

BAB 1

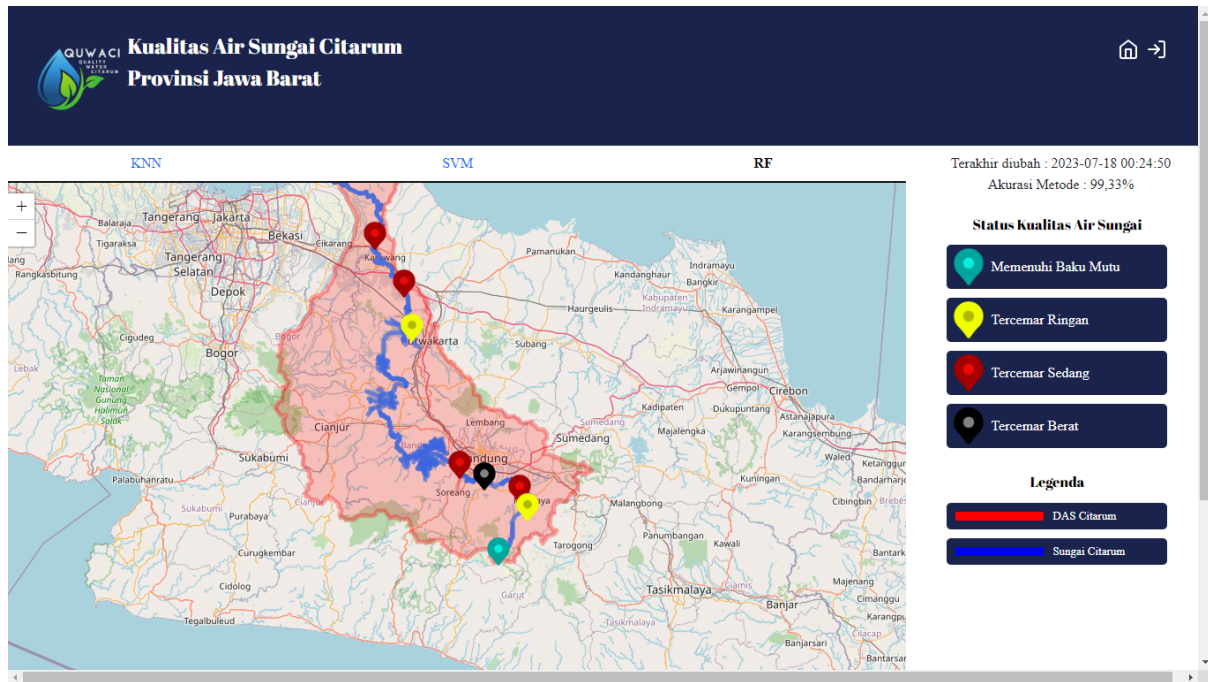
USULAN GAGASAN

1.1 Deskripsi Umum Masalah dan Kebutuhan

Sungai Citarum merupakan sungai terpanjang dan terbesar di Jawa Barat, membentang dari Gunung Wayang, Kabupaten Bandung hingga Muara Gembong, Kabupaten Bekasi[1]. Secara Geografis Wilayah Sungai Citarum terletak pada $106^{\circ} 51'36''$ - $107^{\circ} 51'$ BT dan $7^{\circ} 19'$ - $6^{\circ} 24'$ LS, dengan luas area ± 11.323 Km²[2]. Wilayah Sungai Citarum seluas kurang lebih 12.000 km² mencakup 13 wilayah administrasi Kabupaten/Kota di lingkungan Provinsi Jawa Barat, yaitu: Kabupaten Bandung, Kabupaten Bandung Barat, Kabupaten Bekasi, Kabupaten Cianjur, Kabupaten Bogor, Kabupaten Indramayu, Kabupaten Karawang, Kabupaten Purwakarta, Kabupaten Subang, Kabupaten Sumedang, Kota Bandung, Kota Bekasi dan Kota Cimahi[2].

