

Perbandingan Teknik Pembagian Data untuk Klasifikasi Sarana Akses Air pada Algoritma *K-Nearest Neighbor* dan *Naïve Bayes Classifier*

Said Thaufik Rizaldi¹, Mustakim²

^{1,2}Program Studi Sistem Informasi, Fakultas Sains dan Teknologi
Universitas Islam Negeri Sultan Syarif Kasim Riau

^{1,2}Puzzle Research Data Technology (Predatech), Fakultas Sains dan Teknologi
Universitas Islam Negeri Sultan Syarif Kasim Riau
Jl. HR. Soebrantas Km. 18 Panam Pekanbaru Riau
e-mail: ¹11753101376@students.uin-suska.ac.id, ²mustakim@uin-suska.ac.id

Abstrak

Indonesia menduduki peringkat keempat pertumbuhan penduduk terbesar di dunia. Namun, pertumbuhan penduduk tersebut tidak diikuti dengan pertumbuhan infrastruktur. Sehingga, terjadi fenomena *excess demand* khususnya sarana akses air. Akses air menjadi prioritas pembangunan strategis pada Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2020-2024. Menurut WHO tahun 2019 indikator rekomendasi sarana akses air yakni pelayanan dasar, pelayanan terbatas, dan tidak ada akses. Penelitian ini mengklasifikasikan data sarana akses air menggunakan studi kasus di tingkat sekolah untuk mengetahui indikator kelayakan dalam menentukan rekomendasi akses air. Penelitian ini membandingkan teknik pembagian data *K-Means Clustering*, *K-Medoid* dan *Hold-out* pada algoritma klasifikasi *K-Nearest Neighbor* dan *Naïve Bayes Classifier* menghasilkan *K-Medoid* sebagai pemodelan pembagian data terbaik dengan nilai akurasi 89,39% pada algoritma *K-Nearest Neighbor* dengan parameter $K = 10$ pada penelitian ini.

Kata kunci: Data Mining, Akses Air, Pembagian Data, *K-Nearest Neighbor*, *Naïve Bayes Classifier*.

Abstract

Indonesia is ranked the fourth largest population growth in the world. However, this population growth was not followed by infrastructure growth. There is a phenomenon of *excess demand*, especially water access facilities. Access to water is a strategic development priority in the National Medium-Term Development Plan (RPJMN) 2020-2024. According to WHO, in 2019 the recommended indicators for water access facilities are basic services, limited services, and no access. This study classifies water access facility data using a case study at the school level to determine the feasibility indicator for determining water access recommendations. This study compares the *K-Means Clustering*, *K-Medoid* and *Hold-out* data sharing techniques on the *K-Nearest Neighbor* and *Naïve Bayes Classifier* classification algorithms resulting in *K-Medoid* as the best data sharing modeling with an accuracy value of 89.39% on the *K-Nearest Neighbor* with parameter $K = 10$ in this study.

Keywords: Data Mining, Water Access, Data Sharing, *K-Nearest Neighbor*, *Naïve Bayes Classifier*.

1. Pendahuluan

Air, Sanitasi dan Kebersihan merupakan upaya penting terhadap negara yang memiliki masalah terdapat penghasilan, sejak diluncurkannya program *Water, Sanitation and Hygiene* (WASH), diantaranya komponen akses air menjadi fokus yang prospektif yang mampu menekan masalah kesehatan di daerah terpencil seperti desa [1]. Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2020 – 2024 menekankan pembangunan infrastruktur menjadi fokus yang strategis dalam rangka pertumbuhan dan pemerataan diantaranya bidang pada pelayanan dasar, pemerataan ekonomi dan pembangunan, serta perkotaan dan desa [2]

Air bersih juga sangat dibutuhkan lingkungan sekolah untuk menciptakan suasana yang bersih, sehat, dan aman. Ketersediaan air bersih dan sanitasi sekolah menjadi suatu prioritas pembangunan yang masuk kedalam Tujuan Pembangunan Berkelanjutan (TPB) atau *Sustainable Development Goals* (SDGs). Ketersediaan air bersih dan sanitasi di sekolah dalam SDGs memiliki indikator utama 4a yaitu akses sumber air yang layak, akses jamban yang

berfungsi dan terpisah antara laki-laki dan perempuan, akses sarana cuci tangan, dan akses sabun dan air mengalir [3]

