada April dan 4 pada Oktober. Pada gambar 28 data terendah yaitu Sungai Terusan Cinta Muara Kumbang dengan total bulan April 940 dan Oktober 790, dengan nilai 1 pada bulan April dan 2 pada bulan Oktober, dengan begitu kualitas data air termasuk tidak aman di

wilayah Sungai Terusan Cinta Muara Kumbang. Sedangkan untuk wilayah berikutnya dengan nilai 4 secara keseluruhan yaitu Sungai Burnai Muara Burnai dan Sungai Keramasan Rambutan. Pada bulan April kedua sungai ini mendapatkan nilai 2 dan juga bulan Oktober mendapatkan nilai 4 dengan status (tidak aman).



Gambar 29. Air dengan Nilai Tertinggi pada Tahun 2022



Gambar 30. Air dengan Nilai Terendah pada Tahun 2022

Pada tahun 2021 tidak dilakukannya penelitian sampel dikarenakan covid-19 pada tahun 2021 yang menyebabkan penelitian dihentikan untuk sementara waktu. Dan penelitian dilanjutkan pada tahun 2022. Data tahun 2022 dengan persentase tertinggi dimiliki oleh Sungai Selabung Rantau Nipis, dan Sungai Keramasan Muara Keramasan Kota Palembang. Dengan total pada bulan April Sungai Selabuan mendapatkan 350 total dengan skor 4 (aman) sedangkan Muara Keramasan mendapatkan total pada bulan April 130 dan skor 5 (sangat aman). Pada bulan Oktober Sungai Selabuan Rantau Nipis mendapatkan total 150 dengan skor 5 (sangat aman). Sedangkan bulan Oktober pada Sungai Keramasan Kota Palembang mendapatkan total 400 dengan skor 4 (aman). Sedangkan Sungai Selabung Rantau mendapatkan total 340 pada bulan April dengan skor 4 dan 400 pada bulan Oktober dengan skor 4. Nilai

terendah pada tahun 2022 diperoleh Sungai Lematang Sukamerindu, yang memiliki total 940 dengan skor 1 (sangat tidak aman) sedangkan pada bulan Oktober Sungai Lematang Sukamerindu mendapatkan total 790 dan skor 2 (tidak aman). Selanjutnya di ringi dengan Sungai Komerling Muara Dua dan Sungai Lematang Banyu Ayu yang sama sama mendapatkan nilai pada bulan April 700 dan skor 2 (tidak aman), dan Oktober dengan nilai yang sama 700 dan skor 2 (tidak aman).

Berikut hasil penerapan aplikasi dengan menggunakan laravel 10 dengan bantuan *database* php my admin untuk menyimpan data baru.

The screenshot shows a table with columns for 'id air', 'lokasi', 'bulan', 'tahun', and various water quality parameters (TDS, TSS, DO, PH, BOD, NO2, NO3, NH3, C12, TP, phenol, minyak, deterjen, sulfida, sianida, CN, PB, silenium, COD, seng, mangan, air raksa, CU, cadmium, barium, FE, arsen, crom, cobalt, sulfat, coliform, total, Skor, Status). The table contains three rows of data for different river locations and dates.

Gambar 31. Tampilan Data Utama

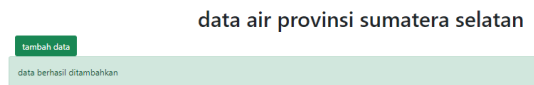
Tampilan data yaitu menampilkan data yang sudah dimasukkan ke *database* yang ada dengan menampilkan *id_air* yang dibuat, lokasi, bulan, tahun, total sampel, dan data status. Pada tampilan ini ada 4 fitur yang ditampilkan, tambah data, cari data, edit data, dan hapus data. pada setiap fitur memiliki akses berbeda-beda hanya saja cari data yang mencari data yang jika digunakan masih akan menampilkan data pada umumnya hanya saja data yang ditampilkan hanya yang dicari.

