



Sistem Prediksi Mutu Air Di Perusahaan Daerah Air Minum Tirta Raharja Menggunakan K – Nearest Neighbors (K – NN)

Rahandanu Rachmat¹, Yulison Herry Chrisnanto², Fajri Rakhmat Umbara³
¹Informatika, Sains dan Informatika, Universitas Jenderal Achmad Yani
rahandanu77@gmail.com

Abstract

PDAM (Perusahaan Daerah Air Minum) Tirta Raharja is the only Regional Business Entity (BUMD) that has the task of providing clean water services to the people of Cimahi City. Clean water is the main requirement that must be consumed by the community and managed in the smooth running of community activities. The development of the city of Cimahi is currently quite fast, with plans to build smart cities, causing the need for clean air as needed. K - Nearest Neighbor (KNN) is a classification algorithm that considers several supporting parameters to carry out a classification process that results in ease of calculation and power. KNN can be considered as one of the most famous non-parametric models. In the research and implementation process of data mining in the regulation of water quality feasibility in PDAM Tirta Raharja using K - the nearest neighbor can be denied as the K - the nearest neighbor implemented in the process of testing the drinking water feasibility in PDAM Tirta Raharja, can be used 93% to be used with the Eligible label Drunk, and 98% for accuracy testing with the label Not Eligible to drink with a K value of 14 where the K value is the most ideal amount that must go through K - Fold Validation Validation of a total of 1,818 data.

Keywords: Information; PDAM Tirta Raharja; Water; Water Quality; K – Nearest Neighbors Algorithm;

Abstrak

PDAM (Perusahaan Daerah Air Minum) Tirta Raharja merupakan satu-satunya Badan Usaha Milik Daerah (BUMD) yang mempunyai tugas memberikan pelayanan air bersih untuk masyarakat Kota Cimahi. Air bersih merupakan kebutuhan pokok yang selalu dikonsumsi masyarakat dan berpengaruh pada kelancaran aktivitas masyarakat. Perkembangan kota Cimahi saat ini cukup pesat, dengan adanya rencana untuk membangun smart city, menyebabkan kebutuhan air bersih secara domestik meningkat. K – Nearest Neighbor (KNN) merupakan sebuah algoritma klasifikasi yang mempertimbangkan beberapa parameter pendukung untuk melakukan proses klasifikasi yang menghasilkan kemudahan waktu perhitungan dan daya. KNN dapat dianggap sebagai salah satu model non-parametrik yang paling terkenal. Dalam penelitian dan proses implementasi data mining pada penentuan kelayakan mutu air pada PDAM Tirta Raharja menggunakan K – Nearest Neighbors dapat disimpulkan bahwa K – Nearest Neighbors yang diimplementasikan pada proses pengujian kelayakan air minum di PDAM Tirta Raharja, menghasilkan akurasi 93% untuk pengujian dengan label Layak Diminum, dan 98% untuk pengujian akurasi dengan label Tidak Layak Diminum dengan nilai K 14 dimana nilai K tersebut adalah jumlah paling ideal yang telah diproses lewat K – Fold Cross Validation dari total data sebanyak 1.818 data.

Kata kunci: Informasi; PDAM Tirta Raharja; Air; Mutu Air; Algoritma K – Nearest Neighbors;

1. Pendahuluan

Air bersih merupakan kebutuhan pokok yang selalu dikonsumsi masyarakat dan berpengaruh pada kelancaran aktivitas masyarakat. Perkembangan kota Cimahi saat ini cukup pesat, dengan adanya rencana untuk membangun smart city, menyebabkan kebutuhan air bersih secara domestik meningkat. Prasarana air bersih merupakan salah satu hal yang penting untuk dikaji mengingat air merupakan kebutuhan pokok yang selalu dikonsumsi oleh masyarakat dan juga berpengaruh besar pada kelancaran aktivitas masyarakat tersebut. Beberapa daerah yang telah dinaungi oleh PDAM telah memiliki kualitas air yang baik, namun sebagian daerah masih memiliki kualitas air yang kurang baik.