Analisa yang terbitkan oleh Pusat Data Statistik Pendidikan dan Kebudayaan (PDSPK) dalam terbitan Profil Sanitasi Sekolah tahun 2017 adalah jenjang pendidikan yang paling banyak memiliki akses air adalah SLB (76,47%) dan SMK (74,65%). Kemudian jenjang pendidikan yang paling banyak tidak memiliki ketersediaan air bersih adalah SD (31,85%), ini berarti sejumlah 46,986 SD diseluruh Indonesia tidak memiliki ketersediaan air bersih. Saat ini Indonesia masih dari standar yang diamatkan oleh Peraturan Menteri Pendidikan Nomor 24 Tahun 2007 tentang Standar Sarana Prasarana SD/MI, SMP/MTs dan SMA/MA hal ini tentu akan menjadi hambatan dalam proses belajar mengajar, dimana siswa tidak mendapatkan akses air bersih yang tidak merata [3]

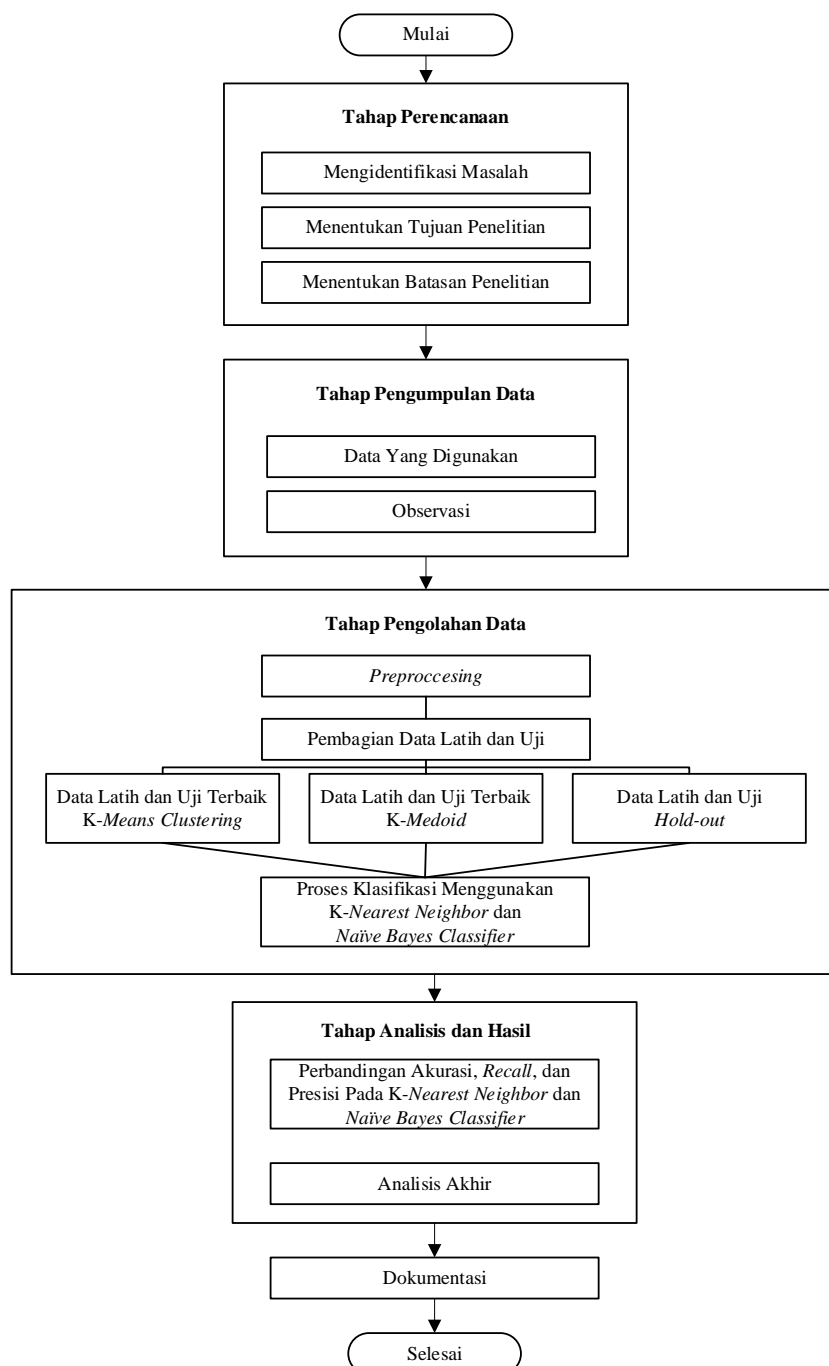
Berdasarkan penjelasan diatas maka perlu adanya klasifikasi berdasarkan rekomendasi kelayakan penentuan rekomendasi dari kelayakan akses air berdasarkan profil sekolah dengan indikator tertentu. Klasifikasi merupakan salah satu teknik dalam data mining yang memiliki kegunaan untuk membangun fungsi dan model klasifikasi pada sebuah objek dalam data yang menyesuaikan dengan karakteristik data tersebut demi memperoleh kelompok masing-masing objek yang sesuai dengan karakteristik kelasnya [4]

