

BAB 1

ANALISIS KEBUTUHAN

1.1 Latar Belakang Masalah

Air merupakan salah satu kebutuhan manusia yang sangat penting. Kebutuhan air dapat dilihat dari segala aspek kehidupan manusia salah satunya digunakan untuk minum. Kualitas air yang digunakan untuk dikonsumsi, merupakan hal yang penting mengingat tubuh manusia memiliki kadar air sebesar 68% [1]. Air yang digunakan untuk memenuhi kebutuhan manusia disebut air baku, yang dapat diperoleh dari berbagai sumber seperti air tanah, sungai, dan air hujan. Namun, tidak semua sumber air baku dapat langsung digunakan karena mudah terkontaminasi dan berpotensi membahayakan kesehatan manusia baik dalam jangka pendek maupun panjang. Oleh karena itu, sebelum air baku dapat dikonsumsi, perlu dilakukan proses pengolahan yang tepat untuk memastikan keamanannya. WHO telah menetapkan standar kualitas air untuk memastikan air yang aman dikonsumsi. Standar ini mencakup bahwa air yang digunakan untuk kebutuhan manusia harus bersifat tidak berbau, tidak berwarna, dan tidak berasa. Selain itu, Permenkes juga telah menetapkan batas aman untuk berbagai parameter kandungan dalam air.

Penentuan kelayakan air minum, umumnya dilakukan pengujian berdasarkan kandungannya. Dua metode yang sering digunakan adalah metode STORET dan Indeks Pencemaran. Namun, kedua metode ini memiliki kelemahan dalam hal efisiensi waktu, biaya, dan tenaga [2]. *Machine learning* dipilih sebagai pendekatan untuk mengklasifikasikan air minum karena dapat meningkatkan efisiensi waktu, mengurangi biaya, dan meminimalkan keterlibatan tenaga manusia [3].

Machine learning adalah cabang ilmu kecerdasan buatan yang melibatkan proses pelatihan komputer atau mesin menggunakan data, dengan tujuan mencapai tingkat kecerdasan yang serupa dengan manusia, termasuk kemampuan yang akurat dalam melakukan prediksi. Pemanfaatan *machine learning* dalam berbagai bidang telah terbukti efektif, terutama dalam melakukan klasifikasi dengan cepat dan akurat [1][4]. Dalam konteks evaluasi kualitas air, penggunaan *machine learning* menjadi solusi yang efisien untuk mengklasifikasikan air minum, karena dapat menghemat waktu, biaya, dan tenaga dengan tingkat akurasi yang tinggi. Dalam rangka mengatasi permasalahan tersebut, *Capstone Project* ini bertujuan untuk mengembangkan sistem klasifikasi air minum berbasis *website*. Dengan memanfaatkan *machine learning*, hasil klasifikasi dapat divisualisasikan melalui *website* yang mudah digunakan oleh pengguna.

1.2 Informasi Pendukung

Tabel 1.1 Informasi Pendukung

No	Informasi Pendukung
1	Studi yang dilakukan oleh Su dan Shen (2020) menyimpulkan bahwa penggunaan <i>machine learning</i> dalam evaluasi kualitas air memiliki potensi yang besar. Teknik <i>machine learning</i> dapat digunakan untuk mengembangkan model prediksi yang akurat dan efisien dalam mengevaluasi kualitas air, termasuk identifikasi kontaminan, prediksi pencemaran, dan deteksi anomali. Metode-metode <i>machine learning</i> seperti pengklasifikasi, regresi, dan jaringan saraf tiruan telah terbukti berhasil dalam menganalisis data kualitas air dan memberikan hasil yang kompetitif dibandingkan dengan metode tradisional. Selain itu, <i>machine learning</i> juga memungkinkan penggabungan data dari berbagai sumber seperti sensor, pengukuran lapangan, dan <i>database</i> historis, yang dapat meningkatkan ketepatan dan keakuratan evaluasi kualitas air. Secara keseluruhan, penggunaan <i>machine learning</i> dalam evaluasi kualitas air menjanjikan untuk memberikan kontribusi signifikan dalam pemantauan dan pengelolaan sumber daya air[5].
2	<i>Machine learning</i> dapat digunakan dalam penilaian kualitas air minum untuk mengembangkan model prediksi efektif yang menggunakan data multivariat kompleks dalam mengevaluasi kualitas air minum[5].
3	<i>Machine learning</i> digunakan secara luas dalam penilaian kualitas air untuk memprediksi berbagai parameter seperti pH, kekeruhan, kandungan bahan organik, dan kontaminan lainnya. Penggunaan metode ini menghasilkan model yang akurat dan dapat diandalkan dalam menilai kualitas air[5].