PDAM (Perusahaan Daerah Air Minum) Tirta Raharja merupakan satu-satunya Badan Usaha Milik Daerah (BUMD) yang mempunyai tugas memberikan pelayanan air bersih untuk masyarakat Kota Cimahi. Pada saat ini PDAM Tirta Raharja terbagi dalam 5 cabang dan 20 kota pelayanan yang tersebar di Kabupaten Bandung Barat, Kota Cimahi dan Kabupaten Bandung. Saat ini PDAM kota Cimahi mampu melayani kurang lebih 60% dari seluruh penduduk kota Cimahi, sedangkan target pelayanan air bersih untuk skala kota besar adalah 80%.

Proses uji laboratorium yang perlu dilakukan memiliki analisis yang cukup rumit, sehingga tidak sembarang orang dapat menganalisa label data mutu air, hal ini membuat proses uji kelayakan memiliki waktu yang

panjang dalam menentukan mutu air yang cukup bagus untuk didistribusikan kepada masyarakat.

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *Machine Learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakrit dari berbagai *database* besar.

Dalam kaitannya data mining memiliki beberapa teknik salah satunya adalah teknik klasifikasi. Klasifikasi adalah pemrosesan untuk menemukan sebuah model atau fungsi yang menjelaskan dan mencirikan konsep atau kelas data, untuk kepentingan tertentu. Klasifikasi memiliki beberapa Algoritma klasifikasi data mining salah satunya yaitu *K-Nearest Neighbors*[1].

Salah satu teknik dalam data mining adalah klasifikasi, klasifikasi mempunyai sebuah tugas untuk memprediksi kategori label yang tidak diketahui sebelumnya untuk membedakan antara satu objek dengan objek lainnya[2]. Teknik klasifikasi yang paling populer adalah *K - Nearest Neighbor* (KNN), KNN disebut sebagai salah satu dari sepuluh algoritma populer dan paling penting. *K - Nearest Neighbor* (KNN) merupakan sebuah algoritma klasifikasi yang mempertimbangkan beberapa parameter pendukung untuk melakukan proses klasifikasi yang menghasilkan kemudahan waktu perhitungan dan daya[3]. *K - NN* dapat dianggap sebagai salah satu model non-parametrik yang paling terkenal. *K - NN* memberikan metrik pada prediktor untuk menemukan set *K* melewati tetangga terdekat untuk saat ini kondisi di mana tetangga terdekat memiliki yang terendah jarak[4]. Sudah banyak penerapan yang menggunakan KNN seperti penentuan status gunung berapi[5], prediksi penyakit kanker[6], prediksi predikat mahasiswa[7], prediksi kelulusan mahasiswa[8].

Untuk menangani masalah efektifitas analisis uji kelayakan mutu air, maka dibuatlah sistem prediksi untuk mengetahui mutu air yang dibuat menggunakan *K - Nearest Neighbors*.

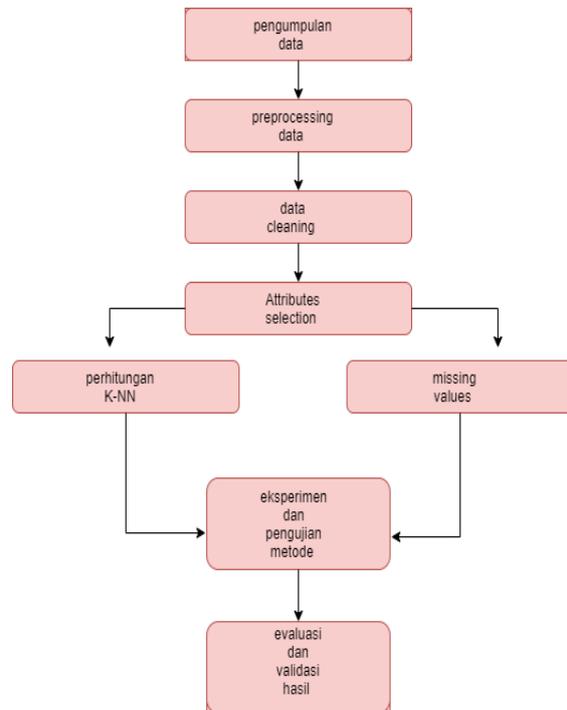
SCADA System merupakan sebuah sistem yang bertujuan untuk mengukur kualitas mutu air yang menjadi sumber air untuk PDAM Tirta Raharja, namun *SCADA System* hanya mampu melakukan perhitungan dengan beberapa parameter saja, sehingga diperlukan proses verifikasi ulang untuk perhitungan kualitas mutu air yang valid. Selain itu beberapa cabang PDAM Tirta Raharja tidak memiliki *SCADA System* untuk proses perhitungan kualitas air, sehingga proses tersebut perlu dilakukan secara manual.