Saat ini Sungai Citarum telah mengalami degradasi kualitas dan kuantitas yang sangat memprihatinkan. Seiring dengan peningkatan aktivitas masyarakat di sekitarnya, pencemaran Sungai Citarum terus meningkat[3]. Diketahui bahwa sepanjang 127 km atau 47,1% dari panjang sungai telah dikategorikan tercemar berat[3]. Sungai Citarum dicemari oleh kurang lebih 20,000 ton sampah dan 340,000 ton air limbah dengan mayoritas penyumbang limbah tersebut berasal dari 2,000 industri tekstil[7].

Sebuah sistem informasi yang bertujuan untuk mengklasifikasikan kualitas air sungai Citarum harus dikembangkan, bersamaan dengan upaya pemetaan sungai tersebut. Tujuannya adalah untuk mengatasi masalah pencemaran dan dampaknya pada lingkungan dan masyarakat setempat. Melalui sistem informasi ini, masyarakat akan memperoleh akses lebih baik terhadap data mengenai kualitas air sungai, yang pada gilirannya diharapkan dapat meningkatkan kesadaran mereka tentang kondisi sungai. Dengan peningkatan kesadaran ini, diharapkan masyarakat akan lebih proaktif dalam menjaga dan melindungi kesehatan sungai, serta menghindari aktivitas yang dapat merusaknya.



Gambar 1.1 Tampilan *Website* Quwaci

Berdasarkan beberapa permasalahan yang terjadi di Sungai Citarum, mengusulkan solusi berupa pengembangan Sistem Informasi mengenai klasifikasi mutu air sungai. Ini adalah hasil kerja sama antara rekan kelompok, pembimbing Prodi S1 teknik komputer, pembimbing prodi S1 teknik elektro, serta dari Dinas Lingkungan Hidup dan Kebersihan Provinsi Jawa Barat. Pada bagian Gambar 1.1 terdapat KNN, SVM, dan RF. Tiga algoritma *machine learning* yang digunakan untuk berbagai tugas dalam *data science*. KNN adalah algoritma yang berdasarkan pada prinsip bahwa objek dengan atribut serupa cenderung berada dalam jarak yang dekat satu sama lain. SVM, merupakan algoritma yang digunakan untuk pemisahan data dengan mencari *hyperplane* terbaik yang memisahkan dua kelas. Sedangkan *Random Forest* (RF) adalah algoritma *ensemble* yang terdiri dari sejumlah pohon keputusan yang digunakan untuk menggabungkan prediksi dari masing-masing pohon keputusan dalam rangka meningkatkan akurasi prediksi dan mengatasi *overfitting*.

1.2 Analisa Masalah

1.2.1 Aspek Lingkungan

Dalam Peraturan Presiden No. 15 pada tahun 2018 Tentang percepatan Pengendalian Pencemaran dan Kerusakan Daerah Aliran Sungai Citarum untuk mempercepat pemulihan Sungai Citarum dari pencemaran yang terjadi. Dimana Data menunjukkan kondisi air 54% sungai Citarum tercemar berat, 23% tercemar sedang, 20% tercemar ringan dan hanya 3%

yang memenuhi baku mutu[4]. Berdasarkan beban pencemaran eksisting sudah melampaui DTBP, dimana contohnya seperti di Sub DAS Cikapundung yang sudah mencapai 77.341,19 kg/hari, sementara daya tampungnya hanya 19.335,30 kg/hari. Sehingga apabila ditotalkan keseluruhannya dari hulu ke hilir penurunan beban pencemaran Sungai Citarum harus mencapai 303.552,30 kg/hari[4].