1.3 Constraint

1.3.1 Aspek Ekonomi

Proses penentuan kelayakan air minum berdasarkan sampel air yang sudah dilakukan pengujian di laboratorium PDAM atau perusahaan air minum yang dilakukan secara manual atau pengolahan data secara tradisional mungkin memakan waktu dan sumber daya yang besar. Dengan menggunakan algoritma *machine learning*, proses klasifikasi dapat dilakukan secara cepat, karena algoritma dapat otomatis menganalisis data pengujian yang sudah dilakukan sebelumnya oleh laboratorium maupun perusahaan air minum. Hal ini menyebabkan pihak

laboratorium PDAM maupun perusahaan air minum dapat meningkatkan efisiensi dalam proses tersebut. Selain itu, metode analisis secara manual membutuhkan sumber daya manusia yang relevan dengan bidang tersebut (berkualitas), peralatan laboratorium yang canggih dan bahan kimia yang mahal. Penggunaan *machine learning* dapat membantu membuat beberapa tahapan diotomatisasi, sehingga dapat mengurangi sumber daya manusia dan biaya operasional yang tinggi, terutama bagi laboratorium PDAM atau perusahaan air minum yang melakukan pengujian dalam skala besar.

1.3.2 Aspek Manufakturabilitas

Pengujian kelayakan air minum yang dilakukan oleh laboratorium pada umumnya memiliki kapasitas berbeda yang dipengaruhi beberapa hal seperti tenaga kerja dan peralatan yang memadai. Tidak semua laboratorium memiliki tenaga kerja yang berpengalaman dalam mengoperasikan dan menginterpretasikan hasil pengujian air minum, karena dibutuhkannya keahlian khusus. Keterbatasan ini mempengaruhi kapasitas laboratorium dalam menguji sampel air terutama dalam sampel yang besar secara bersamaan. Tidak hanya itu, keterbatasan jumlah alat dan beberapa alat memerlukan waktu yang cukup lama untuk mendapatkan hasil yang akurat. Kedua hal ini tentunya sangat berpengaruh terutama efisiensi waktu saat pengujian air minum.

1.3.3 Aspek Keberlanjutan (*Sustainability*)

Pada pengujian beberapa parameter dibutuhkan bahan kimia yang mungkin mahal dan langka. Penggunaan ini tentunya menyebabkan laboratorium menjadi bergantung terhadap bahan tersebut, di sisi lain penggunaan bahan ini menyebabkan adanya limbah bahan kimia yang berdampak serius bagi lingkungan jika tidak diolah dengan benar sesuai langkah-langkah yang sudah ditetapkan.

1.3.4 Aspek Kesehatan dan Keselamatan

Penentuan kelayakan air minum sangat penting dalam menjaga kesehatan dan keselamatan manusia atau makhluk hidup. Algoritma *machine learning* mampu menganalisis data pengujian air minum dengan tingkat akurasi yang tinggi. Dengan menggunakan metode ini, algoritma dapat mengidentifikasi dan mengklasifikasikan kontaminan yang mungkin ada dalam air minum dengan lebih akurat. Hal ini memungkinkan untuk mendeteksi kontaminan yang berpotensi membahayakan kesehatan manusia dengan lebih efektif.

1.3.5 Aspek Sosial

Keberadaan sistem klasifikasi yang mampu mengidentifikasi kontaminan dengan tingkat akurasi yang tinggi dapat memberikan manfaat sosial yang signifikan bagi masyarakat.

Dengan adanya sistem klasifikasi yang akurat, konsumen akan dapat dengan mudah mengakses air minum yang aman dan berkualitas. Hal ini memiliki dampak langsung terhadap kualitas hidup dan kesejahteraan masyarakat. Air minum yang bebas dari kontaminan akan membantu menjaga kesehatan masyarakat secara keseluruhan, terutama bagi kelompok rentan seperti anak-anak, lansia, dan individu dengan sistem kekebalan yang lemah. Selain itu, penentuan kelayakan air minum yang akurat juga dapat berkontribusi pada pemenuhan hak asasi manusia terkait air bersih. Pasokan air minum yang aman dan berkualitas adalah hak dasar setiap individu, dan dengan menggunakan sistem klasifikasi yang efektif, hak tersebut dapat terpenuhi dengan lebih baik. Masyarakat akan merasa lebih percaya dan nyaman dalam mengonsumsi air minum yang telah melalui proses pengujian dan mendapatkan label kelayakan.

1.4 Kebutuhan yang Harus Dipenuhi

Berdasarkan masalah, latar belakang, dan analisis yang telah dipaparkan, maka kebutuhan yang harus dipenuhi dari solusi yang akan diajukan antara lain:

- 1) kemampuan *machine learning* dan algoritma yang digunakan yaitu *Decision Tree*, *K-Nearest Neighbors*, dan *Extreme Learning Machine* untuk menganalisis hasil pengujian laboratorium PDAM maupun perusahaan air minum berupa kandungan air minum secara otomatis, sehingga proses penentuan kelayakan air minum dapat dilakukan dengan cepat dan efisien,
- 2) diperlukan teknik pengolahan data yang tepat dalam tahap preprocessing untuk menangani ketidakseimbangan data yang diperoleh dari laboratorium PDAM atau perusahaan air minum. Salah satu teknik yang efektif untuk mengatasi hal ini adalah menggunakan metode SMOTE (*Synthetic Minority Over-sampling Technique*). Dengan menerapkan teknik SMOTE, data pada kelas target menjadi seimbang, sehingga dapat mengoptimalkan kinerja algoritma klasifikasi seperti *Decision Tree*, *K-Nearest Neighbors*, dan *Extreme Learning Machine*. Dengan menggunakan teknik pengolahan data yang tepat, dapat meningkatkan akurasi dan efektivitas model dalam mengklasifikasikan air minum,
- 3) diperlukan adanya implementasi model yang sudah dibuat ke dalam sebuah *website*. Hal ini bertujuan agar *machine learning* yang telah dibuat sebelumnya dapat diakses dan digunakan dengan mudah oleh pengguna. Pengguna dapat mengakses *website* tersebut, untuk mendapatkan informasi yang akurat mengenai kelayakan air minum sesuai data yang telah dimasukkan oleh pengguna. Pengguna dapat mengunggah data pengujian air minum ke dalam *website* dan menerima hasil analisis secara cepat dan efisien, dan

4) pengembangan sistem klasifikasi berbasis *website* ini dirancang memiliki berbagai macam halaman, diantaranya Beranda, Informasi dan Credits. Halaman Beranda memungkinkan pengguna untuk melakukan uji kelayakan air minum menggunakan dua opsi, yaitu secara manual atau dengan mengunggah *file*. Pengguna juga memiliki fleksibilitas dalam memilih algoritma yang akan digunakan, seperti *Decision Tree*, *K-Nearest Neighbors*, atau *Extreme Learning Machine*. Halaman Informasi menyediakan penjelasan yang mudah dipahami tentang topik-topik terkait. Informasi ini mencakup aspek kesehatan terkait air minum, penjelasan tentang algoritma klasifikasi, algoritma *boosting*, serta sumber *dataset* yang digunakan dalam pembuatan model pada *machine learning*. Sedangkan halaman Credits pada *website* ini menyajikan profil singkat dari tim *Capstone Project* serta dosen pembimbing yang berperan penting dalam pengembangan dan terwujudnya *website* ini.

1.5 Tujuan

Berdasarkan kebutuhan yang harus dipenuhi, tujuan yang ingin dicapai untuk penentuan solusi yang akan diusulkan adalah sebuah solusi yang memadukan teknologi *machine learning* dan algoritma untuk meningkatkan efisiensi dan kecepatan dalam menentukan kelayakan air minum. Dengan implementasi dalam *machine learning*, semua proses akan diotomatisasi sehingga lebih cepat dan efisien. Model tersebut akan diimplementasikan ke dalam sebuah *website* yang akan memungkinkan pengguna untuk melakukan proses klasifikasi kelayakan air minum secara interaktif dan menerima hasil analisis yang akurat dengan cepat dan efisien.

1.6 Solusi Sistem yang Diusulkan

1.6.1 Karakteristik Produk

Produk yang dihasilkan yaitu mengintegrasikan teknologi *machine learning* dan algoritma yang dapat mengklasifikasikan kelayakan air minum. Teknologi *machine learning* dan algoritma tersebut diimplementasikan melalui *website*. Hal ini membuat pengguna baik itu laboratorium di PDAM, perusahaan air minum maupun masyarakat untuk melakukan analisis secara otomatis dan mendapatkan hasil yang akurat dalam waktu yang relatif singkat.

a. Fitur Utama

- 1) Kemampuan *machine learning* dan model yang telah dibuat untuk mengklasifikasikan kelayakan air minum secara tepat dan akurat. Sistem ini dapat melakukan analisis dengan cepat dan efisien, sehingga pengguna dapat memperoleh hasilnya dalam waktu yang singkat.

2) Pengguna diberikan fleksibilitas dalam mengunggah *dataset* pengujian air minum ke dalam *website*, dengan dua opsi yang tersedia yaitu secara manual atau melalui *upload file*. Model yang sudah dibuat pada *machine learning* dan diimplementasikan ke dalam *website* akan langsung memproses data tersebut dan memberikan informasi mengenai kelayakan air minum dengan cepat dan efisien.

b. Sifat Solusi yang diusulkan

- 1) *Website* dapat diakses oleh berbagai pihak tanpa memerlukan biaya, selama pengguna memiliki akses internet.
- 2) Tampilan *website* dirancang dengan tampilan sederhana agar pengguna dari berbagai kalangan dapat menggunakannya dengan mudah.
- 3) *Website* yang dirancang dapat dengan mudah dikembangkan oleh *developer*. *Developer* dapat dengan mudah memperluas dan memodifikasi fitur-fitur yang ada, serta mengintegrasikan fungsi baru sesuai kebutuhan. Kemudahan ini memungkinkan pengembang untuk mengoptimalkan *website* dan memenuhi tuntutan yang berubah dari waktu ke waktu.