Penelitian ini dimaksudkan untuk membuat sebuah sistem prediksi yang akan melakukan proses klasifikasi mutu air pada data hasil perhitungan *SCADA System* dengan tujuan memberikan informasi yang valid terhadap perhitungan data mutu air.

Penelitian terdahulu didapatkan dari pengujian berdasarkan jumlah data training metode *K - Nearest Neighbors* memiliki rata-rata akurasi keseluruhan sebesar 83.32%[9]

2. Metode Penelitian

Dalam penelitian ini data yang digunakan adalah data mutu air yang diambil dari laboratorium PDAM Tirta Raharja. Data mutu air tersebut akan diolah menggunakan metode KNN sehingga hasilnya dapat digunakan sebagai *rules* dalam memprediksi kualitas mutu air. Dalam penelitian ini akan dilakukan beberapa langkah-langkah atau tahapan seperti pada yang digambarkan pada gambar 2.1



gambar 2. 1. Tahapan penelitian

Untuk mendukung penelitian, maka perlu mengkaji berbagai referensi tentang tugas dari data mining, di mana salah satu tugas dari data mining adalah klasifikasi kasi. Teknik klasifikasi yang umum digunakan adalah *K - NN*[10]. Metode *K - NN* dapat digunakan untuk melakukan pemilihan input untuk model nonlinear dan juga menyediakan perkiraan akurat untuk deret waktu dalam masalah prediksi[14]. Jika *K - NN* digunakan untuk klasifikasi, objek diklasifikasikan dalam kelas paling banyak diminati di antara *k* tetangganya. Jika digunakan untuk regresi, nilai diberikan ke properti objek sebagai rata-rata dari nilai yang diketahui dari *K* tetangga terdekatnya[11].

Jelaskan metode preparasi dan teknik karakterisasi yang digunakan. Jelaskan dengan ringkas, tetapi tetap akurat seperti ukuran, volume, replikasi dan teknik pengerjaan. Untuk metode baru harus dijelaskan secara rinci agar peneliti lain dapat mereproduksi percobaan. Sedangkan metode yang sudah mapan bisa dijelaskan dengan memetik rujukan[4-6].

2.1. Data Mining

Data Mining merupakan proses ekstraksi data menjadi informasi yang sebelumnya belum tersampaikan, dengan teknik yang tepat proses data mining akan memberikan hasil yang optimal[12]. Proses Data Mining yang akan diimplementasikan adalah clustering. Data yang telah dihasilkan akan diolah terlebih dahulu untuk dapat diclustering[13]. Beberapa orang berpendapat bahwa data mining tidak lebih dari machine learning atau statistical analysis yang berjalan pada database. Dalam jurnal ilmiah, data mining juga dikenal sebagai Knowledge Discovery in Databases[14].

2.2. Pengumpulan data

Metode pengumpulan data adalah teknik atau cara yang dilakukan oleh peneliti untuk mengumpulkan data. Pengumpulan data dilakukan untuk memperoleh informasi yang dibutuhkan dalam rangka mencapai tujuan penelitian. Sementara itu instrumen pengumpulan data merupakan alat yang digunakan untuk mengumpulkan data. Pada tahap pengumpulan data akan dilakukan dengan 2 cara, yaitu observasi dan wawancara. Data penelitian diambil data hasil klasifikasi Laboratorium PDAM Tirta Raharja.