Pemantauan terakhir di beberapa titik Sungai Citarum menunjukkan Kebutuhan oksigen biologis maksimal mencapai 2,284 kilogram per hari dan kebutuhan oksigen kimia mencapai 10,673 kilogram per hari. Sedangkan kadar besi (Fe) terlarut maksimal 23,4 kilogram per hari, mangan (Mn) 8,29 kilogram per hari, tembaga (Cu) mencapai 51,1 kilogram per hari dan seng (Zn) mencapai 57,3 kilogram per hari[5]. Melihat angka-angka tersebut, terbukti kualitas air Sungai Citarum sangat rawan dan sudah tidak layak digunakan sebagai sumber air minum, dan untuk perikanan, walaupun masih layak dimanfaatkan untuk pertanian. Kondisi tersebut merupakan faktor-faktor yang menyebabkan terjadinya pencemaran di perairan Sungai Citarum[5]. Akibat pencemar ini mengakibatkan terjadinya masalah lingkungan di perairan Sungai Citarum yaitu terlampauinya Daya dukung lingkungan hidup, Daya tampung lingkungan hidup dan daya lenting lingkungan[5].

1.2.2 Aspek Kesehatan

Masyarakat yang tinggal di bantaran Sungai Citarum memiliki permasalahan kesehatan yang memprihatinkan, akibat mengkonsumsi air sungai yang tidak tepat sebagai akses dari pola perilaku yang buruk. Wangsaatmaja dalam Cahyaning pada tahun 2009 mengungkapkan, di bantaran Sungai Citarum telah dibuktikan ada hubungan yang nyata antara lokasi di sepanjang sungai (hulu-hilir) dengan penyakit bawaan air [6]. Resiko menderita penyakit bawaan tertinggi ada di Cisirung dengan nilai sebesar 276 untuk penyakit kulit, dan 14,636 untuk diare. Katapang dan Nanjung, nilai penyakit kulit 0,866 dan 0,479 dan diare sebesar 1,178 dan 2,029. Andir dan Cijeruk, untuk penyakit kulit 26,833 dan 1,568 serta diare sebesar 5,664 dan 1,178[6].

Sehingga dapat kita simpulkan bahwa masyarakat di sepanjang Sungai Citarum memiliki masalah kesehatan yang serius akibat mengkonsumsi air sungai yang tidak layak, dengan tingkat risiko tertinggi terkait penyakit kulit dan diare di beberapa lokasi seperti Cisirung, Katapang, Nanjung, Andir, dan Cijeruk. Hal ini didukung oleh penelitian Wangsaatmaja dalam Cahyaning pada tahun 2009 yang menunjukkan adanya hubungan yang signifikan antara lokasi di sepanjang sungai dengan penyakit bawaan air.

1.3 Analisa Solusi yang Ada

Pada penelitian yang sudah dilakukan sebelumnya digunakan 3 metode untuk mengklasifikasikan Air Sungai Citarum. Ketiga metode algoritma *machine learning* yaitu KNN, SVM, dan Random Forest serta ketiga hasil klasifikasi akan ditampilkan pada *website*. Berdasarkan penelitian dari Siti An-nisa, Putri Zahrani, dan Delatifa Sugandi pada tahun 2022 didapatkan akurasi KNN sebesar 92%, SVM didapatkan akurasi sebesar 95%, dan *Random Forest* menghasilkan akurasi sebesar 99%.

1.3.1 Karakteristik Produk *Websites*

Sistem Informasi dapat menyajikan informasi terkait visualisasi data klasifikasi kualitas air Sungai Citarum melalui platform *website* yang dapat diakses melalui PC/Laptop dan Handphone. Solusi yang dapat digunakan untuk menyelesaikan permasalahan yang terjadi adalah dengan mempertimbangkan tiga aspek penting seperti keunggulan (*strength*), kekurangan (*weakness*), dan keterbatasan (*limitation*).

1.3.1.1 Keunggulan Produk *Websites*

Dalam informasi kualitas air sungai dengan *website*, pengolahan data yang kompleks pada sistem mampu untuk mengumpulkan, mengintegrasikan, dan menganalisis berbagai variabel yang mempengaruhi kualitas air. Ini mencakup data yang sangat beragam pada target parameter lingkungan air sungai.