1.6.1.1 Solusi 1 ("Penggunaan Algoritma Klasifikasi serta Implementasi Model *Machine Learning* dalam *Website*")

Machine Learning memiliki beberapa algoritma yang digunakan untuk melakukan klasifikasi. Algoritma klasifikasi yang digunakan pada *Capstone Project* kali ini yaitu *Decision Tree*, *K-Nearest Neighbors*, dan *Extreme Machine Learning*.

1) *Decision Tree*

Algoritma *Decision Tree* bersifat sangat kuat, populer, berbasis logika, dan mudah dipahami. Hal yang menarik dari *Decision Tree* adalah penggunaan struktur pohon (*tree*) yang berfungsi untuk merepresentasikan aturan yang terbentuk dari hasil klasifikasi. Dalam *tree*, atribut direpresentasikan oleh sebuah *node*, dan kelas direpresentasikan oleh daun (*leaf*). Setiap pohon memiliki akar (*root*) yaitu *node* yang berada di paling atas[6]. Metode ini diimplementasikan pada saat pengujian *dataset* yang sudah tersedia dan akan diuji keakuratannya. Kelebihan dari metode ini dapat menghilangkan perhitungan yang tidak diperlukan. Karena pada metode *Decision Tree* data akan menguji atau memilih berdasarkan kelas tertentu. Selain itu, metode ini mudah dipahami walaupun bukan kalangan yang tidak memiliki kemampuan khusus dibidang statistik karena sebuah item akan dikelompokkan menggunakan sebuah pohon keputusan.

Stakeholder yang terlibat:

Tabel 1.2 Stakeholder pada Algoritma Decision Tree

No	Stakeholder
1.	Kelompok Tugas Akhir <i>Capstone</i> sebagai pelaksana proyek
2.	Dinas Kesehatan Banda Aceh
3.	Dinas Kesehatan Provinsi Sulawesi Utara
4.	Dinas Kesehatan Kota Bogor
5.	Dinas Kesehatan Provinsi Bandar Lampung
6.	Dinas Kesehatan Kota Bekasi
7.	Dinas Kesehatan Kabupaten Bekasi
8.	Balai Laboratorium Kesehatan Provinsi Sumatera Barat
9.	Dinas Kesehatan Pemerintah Provinsi Kalimantan Barat
10.	Dinas Kesehatan Pemerintah Kota Balikpapan
11.	Dinas Kesehatan Kabupaten Kapuas Kalimantan Tengah
12.	Balai Laboratorium Kesehatan Jayapura Papua
13.	Institut Teknologi Bandung Kabupaten Malang
14.	Dinas Kesehatan Pemerintah Kabupaten Blora Jawa Timur
15.	Dinas Kesehatan Pemerintah Kabupaten Malang Jawa Timur
16.	UPT Laboratorium Kesehatan dan Lingkungan Riau
17.	Laboratorium Mikrobiologi Departemen Teknik Lingkungan Bandung
18.	Poltek Kemenkes Kupang - Jurusan Kesehatan Lingkungan Kabupaten Sabu
19.	UPT Laboratorium Kesehatan Kota Bandung
20.	BBTKLLP Surabaya Kabupaten Malang
21.	BBTKLLP Surabaya Kabupaten Sidoarjo
22.	Unit Laboratorium Jasa Pengujian, Kalibrasi dan Sertifikasi Institut Pertanian Bogor
23.	PT Coway Internasional Indonesia
24.	PDAM Surya Sembada
25.	PDAM Titrawening

2) *K-Nearest Neighbors*

K-Nearest Neighbor merupakan salah satu metode klasifikasi dalam *data mining* yang termasuk ke dalam *supervised learning*. Pengklasifikasian yang dilakukan berdasarkan

atribut dan data *training*, sehingga proses pengklasifikasian data baru dilakukan berdasarkan perbandingan kemiripan mayoritas pada data *training*. Dalam K-NN nilai jarak ditentukan dengan pengujian data *testing* terhadap data *training* kemudian menggunakan nilai terkecil dari nilai ketetanggan terdekat[6]. Metode ini diimplementasikan pada saat pengujian *dataset* yang sudah tersedia dan akan diuji keakuratannya. Kelebihan dari metode ini adalah tangguh terhadap *noise*.

Stakeholder yang terlibat:

Tabel 1.3 Stakeholder pada Algoritma K-Nearest Neighbors

No	Stakeholder
1.	Kelompok Tugas Akhir <i>Capstone</i> sebagai pelaksana proyek
2.	Dinas Kesehatan Banda Aceh
3.	Dinas Kesehatan Provinsi Sulawesi Utara
4.	Dinas Kesehatan Kota Bogor
5.	Dinas Kesehatan Provinsi Bandar Lampung
6.	Dinas Kesehatan Kota Bekasi
7.	Dinas Kesehatan Kabupaten Bekasi
8.	Balai Laboratorium Kesehatan Provinsi Sumatera Barat
9.	Dinas Kesehatan Pemerintah Provinsi Kalimantan Barat
10.	Dinas Kesehatan Pemerintah Kota Balikpapan
11.	Dinas Kesehatan Kabupaten Kapuas Kalimantan Tengah
12.	Balai Laboratorium Kesehatan Jayapura Papua
13.	Institut Teknologi Bandung Kabupaten Malang
14.	Dinas Kesehatan Pemerintah Kabupaten Blora Jawa Timur
15.	Dinas Kesehatan Pemerintah Kabupaten Malang Jawa Timur
16.	UPT Laboratorium Kesehatan dan Lingkungan Riau
17.	Laboratorium Mikrobiologi Departemen Teknik Lingkungan Bandung
18.	Poltek Kemenkes Kupang - Jurusan Kesehatan Lingkungan Kabupaten Sabu
19.	UPT Laboratorium Kesehatan Kota Bandung
20.	BBTKLLP Surabaya Kabupaten Malang
21.	BBTKLLP Surabaya Kabupaten Sidoarjo
22.	Unit Laboratorium Jasa Pengujian, Kalibrasi dan Sertifikasi Institut Pertanian Bogor
23.	PT Coway Internasional Indonesia

24.	PDAM Surya Sembada
25.	PDAM Titrawening

3) *Extreme Learning Machine*

ELM adalah metode yang bekerja dengan konsep SLFNs, metode ini diciptakan untuk mengatasi kelemahan metode jaringan syaraf tiruan *feedforward* terlebih pada proses *learning speed*. ELM baik digunakan untuk mengatasi masalah peramalan dengan waktu pelatihan cepat, mudah untuk diaplikasikan pada masalah kompleks dan diterapkan pada kehidupan nyata[7].

Stakeholder yang terlibat:

Tabel 1.4 Stakeholder pada Algoritma Extreme Learning Machine

No	<i>Stakeholder</i>
1.	Kelompok Tugas Akhir <i>Capstone</i> sebagai pelaksana proyek
2.	Dinas Kesehatan Banda Aceh
3.	Dinas Kesehatan Provinsi Sulawesi Utara
4.	Dinas Kesehatan Kota Bogor
5.	Dinas Kesehatan Provinsi Bandar Lampung
6.	Dinas Kesehatan Kota Bekasi
7.	Dinas Kesehatan Kabupaten Bekasi
8.	Balai Laboratorium Kesehatan Provinsi Sumatera Barat
9.	Dinas Kesehatan Pemerintah Provinsi Kalimantan Barat
10.	Dinas Kesehatan Pemerintah Kota Balikpapan
11.	Dinas Kesehatan Kabupaten Kapuas Kalimantan Tengah
12.	Balai Laboratorium Kesehatan Jayapura Papua
13.	Institut Teknologi Bandung Kabupaten Malang
14.	Dinas Kesehatan Pemerintah Kabupaten Blora Jawa Timur
15.	Dinas Kesehatan Pemerintah Kabupaten Malang Jawa Timur
16.	UPT Laboratorium Kesehatan dan Lingkungan Riau
17.	Laboratorium Mikrobiologi Departemen Teknik Lingkungan Bandung
18.	Poltek Kemenkes Kupang - Jurusan Kesehatan Lingkungan Kabupaten Sabu
19.	UPT Laboratorium Kesehatan Kota Bandung
20.	BBTKLLP Surabaya Kabupaten Malang
21.	BBTKLLP Surabaya Kabupaten Sidoarjo

22.	Unit Laboratorium Jasa Pengujian, Kalibrasi dan Sertifikasi Institute Pertanian Bogor
23.	PT Coway Internasional Indonesia
24.	PDAM Surya Sembada
25.	PDAM Titrawening

Selain penggunaan *machine learning* dalam pengklasifikasian menggunakan algoritma *Decision Tree*, *K-Nearest Neighbors*, dan *Extreme Learning Machine*, model-model ini juga diimplementasikan melalui sebuah *website* yang interaktif. *Website* ini memungkinkan pengguna untuk memeriksa kelayakan dengan mudah tanpa perlu memiliki pengetahuan mendalam tentang *machine learning*.