2.3. Preprocessing data

Preprocessing Data menerangkan tipe-tipe proses yang melaksanakan data mentah untuk mempersiapkan proses prosedur yang lainnya. Pre Processing yang dilakukan yaitu proses pembersihan data mengalami tiga tahap pembersihan yaitu: Data Cleaning, dan Transformasi. Data yang telah didapatkan, akan dilakukan proses Preprocessing Data guna mengatasi atribut yang tidak memiliki data serta beberapa pembersihan data lainnya yang dibutuhkan untuk proses data mining.

2.4. Classification

Classification atau Klasifikasi adalah teknik di mana data dikategorikan ke dalam kelas dengan memprediksi label kelas atau kategori data[15]. Dalam tulisan ini, klasifikasi dilakukan menggunakan KNN untuk memprediksi kualitas mutu air. Proses klasifikasi adalah dengan menganalisis parameter yang ada dan menyesuaikan nilai parameter dengan standar kebutuhan yang ada dengan tujuan untuk membuat asumsi yang lebih kuat pada label data.

2.5. Supervised Learning

Supervised learning merupakan sebuah pembelajaran yang difokuskan pada data – data yang memiliki label[16]. Supervised learning tidak hanya mempelajari classifier, tetapi juga mempelajari fungsi yang dapat memprediksi suatu nilai numerik[17]. Sebagian besar telah memiliki Di sebagian besar sudah ada yang diawasi algoritma pembelajaran, hanya informasi kelas yang digunakan tetapi informasi kendala dibuang[18].

$$Accuracy = \frac{\text{jumlah nilai benar}}{\text{jumlah data keseluruhan}} \times 100 \% \quad (2)$$

2.6. K – Nearest Neighbors

K – Nearest Neighbors adalah algoritma klasifikasi yang digunakan untuk menentukan kelompok berdasarkan mayoritas pada k tetangga paling dekat. Jika d adalah sekumpulan data training, maka ketika data testing disajikan, algoritma akan menghitung jarak antar bagian data dalam d dengan data testing[10]. Metode KNN dapat digunakan untuk melakukan pemilihan input untuk model nonlinear dan juga menyediakan perkiraan akurat untuk deret waktu dalam masalah prediksi[19]. Jika KNN digunakan untuk klasifikasi, objek diklasifikasikan dalam kelas paling banyak diminati di antara k tetangganya. Jika digunakan untuk regresi, nilai diberikan ke properti objek sebagai rata-rata dari nilai yang diketahui dari k tetangga terdekatnya[11]. Bilangan bulat positif kecil Dalam sebuah penelitian yang membandingkan kinerja antara bayes naif dan K-NN. Menggunakan dataset atribut nominal penelitian ini menghasilkan KNN memiliki kinerja yang lebih baik daripada Naive Bayes. Hasil akurasi naive bayes 87,24% sementara K-NN memiliki akurasi terbaik 90,55% [20]. Pada penelitian Electroencephalography (EEG)-based epileptic seizure prediction using entropy and K-nearest neighbor (KNN) menyebutkan bahwa entropi Shannon dapat membantu diagnosis epilepsi dengan kombinasi dari algoritma K – Nearest Neighbors proses perhitungan prediksi pasien yang menderika kejang menjadi luar biasa cepat daripada menghitung LLE[21].

Pada penelitian Predicting kidney transplantation outcome based on hybrid feature selection and KNN classifier menunjukkan bahwa K – Nearest Neighbors memiliki hasil eksperimen prediksi terhadap transplantasi ginjal yang diusulkan mengungguli metode klasifikasi lainnya dengan skor yang dimiliki adalah 80,77, dan 73,5 untuk akurasi rata-rata dan rata-rata F-Measure masing-masing[22].

Data yang telah dilakukan preprocessing akan menjadi data latih untuk data yang akan diuji dengan mengukur kedekatannya dengan data yang ada, perhitungan *distance* yang akan digunakan adalah *euclidian distance*.

Data yang telah dihitung kedekatannya terhadap data latih, akan dilakukan pemilihan kelas dengan menguji akurasi yang terhadap data latih.