Kemampuan prediksi model berbasis *website* termasuk kemampuan untuk memprediksi perubahan kualitas air sungai di masa depan dengan menggunakan model *machine learning* yang dibuat. Ini berarti bahwa model ini tidak hanya dapat menganalisis data masa lalu tetapi juga dapat membuat perkiraan tentang bagaimana kualitas air akan berkembang di masa depan. Ini sangat penting untuk Sungai Citarum karena memungkinkan pengawasan yang lebih baik atas perubahan kualitas air.

Deployment machine learning di *website* memiliki beberapa kelebihan dibandingkan dengan aplikasi *mobile*. Berikut adalah beberapa kelebihan dari *deployment machine learning* di *website* yaitu akses lintas berbagai platform *gadget*, tidak perlu mengunduh atau menginstal aplikasi, mudah memperbarui model *machine learning* atau fitur dalam web, dapat menambahkan sumber daya server agar memudahkan aplikasi web untuk diskalakan sehingga lebih banyak pengguna yang dapat ditangani secara bersamaan, dan aplikasi web

mudah untuk diintegrasikan karena lebih lancar dengan layanan lain seperti layanan pihak ketiga atau disebut sebagai *Application Programming Interface* (API) karena mengikuti standar protokol web.

1.3.1.2 Kekurangan Produk *Websites*

Website berbasis prediksi model memiliki beberapa kekurangan yang perlu diperhatikan. Pertama, akurasi model sangat bergantung pada kualitas data pelatihan; data yang tidak representatif atau akurat dapat menyebabkan prediksi yang tidak akurat. Selain itu, model harus diperbarui secara berkala untuk mempertahankan akurasi, yang membutuhkan pengawasan dan pemeliharaan yang berkelanjutan. Selain itu, kemampuan model untuk memprediksi perubahan kualitas air yang tak terduga, seperti cuaca ekstrem atau adanya insiden polusi.

Pengelolaan model ini memerlukan keahlian dalam analisis data dan statistik, yang dapat membuat pengembangan dan pemeliharaan menjadi lebih sulit. Data yang dihasilkan oleh model tidak selalu real-time, dan pengguna harus memiliki koneksi internet, yang dalam beberapa kasus dapat menjadi tantangan. Walaupun ada kekurangan, memahami dan mengatasi masalah ini dapat membantu memaksimalkan manfaat dari *website* berbasis prediksi model untuk pemantauan dan manajemen kualitas air.

Berikut kekurangan *deployment machine learning* pada *website* seperti keterbatasan akses *online* yang memerlukan koneksi internet stabil, kinerja yang mungkin kurang responsif, pengalaman pengguna yang tidak konsisten bergantung pada perangkat dan peramban, kemungkinan kerentanan keamanan seperti injeksi XSS dan serangan CSRF, dan masalah dalam pengembangan UI/UX yang lebih kompleks.

1.3.1.3 Keterbatasan Produk *Websites*

Keterbatasan yang muncul dalam memprediksi kualitas air adalah tidak mungkin untuk memperkirakan perubahan yang tak terduga setiap saat. Meskipun telah dilakukan banyak upaya untuk mengembangkan teknologi dan metode untuk memonitor dan memprediksi kualitas air, faktor-faktor seperti perubahan cuaca ekstrem, aktivitas manusia yang tidak terduga, atau masalah teknis dengan peralatan pemantauan dapat menyebabkan prediksi yang tidak akurat. Oleh karena itu, penting untuk terus melakukan pencegahan dan memantau kualitas air dengan cermat, meskipun kita tidak dapat memprediksi perubahan tak terduga ini.