The form is titled 'Tambah Data Air' and contains input fields for ID Air, Lokasi, Bulan, Tahun, and various water quality parameters (TDS, TSS, DO, PH, BOD, NO2, NO3, NH3, C12, TP, phenol, minyak, deterjen, sulfida, sianida, CN, PB, silenium, COD, seng, mangan, air raksa, CU, cadmium, barium, FE, arsen, crom, cobalt, sulfat, coliform, total, Skor, Status). There is a 'Pilih Status Air' dropdown menu and a 'Submit' button at the bottom.

Gambar 32. Tampilan Pada Input Data

Pada tampilan tambah data, tampilan ini hasil dari tombol tambah data, data yang akan ditambah akan melakukan penambahan pada *database*, dengan menginput *id_air*, lokasi, bulan, tahun, total, skor

dan status. Ketika di-*submit* akan menambahkan secara otomatis data baru yang ada di *database*. Jika data sudah ditambahkan maka akan tampil pada bagian utama yaitu akan memberikan notif pada web yang memberitahukan bahwa data sudah ditambahkan.

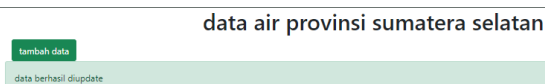


Gambar 33. Tampilan Data Berhasil Ditambahkan

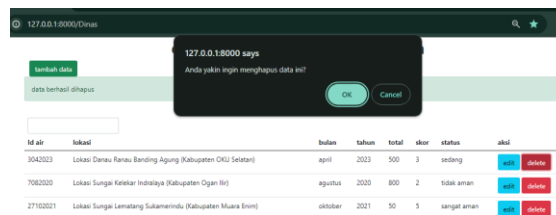
The form is titled 'update data air' and contains input fields for 'id air' (2042023), 'Lokasi' (Lokasi Sungai Batanghari Leko Teluk (Kabupaten Musi Banyuasin)), 'bulan' (april), 'tahun' (2023), 'total' (400), 'skor' (3), and 'status' (sedang). There is a 'Submit' button at the bottom.

Gambar 34. Update Data

Pada tampilan *update* data yaitu jika ada kesalahan pada input yang dilakukan oleh anggota maka data masih bisa di-*edit* atau di-*update*, data akan di-*update* jika diperlukan dan seluruh data bisa di-*update* *id_air*, lokasi, bulan, tahun, total skor dan status. Data jika sudah di-*update* maka *database* juga akan di-*update* secara otomatis. Jika data sudah di-*update* tampilan yang sama akan ditampilkan pada tabel tersebut, dan menampilkan ulang data air yang ada.



Gambar 35. Data Berhasil di-update



Gambar 36. Tampilan Konfirmasi Hapus Data

Pada tampilan *delete* data yang dilakukan jika data sudah tidak diperlukan lagi, data akan dihapus jika diperlukan, *delete* data akan mengkonfirmasi terlebih dahulu data yang akan dihapus karena meminimalisir data yang dihapus dengan tidak sengaja. jika dipilih oke akan dilakukan penghapusan data dan jika tidak maka data yang ada di *database* tidak akan dihapus.

V. KESIMPULAN DAN SARAN

Kesimpulan

Selama lima tahun terakhir dinas lingkungan hidup dan pertanahan telah melakukan penelitian terkait sampel kualitas air di Sumatera Selatan. Dengan melakukan dua kali survei yaitu pergantian musim hujan ke musim kemarau pada bulan April dan bulan Oktober dengan pergantian musim kemarau ke musim hujan. Walaupun demikian setidaknya pelaksanaan penelitian yang dilakukan dinas lingkungan hidup dan pertanahan sudah membantu masyarakat agar mengerti kualitas alam dan air yang ada di wilayah mereka masing-masing.

Dengan begitu, dapat disimpulkan bahwa sungai yang ada di Sumatera Selatan dengan 73 titik yang berbeda Sungai Sumatera Selatan dominan dengan 89 % memiliki kualitas sedang, artinya masih banyak masyarakat yang masih rendah untuk memahami alam mereka sendiri, kurangnya kesadaran membuang limbah dan kotoran ke tempat yang baik dan kurangnya pemahaman masyarakat untuk menjaga sungai mereka sendiri. Sedangkan sungai dengan kualitas aman dengan 10 titik saja yaitu 11%, diharapkan masyarakat bisa bertahan dan terus berkembang kualitas sungainya agar mereka bisa memotivasi dan mendorong masyarakat lainnya. Agar mendapatkan kualitas air yang baik pada sungai lainnya. Dengan begitu masyarakat bisa menikmati air yang baik dimanapun berada.