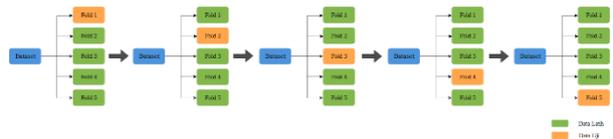
2.7. Data Cleaning

Data yang telah didapatkan dari PDAM Tirta Raharja akan dilakukan proses cleaning dengan maksud untuk membersihkan data dari atribut yang tidak diperlukan serta mengisi atribut data yang tidak memiliki nilai. Terdapat dua proses yang akan dilakukan pada *Data Cleaning* yaitu *Attributes Selection* dan *Missing Values*. *Attributes Selection*, Data yang telah dilakukan proses cleaning, akan dilakukan proses *attributes selection*

untuk memilih atribut yang memiliki tipe *numeric* dengan tujuan agar data dapat dihitung berdasarkan formula $K - NN$. Setelah itu akan dilakukan proses *Missing Values* yaitu data yang telah dilakukan proses *attributes selection*, akan dilakukan proses *missing value* dengan tujuan untuk mengisi data dengan cara mencari nilai modus dari seluruh data dan menginputkan nilai modus tersebut pada atribut data yang kosong.

2.8. *K – Fold Cross Validation*

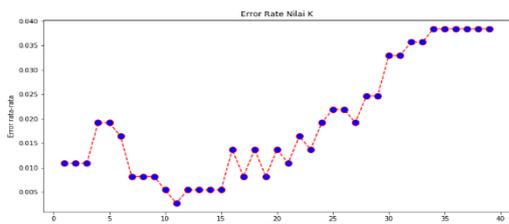
K-Fold adalah salah satu metode Cross Validation yang populer dengan melipat data sebanyak K dan melakukan iterasi terhadap percobaan sebanyak K[23]. Dalam K - Fold Cross-Validation, set data pertama kali dibagi menjadi k partisi yang tidak tumpang tindih. Model diestimasi berdasarkan pada partisi k 1, dan kemudian model dengan parameter yang tidak diketahui diperbaiki pada nilai estimasi sebelumnya yang kemudian akan diterapkan ke partisi yang tersisa[24]. Biasanya *Cross Validation* K - Fold digunakan karena dapat mengurangi waktu komputasi dengan tetap menjaga keakuratan estimasi. Simulasi K – Fold Cross dapat dilihat pada gambar 2.



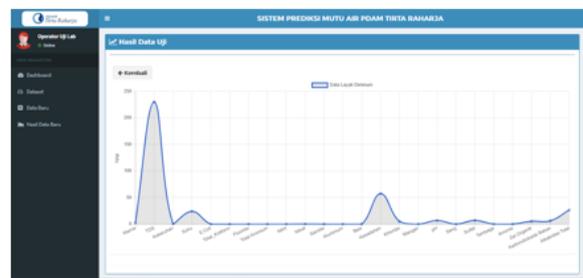
gambar 2. Pembentukan model K-NN

3. Hasil dan Pembahasan

Kebutuhan perangkat lunak pada penelitian pengujian dilakukan menggunakan komputer pribadi. Lingkungan operasi aplikasi yang digunakan adalah lingkungan yang digunakan untuk menerapkan hasil rancangan pada bab analisis dan perancangan. Dalam penelitian ini perangkat lunak berbasis web dengan menggunakan bahasa pemrograman Python, dengan text editor menggunakan *Microsoft Visual Studio Code*, *Basis Data MySql*, Web Server Apache yang ada pada aplikasi *Laragon*, web browser menggunakan Google Chrome dan Firefox serta Framework menggunakan Flask untuk Python. Dalam implementasi program penulis membuat dataset yang akan digunakan untuk proses perhitungan pada data baru. Nantinya data tersebut akan dihitung menggunakan proses K-NN dan dari hasil perhitungan tersebut akan diketahui berapa jumlah K – ideal yang ditampilkan ke dalam bentuk *line charts*.



Gambar 3. Implementasi Program



Gambar 4 implementasi visual hasil data baru

Pada gambar 4 merupakan tampilan yang digunakan untuk mengetahui nilai data hasil prediksi dari masukan data baru, tampilan dari implementasi program tersebut dibuat ke dalam dua tampilan yaitu tampilan yang berbentuk tabel dan tampilan yang berbentuk line chart. Data yang ditampilkan berupa ukuran dari dataset dan status kelayakan dari mutu air tersebut.