Keterbatasan *deployment machine learning* dengan *website* dibanding dengan aplikasi *mobile* yang dapat berdampak pada pengembangan dan pengalaman pengguna yaitu aplikasi *mobile* memungkinkan sistem operasi perangkat berinteraksi lebih langsung, memungkinkan penggunaan notifikasi, dan memungkinkan pengembang membuat pengalaman yang lebih terintegrasi. Selain itu, aplikasi *mobile* dapat menyimpan data secara lokal bahkan saat perangkat tidak terhubung ke internet. Sementara aplikasi *website* mungkin bergantung pada penyimpanan *cloud* atau keterbatasan penyimpanan browser yang kurang fleksibel. Akibatnya, aplikasi *website* mungkin tidak dapat menampilkan konten atau hasil *machine learning* saat perangkat tidak terhubung ke internet.

Pengembang aplikasi *mobile* memiliki lebih banyak kebebasan untuk membuat pengalaman pengguna yang mirip dengan aplikasi *native*, termasuk antarmuka pengguna yang lebih kaya, animasi yang lebih halus, dan navigasi yang lebih responsif. Dalam beberapa situasi, ini dapat membuat aplikasi *mobile* lebih menarik dan efektif. Sebaliknya, aplikasi *website* mungkin menghadapi batasan dalam mencapai tingkat kehalusan dan kedalaman ini. Pada aplikasi *mobile* dapat lebih mudah terintegrasi dengan ekosistem perangkat, seperti asisten suara, yang memberikan pengalaman pengguna yang lebih terhubung dengan perangkat. Aplikasi *website* mungkin mengalami keterbatasan dalam hal integrasi semacam ini. Selain itu, perlu diperhatikan bahwa aplikasi web lebih rentan terhadap serangan keamanan, seperti XSS dan CSRF. Akibatnya pengembang harus mempertimbangkan aspek-aspek ini ketika memilih platform untuk menerapkan *machine learning* di dalam aplikasi yang ingin dikembangkan nantinya.

1.3.2 Karakteristik Sistem Model *Machine Learning*

Model *machine learning* dapat digunakan untuk klasifikasi air sungai untuk memprediksi kualitas atau kebersihan air berdasarkan data seperti konsentrasi zat terlarut, suhu, pH, dan parameter lain dari lingkungan. Proses pelatihan melibatkan penggunaan data historis atau saat ini untuk mengidentifikasi pola yang terkait dengan tingkat kebersihan air. Model ini dapat digunakan untuk memantau kualitas air secara *real-time*, menemukan pencemaran atau perubahan yang tidak diinginkan, dan memberikan peringatan dini kepada pihak berwenang atau pemangku kepentingan untuk memastikan keberlanjutan dan keamanan sumber daya air sungai. Model algoritma *machine learning* yang dipakai yaitu *Weighted KNN with Euclidean Distance*, *Gaussian Naive Bayes*, dan *Artificial Neural Network*.

1.3.2.1 Keunggulan Setiap Algoritma *Machine Learning*

Dalam klasifikasi kualitas air sungai, *Weighted K-Nearest Neighbors (Weighted KNN)* dengan *Euclidean Distance* memiliki keunggulan dalam menyesuaikan bobot tetangga terdekat yang digunakan untuk mengklasifikasikan data berdasarkan variasi data sampel air. Algoritma ini memberikan bobot yang lebih besar pada tetangga yang lebih dekat, menyiratkan bahwa pengaruh tetangga terdekat terhadap keputusan klasifikasi lebih signifikan daripada tetangga yang lebih jauh. Hal ini membuat *Weighted KNN* dengan *Euclidean Distance* menjadi pilihan yang tepat untuk aplikasi pemantauan kualitas air sungai yang sering mengalami perubahan lingkungan [70].