4) Aplikasi Berbasis Web

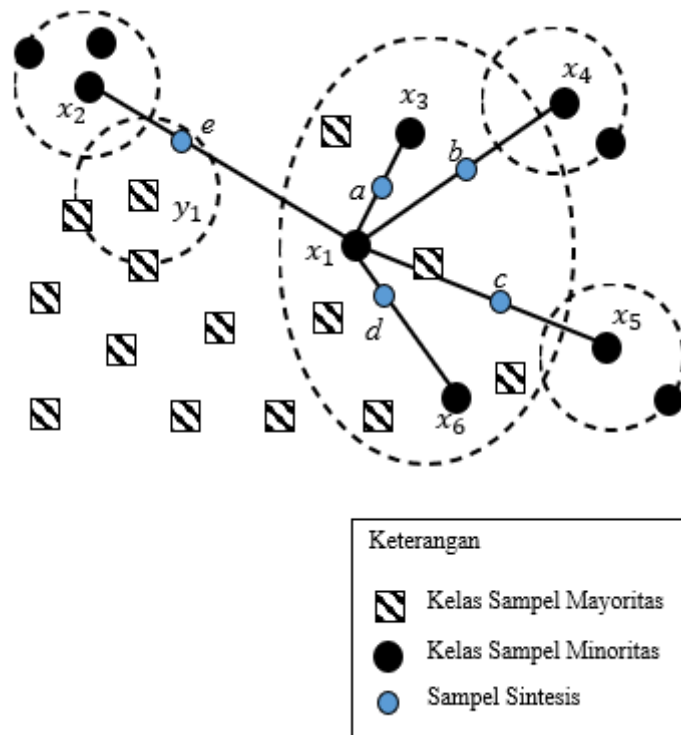
Aplikasi Berbasis Web adalah sebuah aplikasi yang mudah diakses melalui *browser*. Dilihat dari sisi pengguna aplikasi berbasis web ini tidak terlalu banyak memerlukan sumber daya baik dari sisi *software* maupun *hardware*. Pengguna hanya perlu menyediakan internet dan *browser* saja untuk mengaksesnya tanpa harus mengunduh aplikasi tambahan untuk menjalankannya. Dilihat dari sisi *developer*, web ini dibangun menggunakan *framework* yaitu Streamlit. Streamlit merupakan *framework open source* berbasis Python yang bertujuan untuk membangun aplikasi web pada bidang data sains dan *machine learning*.

1.6.1.2 Solusi 2 ("Mengoptimalkan Kinerja Algoritma Klasifikasi dengan Teknik SMOTE untuk Mengatasi Ketidakseimbangan Data, dan Implementasi Model *Machine Learning* dalam *Website*")

Kehandalan dan kinerja model *machine learning* sangat tergantung pada ketersediaan sampel (*dataset*) yang mencakup air yang layak minum maupun tidak dalam jumlah yang cukup dan seimbang. Terdapat masalah ketidakseimbangan data dalam klasifikasi kualitas air, di mana kelas air yang tidak layak minum mendominasi *dataset*. Hal ini dapat menyebabkan model *machine learning* cenderung memprediksi kelas mayoritas dan menghasilkan bias[8]. Untuk mengatasi masalah ini, alternatif yang dapat dilakukan adalah melakukan *resampling*, seperti menggunakan teknik SMOTE[9].

SMOTE adalah salah satu turunan dari *oversampling*. SMOTE pertama kali diperkenalkan oleh Nithes V. Chawla. Pendekatan ini bekerja dengan membuat replikasi dari data minoritas. Replikasi tersebut dikenal dengan data sintesis (*syntetic data*). Metode SMOTE bekerja dengan mencari *K-Nearest Neighbors* (yaitu ketetanggaan terdekat data sebanyak *k*)

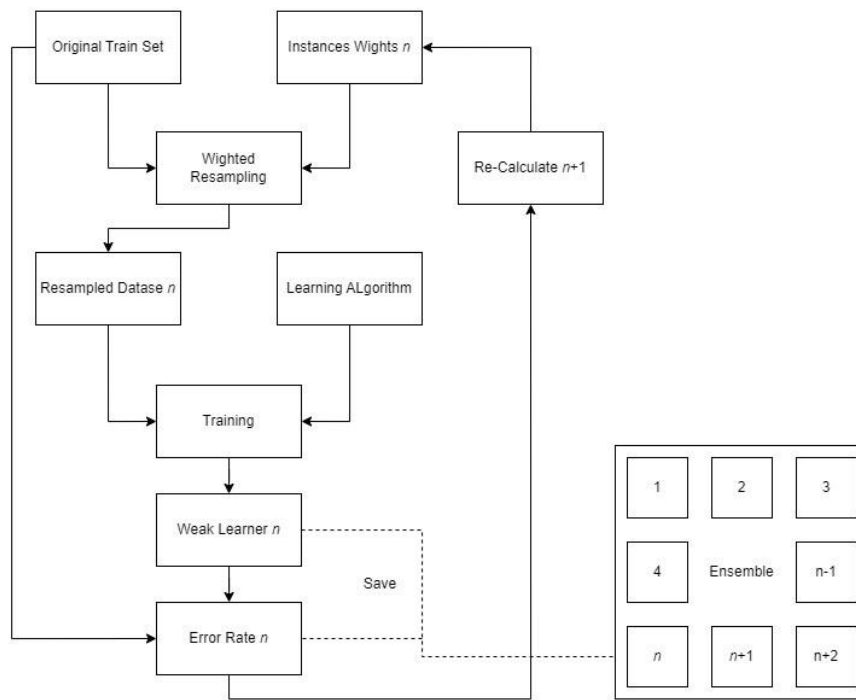
untuk setiap data di kelas minoritas, setelah itu dibuat data sintetis sebanyak persentase duplikasi yang diinginkan antara data *minor* dan *K-Nearest Neighbors* yang dipilih secara acak[10].