Pengujian Akurasi

Pengujian akurasi dimaksudkan untuk menguji kesesuaian label dengan data uji yang diuji terhadap data latih. Pengujian dilakukan dengan menggunakan teknik *K-Fold Cross Validation* dengan mengevaluasi kinerja model atau algoritma dimana data dipisahkan menjadi dua subset yaitu data proses pembelajaran dan data validasi atau evaluasi. Hasil dari pengujian akurasi tersebut dapat dilihat pada tabel 1

Tabel 1. Tabel Pengujian

No	Label	Jumlah Data	Akurasi Pengujian
1.	Layak Diminum	1454	93%
2.	Tidak Layak Diminum	364	98%

Pada pengujian dapat kita lihat sebanyak 1454 data mutu air yang memiliki label layak diminum memiliki akurasi pengujian sebesar 93%, sedangkan sebesar 364 data dengan label tidak layak diminum memiliki akurasi pengujian sebesar 98%.

4. Kesimpulan

Dalam penelitian dan proses implementasi data mining pada penentuan kelayakan mutu air pada PDAM Tirta Raharja menggunakan K – Nearest Neighbors dapat disimpulkan bahwa K – Nearest Neighbors yang diimplementasikan pada proses pengujian kelayakan air minum di PDAM Tirta Raharja, menghasilkan akurasi 93% untuk pengujian dengan label Layak Diminum, dan 98% untuk pengujian akurasi dengan label Tidak Layak Diminum dari total data sebanyak 1.818 data.