Gaussian Naive Bayes dapat mengelola tingkat ketidakseimbangan yang umum terjadi pada data lingkungan seperti air sungai. *Gaussian Naive Bayes* dapat membuat perkiraan yang lebih akurat untuk kategori kualitas air dengan mengambil pendekatan yang lebih fleksibel terhadap data yang tidak seimbang. Hal ini sangat penting terutama ketika kelompok minoritas, seperti mereka yang terkena dampak air kotor, kurang terwakili dalam statistik. Teknik ini sangat berhasil karena dapat memperhitungkan distribusi Gaussian dari fitur-fiturnya, sehingga ideal untuk aplikasi pemantauan kualitas air sungai yang memiliki ketidakseimbangan tingkat yang lazim [71].

Artificial Neural Network (ANN) dapat menginterpretasikan dan memahami pola data lingkungan sungai yang rumit. Dengan metode ini dan data yang cukup, ANN dapat mengenali faktor kualitas air seperti tingkat polusi, suhu, dan pH. Temuan ini memungkinkan para peneliti dan pengelola sumber daya air untuk melakukan upaya yang tepat dalam melestarikan dan meningkatkan kualitas air sungai, sehingga menghasilkan ekosistem yang lebih sehat dan berkelanjutan. ANN sangat berhasil dalam menangani data non-linear dan dapat belajar dan beradaptasi dengan perubahan data, menjadikannya alat yang sangat baik untuk memantau kualitas air sungai [72].

1.3.2.2 Kekurangan Setiap Algoritma *Machine Learning*

Weighted K-Nearest Neighbors (Weighted KNN) yang menggunakan *Euclidean Distance* memiliki keterbatasan dalam hal skalabilitas dan efisiensi pemrosesan, terutama ketika digunakan dengan set data yang besar. Pendekatan ini mengharuskan penghitungan jarak antara semua titik dalam data pelatihan untuk setiap prediksi, yang dapat menghabiskan banyak waktu dan sumber daya. Selain itu, *Weighted KNN* sensitif terhadap fitur dengan skala yang berbeda-beda, sehingga memerlukan standarisasi data sebelum diterapkan. Dan

juga algoritma ini rentan terhadap outlier yang bisa mempengaruhi keputusan klasifikasi secara signifikan [70].

Gaussian Naive Bayes membuat asumsi yang signifikan tentang independensi fitur, yang jarang terjadi dalam praktiknya. Ketika karakteristik data sangat terkait, kinerja *Gaussian Naive Bayes* dapat menurun. Selain itu, asumsi bahwa data mengikuti distribusi Gaussian mungkin tidak dapat diterapkan pada semua bentuk data lingkungan, seperti kualitas air sungai. Jika distribusi data aktual tidak sesuai dengan asumsi model, model dapat menghasilkan prediksi yang kurang akurat. Kerugian lainnya adalah kerentanannya terhadap data yang tidak seimbang, namun *Gaussian Naive Bayes* memiliki kinerja yang lebih baik daripada *Naive Bayes* standar [71].

Artificial Neural Network (ANN) memiliki berbagai kekurangan yang harus diatasi. Pertama, ANN membutuhkan sejumlah besar data pelatihan berkualitas tinggi untuk menghasilkan model yang andal. Jika data yang diberikan tidak memadai atau tidak representatif, model dapat membuat prediksi yang tidak akurat. Kedua, proses pelatihan ANN membutuhkan banyak waktu dan sumber daya komputer, terutama untuk jaringan yang rumit. Ketiga, ANN rentan terhadap *overfitting*, yang terjadi ketika model berkinerja sangat baik pada data pelatihan tetapi buruk pada data baru. Terakhir, ANN terkadang disebut sebagai "*black box*" karena sulit untuk memahami dan menjelaskan bagaimana model membuat prediksi, yang dapat menjadi masalah jika penjelasannya sangat penting [72].