Gambar 1.1 SMOTE (*Synthetic Minority Over-sampling Technique*)

1.6.1.3 Solusi 3 (“Peningkatan Akurasi Menggunakan Algoritma AdaBoost”)

Salah satu cara untuk meningkatkan akurasi dalam klasifikasi adalah dengan menggunakan metode penggabungan, seperti algoritma *Adaptive Boosting* (AdaBoost). AdaBoost dapat dengan mudah digabungkan dengan berbagai metode klasifikasi, seperti *Decision Tree*, *K-Nearest Neighbors*, dan *Extreme Learning Machine*, untuk menciptakan model klasifikasi yang lebih kuat. Dengan menggabungkan hasil prediksi dari beberapa model klasifikasi yang lemah, AdaBoost dapat menghasilkan model yang lebih akurat dan dapat diandalkan dalam memprediksi kualitas air. AdaBoost adalah salah satu algoritma *boosting* yang paling populer. Ide dibalik algoritma ini adalah membuat sejumlah *weak learners* yang tidak berkorelasi dan kemudian menggabungkan prediksinya.



Gambar 1.2 Cara Kerja Algoritma AdaBoost

1.6.2 Skenario Penggunaan

1.6.2.1 Solusi 1 ("Penggunaan Algoritma Klasifikasi serta Implementasi Model *Machine Learning* dalam *Website*")

Adapun skenario penggunaan produk dari solusi 1 mencakup langkah-langkah sebagai berikut:

1) Pengumpulan Data

Data kualitas air minum dikumpulkan dari berbagai sumber, seperti pengujian laboratorium atau sumber data lainnya. Data ini berisi parameter-parameter air berupa *E.Colli*, *Coliform*, Arsen, Kromium, Kadmium, Nitrit, Nitrat, Sianida, Selenium, Alumunium, Besi, Kesadahan, Klorida, Mangan, pH, Seng, Sulfat, Tembaga, Amonia, *Chlor*, BOD5, COD, Bau, Rasa, Warna, TDS, Kekeruhan, Suhu dan Potabilitas.

2) Data Preprocessing

Sebelum menerapkan algoritma klasifikasi, data harus diproses dan dibersihkan. Langkah-langkah preprocessing meliputi data *cleaning*, data *transformation* dan lain-lain.

3) Pemodelan menggunakan Algoritma Klasifikasi menggunakan *Decision Tree*, *K-Nearest Neighbors* (KNN), dan *Extreme Learning Machine* (ELM)

Proses ini melibatkan penerapan algoritma tersebut pada data yang telah diproses untuk membuat model klasifikasi.

4) Evaluasi Kinerja Model

Setelah model klasifikasi dibuat, langkah selanjutnya adalah mengevaluasi kinerjanya. Ini melibatkan penggunaan metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-Score* untuk mengukur seberapa baik model dapat mengklasifikasikan kualitas air.

5) Implementasi Model dalam *Website*

Setelah model klasifikasi yang optimal dipilih, model ini diimplementasikan dalam bentuk aplikasi web. Ini berarti model akan diintegrasikan ke dalam *website*, sehingga pengguna dapat mengaksesnya secara *online*.

Skenario ini mengilustrasikan bagaimana pengguna dapat menggunakan *website* yang dibangun dengan mengintegrasikan model *machine learning* berdasarkan algoritma klasifikasi seperti *Decision Tree*, *K-Nearest Neighbors*, dan *Extreme Learning Machine* untuk mengklasifikasikan kualitas air minum. Selain itu, solusi 1 juga berfokus pada aspek implementasi model *machine learning* dalam sebuah *website* agar dapat diakses oleh banyak orang secara *online*.

1.6.2.2 Solusi 2 ("Mengoptimalkan Kinerja Algoritma Klasifikasi dengan Teknik SMOTE untuk Mengatasi Ketidakseimbangan Data dan Implementasi Model *Machine Learning* dalam *Website*")

Adapun skenario penggunaan produk dari solusi 2 mencakup langkah-langkah sebagai berikut.

1) Pengumpulan Data

Data kualitas air minum dikumpulkan dari berbagai sumber, seperti pengujian laboratorium atau sumber data lainnya. Data ini berisi parameter-parameter air berupa *E.Colli*, *Coliform*, Arsen, Kromium, Kadmium, Nitrit, Nitrat, Sianida, Selenium, Aluminium, Besi, Keadahan, Klorida, Mangan, pH, Seng, Sulfat, Tembaga, Amonia, *Chlor*, BOD5, COD, Bau, Rasa, Warna, TDS, Kekeruhan, Suhu dan Potabilitas. Dalam skenario ini, ada masalah ketidakseimbangan data, yaitu kelas Potabilitas 0 merupakan mayoritas yang memiliki jumlah sampel jauh lebih besar daripada kelas minoritas.

2) Data Preprocessing

Sebelum menerapkan algoritma klasifikasi, data harus diproses dan dibersihkan. Langkah-langkah preprocessing meliputi data *cleaning*, data *transformation*, data *reduction* dan data *augmentation* dengan menggunakan teknik *oversampling* SMOTE.

- 3) Pemodelan menggunakan Algoritma Klasifikasi menggunakan *Decision Tree*, *K-Nearest Neighbors* (KNN) dan *Extreme Learning Machine* (ELM)

Proses ini melibatkan penerapan algoritma tersebut pada data yang telah diproses untuk membuat model klasifikasi.