Saran

Mengoptimalkan pemanfaatan teknik Data Mining, terutama menggunakan algoritma SVM, dalam melakukan klasifikasi kualitas air. Rekomendasi pemahaman mendalam terkait dengan parameter kualitas air yang dapat diidentifikasi dan dioptimalkan melalui pendekatan ini. Untuk memperkuat integrasi data dari berbagai sumber yang relevan untuk meningkatkan akurasi dan keberlanjutan model klasifikasi. Dibutuhkan pendekatan yang menyeluruh untuk memastikan keberlanjutan dan keterpercayaan data yang digunakan.

Dengan memperbanyak sosialisasi, edukasi dan praktek pada masyarakat yang dilakukan oleh dinas lingkungan hidup dan pertanahan provinsi sumsel, kemungkinan besar masyarakat bisa terdorong untuk menjaga kualitas air mereka di setiap sungai. Diharapkan agar dinas lingkungan

hidup dan pertanahan Provinsi Sumatera Selatan bisa memperluas pencarian titik terbaru untuk diteliti agar seluruh sungai Provinsi Sumatra Selatan mendapatkan fasilitas dan edukasi lebih leluasa.

DAFTAR PUSTAKA

- Ahmad, T., Purwono, and Gata, W. (2022). "Model Prediksi Kualitas Udara dengan Support Vector Machines dengan Optimasi Hyperparameter GridSearch CV," *Buletin Ilmiah Sarjana Teknik Elektro*, no. 4, pp. 12-21.
- Darmawan, A., Kustian, N. and Rahayu, W. (2018). "Implementasi data mining menggunakan model SVM untuk prediksi kepuasan pengunjung taman tabebuaya." *STRING (Satuan Tulisan Riset dan Inovasi Teknologi) 2.3* (2018): 299-307.
- Fadilah, Ulul, W.R., Agfiannisa, D. and Azhar, Y. (2020). "Analisis Prediksi Harga Saham PT. Telekomunikasi Indonesia Menggunakan Metode Support Vector Machine." *Fountain of Informatics Journal 5.2*: 45-51.
- Batubara, Nabila, D., Windarto, A.P., and Irawan, E.(2022). "Analisis Prediksi Keterlambatan Pembayaran Listrik Menggunakan Komparasi Metode Klasifikasi Decision Tree dan Support Vector Machine." *JURIKOM (Jurnal Riset Komputer) 9.1*: 102-108.
- Irmanda, Nurramdhani, H. and Astriratma, R. (2020). "Klasifikasi Jenis Pantun Dengan Metode Support Vector Machines (SVM)." *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi) 4.5*: 915-922.
- Rooy, T., Kusriani, and Luthf, E.T. (2020). "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma Support Vector Machine." *Jurnal FATEKSA: Jurnal Teknologi dan Rekayasa 5.2*: 20-29.
- Bode, Andi. (2019). "Perbandingan metode prediksi support vector machine dan linear regression menggunakan backward elimination pada produksi minyak kelapa." *Simtek: jurnal sistem informasi dan teknik komputer 4.2*: 104-107.
- Mardewi, et. al. "Klasifikasi Kategori Obat Menggunakan Algoritma Support Vector Machine." *Journal Pharmacy and Application of Computer Sciences 1.1* (2023): 27-32.
- Naldy, Tachi, E. and Andri. (2021). "Penerapan Data Mining untuk Analisis Daftar Pembelian Konsumen Dengan Menggunakan Algoritma Apriori pada Transaksi Penjualan Toko Bangunan MDN." *Jurnal Nasional Ilmu Komputer 2.2*: 89-101.
- Putra, et.al. (2023). *Data Mining: Algoritma dan Penerapannya*. PT. Sonpedia Publishing Indonesia.