1.3.2.3 Keterbatasan Setiap Algoritma *Machine Learning*

Weighted K-Nearest Neighbors (*Weighted KNN*) yang menggunakan *Euclidean Distance* memiliki keterbatasan ketika berhadapan dengan *dataset* yang besar karena perlu menghitung jarak untuk setiap titik data dalam proses prediksi, yang mungkin memakan waktu dan sumber daya yang intensif. Pendekatan ini juga kurang berhasil ketika berhadapan dengan data berdimensi tinggi karena dapat mengakibatkan "*curse of dimensionality*", di mana jarak antara titik data menjadi kurang signifikan. Selain itu, *Weighted KNN* membutuhkan normalisasi fitur sebelum digunakan, karena fitur berskala lebih besar dapat mendominasi perhitungan jarak, dan algoritma ini sensitif terhadap *outlier* yang dapat mempengaruhi hasil klasifikasi [70].

Gaussian Naive Bayes memiliki batasan yang signifikan dalam asumsi independensi fitur, yang jarang terjadi pada data nyata. Jika fitur-fitur dalam dataset berkorelasi, kinerja *Gaussian Naive Bayes* dapat menurun drastis. Selain itu, pendekatan ini mengasumsikan bahwa data memiliki distribusi Gaussian, yang mungkin tidak berlaku untuk semua bentuk

data, termasuk data kualitas air sungai. Kelemahan lainnya adalah kerentanan model terhadap data yang tidak seimbang, di mana kelas-kelas minoritas mungkin kurang terwakili, yang menghasilkan prediksi yang kurang akurat untuk kelas-kelas tersebut [71].

Artificial Neural Network (ANN) memiliki banyak kendala yang harus dipertimbangkan. Pertama-tama, tergantung pada ukuran dan kompleksitas model, ANN dapat menjadi sangat rumit, membutuhkan banyak parameter. Hal ini dapat menyebabkan durasi pelatihan yang lama, terutama pada *dataset* yang besar. Kedua, ANN rentan terhadap *overfitting*, di mana model dapat "menghafal" data pelatihan daripada mempelajari pola umum, yang mengakibatkan kinerja yang buruk pada data baru. Ketiga, dalam keadaan tertentu, menginterpretasikan model ANN dapat menjadi masalah karena hubungan antara *input* dan *output* tidak selalu jelas, terutama pada model yang sangat dalam. Akibatnya, penggunaan ANN harus dinilai secara hati-hati dengan mempertimbangkan fitur dan persyaratan *dataset* yang bersangkutan [72].

1.4 Kesimpulan

Sungai Citarum merupakan sungai terpanjang dan terbesar di Jawa Barat, membentang dari Gunung Wayang, Kabupaten Bandung hingga Muara Gembong, Kabupaten Bekasi. Wilayah Sungai Citarum mencakup 13 wilayah administrasi Kabupaten/Kota di Provinsi Jawa Barat, dengan luas area sekitar 11.323 km². Saat ini, sungai ini mengalami degradasi kualitas dan kuantitas yang sangat memprihatinkan karena peningkatan aktivitas masyarakat di sekitarnya, yang telah menyebabkan peningkatan pencemaran.

Pada penelitian yang sudah dilakukan sebelumnya digunakan tiga metode untuk mengklasifikasikan Air Sungai Citarum. Ketiga metode algoritma *machine learning* yaitu KNN, SVM, dan *Random Forest* serta ketiga hasil klasifikasi akan ditampilkan pada *website* yang bernama Quwaci. Hasil penelitian yang dilakukan pada tahun 2022 didapatkan akurasi KNN sebesar 92%, SVM didapatkan akurasi sebesar 95%, dan *Random Forest* menghasilkan akurasi sebesar 99%.

Berdasarkan permasalahan diatas, memutuskan untuk menggunakan tiga metode algoritma *machine learning* untuk melakukan klasifikasi air sungai yaitu *Weighted KNN with Euclidean Distance*, *Gaussian Naive Bayes*, dan *Artificial Neural Network*. Solusi ini diharapkan dapat mengatasi permasalahan kualitas air di Sungai Citarum dan bermanfaat dalam memberikan informasi mengenai kualitas air Sungai Citarum serta memprediksi atau mengurangi aktivitas yang dapat merusak Sungai Citarum.