- 4) Evaluasi Kinerja Model

Setelah model klasifikasi dibuat, langkah selanjutnya adalah mengevaluasi kinerjanya. Ini melibatkan penggunaan metrik evaluasi seperti akurasi, presisi, *recall* dan *F1-Score* untuk mengukur seberapa baik model dapat mengklasifikasikan kualitas air.

- 5) Implementasi Model dalam *Website*

Setelah model klasifikasi yang optimal dipilih, model ini diimplementasikan dalam bentuk aplikasi web. Ini berarti model akan diintegrasikan ke dalam *website*, sehingga pengguna dapat mengaksesnya secara *online*.

Skenario ini menyoroti penerapan teknik SMOTE untuk mengatasi ketidakseimbangan data dalam model klasifikasi kualitas air. Implementasi model *machine learning* yang telah dioptimalkan dengan SMOTE dalam *website* bertujuan untuk menyajikan hasil klasifikasi yang lebih andal dan akurat bagi pengguna yang ingin menguji kualitas air.

1.6.2.3 Solusi 3 (“Peningkatan Akurasi Menggunakan Algoritma AdaBoost”)

Adapun skenario penggunaan produk dari solusi 2 mencakup langkah-langkah sebagai berikut.

- 1) Pengumpulan Data

Data kualitas air minum dikumpulkan dari berbagai sumber, seperti pengujian laboratorium atau sumber data lainnya. Data ini berisi parameter-parameter air berupa *E.Colli*, *Coliform*, Arsen, Kromium, Kadmium, Nitrit, Nitrat, Sianida, Selenium, Aluminium, Besi, Kesadahan, Klorida, Mangan, pH, Seng, Sulfat, Tembaga, Amonia, *Chlor*, BOD5, COD, Bau, Rasa, Warna, TDS, Kekeruhan, Suhu dan Potabilitas. Dalam skenario ini, ada masalah ketidakseimbangan data, yaitu kelas Potabilitas 0 merupakan mayoritas yang memiliki jumlah sampel jauh lebih besar daripada kelas minoritas.

- 2) Data Preprocessing

Sebelum menerapkan algoritma klasifikasi, data harus diproses dan dibersihkan. Langkah-langkah preprocessing meliputi data *cleaning*, data *transformation*, data *reduction* dan data *augmentation* dengan menggunakan teknik *oversampling* SMOTE.

- 3) Pemodelan menggunakan Algoritma Klasifikasi menggunakan *Decision Tree*, *K-Nearest Neighbors* (KNN), dan *Extreme Learning Machine* (ELM)

Proses ini melibatkan penerapan algoritma tersebut pada data yang telah diproses untuk membuat model klasifikasi.

4) Evaluasi Kinerja Model

Setelah model klasifikasi dibuat, langkah selanjutnya adalah mengevaluasi kinerjanya. Ini melibatkan penggunaan metrik evaluasi seperti akurasi, presisi, *recall* dan *F1-Score* untuk mengukur seberapa baik model dapat mengklasifikasikan kualitas air. Setelah mengetahui akurasi masing-masing model maka ditambahkan AdaBoost untuk meningkatkan akurasi dari model-model yang ada. Lalu dilakukan evaluasi kembali untuk memastikan bahwa akurasi sudah mengalami peningkatan.

5) Implementasi Model dalam *Website*

Setelah model klasifikasi yang optimal dipilih, model ini diimplementasikan dalam bentuk aplikasi web. Ini berarti model akan diintegrasikan ke dalam *website*, sehingga pengguna dapat mengaksesnya secara *online*.

1.7 Kesimpulan dan Ringkasan CD-1

Proses klasifikasi kualitas air minum menggunakan *machine learning* dengan algoritma *Decision Tree*, *K-Nearest Neighbors*, dan *Extreme Learning Machine* akan dilakukan pada *Capstone Project* ini. Salah satu tantangan yang dihadapi dalam proses ini adalah ketidakseimbangan *dataset*, di mana jumlah sampel pada kelas-kelas tertentu lebih sedikit dibandingkan kelas lainnya. Untuk mengatasi masalah ini, digunakan metode SMOTE yang akan membuat sampel sintetis untuk menyeimbangkan *dataset*. Kelebihan dari SMOTE adalah mampu meningkatkan performa model dengan memperluas variasi data dan mengurangi *bias* pada kelas minoritas.

Setelah proses *machine learning* selesai, model yang telah dibuat diimplementasikan ke dalam sebuah *website*. Hal ini memungkinkan pengguna untuk dengan mudah melakukan klasifikasi kualitas air minum melalui *website*. Dengan mengakses *website* tersebut, pengguna dapat memasukkan data yang ingin diklasifikasikan dan mendapatkan hasil klasifikasi secara cepat. Kemudahan ini memungkinkan pengguna untuk melakukan klasifikasi kualitas air minum secara praktis dan efisien.