Daftar Rujukan

- [1] N. N. P. Galih. Surono, "Journal of technology information," *Http://Jurnal.Kampuswiduri.Ac.Id/*, vol. 5, no. 1, pp. 25–30, 2020.
- [2] Okfalisa, I. Gazalba, Mustakim, and N. G. I. Reza, "Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification," *Proc. - 2017 2nd Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE 2017*, vol. 2018-Janua, pp. 294–298, 2018.
- [3] P. P. D. K. Singh, "Review on Data Mining Techniques for Prediction of Water Quality," *Int. J. Adv. Res. Comput. Sci.*, vol. 8, no. 5, pp. 396–401, 2017.
- [4] D. Dezfooli, S.-M. Hosseini-Moghari, K. Ebrahimi, and S. Araghinejad, "Classification of water quality status based on minimum quality parameters: application of machine learning techniques," *Model. Earth Syst. Environ.*, vol. 4, no. 1, pp. 311–324, 2018.
- [5] F. Tempola, M. Muhammad, and A. Khairan, "Perbandingan Klasifikasi Antara Knn Dan Naive Bayes Pada Penentuan Status Gunung Berapi Dengan K-Fold Cross Validation Comparison of Classification Between Knn and Naive Bayes At the Determination of the Volcanic Status With K-Fold Cross," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, pp. 577–584, 2018.
- [6] T. Praningki and I. Budi, "Sistem Prediksi Penyakit Kanker Serviks Menggunakan CART, Naive Bayes, dan k-NN," *Creat. Inf. Technol. J.*, vol. 4, no. 2, p. 83, 2018.
- [7] G. O. Mustakim, "Algoritma K-Nearest Neighbor Classification," *J. Sains, Teknol. dan Ind.*, vol. 13, no. 2, pp. 195–202, 2016.
- [8] A. Rohman, "Model Algoritma K-Nearest Neighbor (K-Nn) Untuk Prediksi Kelulusan Mahasiswa," *Neo Tek.*, vol. 1, no. 1, 2015.
- [9] M. A. Rahman, N. Hidayat, and A. Afif Supianto, "Komparasi Metode Data Mining K-Nearest Neighbor Dengan Naïve Bayes Untuk Klasifikasi Kualitas Air Bersih (Studi Kasus PDAM Tirta Kencana Kabupaten Jombang)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Vol. 2, No. 12, Desember 2018, hlm. 6346-6353 e-ISSN*, vol. 2, no. 12, pp. 925–928, 2018.
- [10] A. Noviriandini and P. Handayani, "Prediksi Penyakit Liver Dengan Menggunakan Metode Naïve Bayes Dan K- Nearest Neighbour (KNN)," no. November, 2019.
- [11] A. Suárez Sánchez, F. J. Iglesias-Rodríguez, P. Riesgo Fernández, and F. J. de Cos Juez, "Applying the K-nearest neighbor technique to the classification of workers according to their risk of suffering musculoskeletal disorders," *Int. J. Ind. Ergon.*, vol. 52, pp. 92–99, 2014.
- [12] G. Abdillah *et al.*, "Penerapan Data Mining Pemakaian Air Pelanggan Untuk Menentukan Klasifikasi Potensi Pemakaian Air Pelanggan Baru Di Pdam Tirta Raharja Menggunakan Algoritma K-Means," *Sentika 2016*, vol. 2016, no. Sentika, pp. 18–19, 2016.
- [13] A. P. Windarto, "Penerapan Datamining Pada Ekspor Buah-Buahan Menurut Negara Tujuan Menggunakan K-Means Clustering Method," *Techno.Com*, vol. 16, no. 4, pp. 348–357, 2017.
- [14] M. Irfan, W. Uriawan, O. T. Kurahman, M. A. Ramdhani, and I. A. Dahlia, "Comparison of Naive Bayes and K-Nearest Neighbor methods to predict divorce issues," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 434, no. 1, 2018.
- [15] K. Ashwini, J. J. Vedha, and M. D. Priya, "Intelligent model for predicting water quality," *Int. J. Adv. Res. Ideas Innov. Technol. ISSN*, vol. 5, no. 2, pp. 70–75, 2019.
- [16] M. Hardt, E. Price, and N. Srebro, "Equality of opportunity in supervised learning," *Adv. Neural Inf. Process. Syst.*, no. Nips, pp. 3323–3331, 2016.
- [17] T. Dencœur, O. Kanjanatarakul, and S. Sriboonchitta, "A new evidential K-nearest neighbor rule based on contextual discounting with partially supervised learning," *Int. J. Approx. Reason.*, vol. 113, pp. 287–302, 2019.
- [18] P. C. Sen, M. Hajra, and M. Ghosh, *Supervised Classification Algorithms in Machine Learning: A Survey and Review*, vol. 937, 2020.
- [19] J. Sayyad Amin, A. Bahadori, B. Hosseini Nia, S. Rafiee, and N. Kheilnezhad, "Prediction of hydrate equilibrium conditions using k-nearest neighbor algorithm to CO2capture," *Pet. Sci. Technol.*, vol. 35, no. 11, pp. 1070–1077, 2017.
- [20] H. A. Roysid, A. Maulana, and U. Pujiyanto, "Can K-Nearest Neighbor Method Be Used To Predict Success in Indonesia State University Student Selection," *Kursor*, vol. 9, no. 4, pp. 137–144, 2019.
- [21] S. W. Ibrahim, R. Djemal, A. Alsuwailam, and S. Gannouni, "Electroencephalography (EEG)-based epileptic seizure prediction using entropy and K-nearest neighbor (KNN)," *Commun. Sci. Technol.*, vol. 2, no. 1, pp. 6–10, 2017.
- [22] D. M. Atallah, M. Badawy, A. El-Sayed, and M. A. Ghoneim, "Predicting kidney transplantation outcome based on hybrid feature selection and KNN classifier," *Multimed. Tools Appl.*, vol. 78, no. 14, pp. 20383–20407, 2019.
- [23] S. P. Power, O. Prediction, and E. Model, "Short-Term Photovoltaic Power Output Prediction Based on k -Fold Cross-Validation and an Ensemble Model," no. 1, 2019.
- [24] K. J. Grimm, G. L. Mazza, P. Davoudzadeh, K. J. Grimm, G. L. Mazza, and P. Davoudzadeh, "Model Selection in Finite Mixture Models : A k-Fold Cross-Validation Approach Model Selection in Finite Mixture Models : A k -Fold Cross-Validation Approach," *Struct. Equ. Model. A Multidiscip. J.*, vol. 00, no. 00, pp. 1–11, 2016.