Metode Penelitian sebelumnya pernah dilakukan Sabilla dan Putri tahun 2017 menggunakan *K-Nearest Neighbor* dan *Naïve Bayes Classifier* untuk memprediksi waktu lulus mahasiswa dengan *K-Nearest Neighbor* sebagai akurasi tertinggi yang diperoleh menggunakan "cityblok" untuk menghitung 2 buah *neighbor* atau jarak pada penelitian tersebut [5]. Namun, penelitian ini hanya menggunakan Teknik pembagian data *hold-out* yakni 80%:20%, 60:40% dan 50%:50%. Penelitian lain dilakukan oleh Tempola, dkk. tahun 2018 dalam membandingkan metode klasifikasi *K-Nearest Neighbor* dan *Naïve Bayes Classifier* menggunakan data status gunung berapi dimana teknik pembagian data yang digunakan adalah *K-Fold Coss Validation* menghasilkan *Naïve Bayes Classifier* memperoleh akurasi tertinggi dalam penelitian tersebut [6].

Berdasarkan penelitian sebelumnya, penelitian ini menggunakan dua algoritma klasifikasi yakni *K-Nearest Neighbor* dan *Naïve Bayes Classifier* dengan membandingkan teknik pembagian data *K-Means Clustering*, *K-Medoid* dan *Hold-out* (70%:30%). Pada tahun 2017, Mustakim menggunakan *K-Means Clustering* dan *K-Fold Cross Validation* menghasilkan *K-Means Clustering* sebagai teknik pembagian data terbaik untuk memaksimalkan akurasi pada algoritma klasifikasi *K-Nearest Neighbor* [7]. Pada tahun 2019, Mustakim, dkk. Pada algoritma *Naïve Bayes Classifier* melakukan teknik pembagian data menggunakan *Hold-out* (70%:30%), *K-Means Clustering* dan 10 Fold Cross Validation menghasilkan 10 Fold Cross Validation sebagai permodelan pembagian data yang memaksimalkan akurasi pada algoritma tersebut [8]. Penelitian ini menerapkan data Sarana Akses Air untuk mengetahui perbandingan teknik pembagian data terbaik pada algoritma klasifikasi *K-Nearest Neighbor* dan *Naïve Bayes Classifier*.

2. Metode Penelitian

Tahap perencanaan yang digunakan pada penelitian ini adalah data Sanitasi Sekolah berupa komponen Akses Air pada situs web Sarana dan Prasarana di Kementerian Pendidikan dan Kebudayaan (<https://sarpras.dikdasmen.kemdikbud.go.id>) dimana tahap pengambilan data yakni mengambil batasan data Sekolah Menengah Atas (SMA) dan Sekolah Menengah Kejuruan (SMK) di Provinsi Kepulauan Riau dengan jumlah 215 data sesuai dengan kewenangannya. Selanjutnya tahap pengolahan data yakni *preprocessing* dan melakukan pembagian data latih dan data uji dengan menggunakan *K-Means Clustering*, *K-Medoid* dan *Hold-out*. Selanjutnya dilakukan proses klasifikasi menggunakan *K-Nearest Neighbor* (KNN) dan *Naïve Bayes Classifier* (NBC). Pada proses Analisis dan Hasil dilakukan penghitungan Akurasi, *Recall* dan Presisi pada hasil klasifikasi serta melakukan analisis akhir. Terakhir dilakukan dokumentasi pada penelitian. Berikut adalah Metodologi Penelitian yang terdapat pada Gambar 1 di bawah ini.



Gambar 1. Metodologi Penelitian

2.1. Data Mining

Data Mining adalah serangkaian proses untuk menemukan informasi yang belum diketahui sebelumnya dari suatu basis data. Informasi tersebut didapat dengan cara ekstraksi dan pengenalan pola penting atau menarik dari data yang terdapat dalam basis data [9]. Dalam proses penggalian tersebut data mining menerapkan teknik statistika, matematika, *machine learning*, maupun kecerdasan buatan untuk mendapat pengetahuan yang bermanfaat dari berbagai data berukuran besar [7].

2.2. K-Means Clustering

K-Means Clustering merupakan metode analisis kluster yang berfungsi untuk memecah objek menjadi k cluster kemudian posisi setiap objek teramati melalui rata-rata terdekat [10]. Algoritma ini mengelompokkan pengamatan kedalam kluster parameter input (k), kemudian

setiap data ditetapkan pada setiap pengamatan *cluster* berdasarkan kedekatan pengamatan nilai rata-rata *cluster* yang selanjutnya pada proses awal akan dilakukan perhitungan berulang [11]. Persamaan dari kedua titik X1 dan X2 dihitung menggunakan persamaan:

$$D_{L_2}(x_2, x_1) = \|x_2 - x_1\|_2 \quad (1)$$

Sehingga, jarak ruang Euclidean antara dua titik dihitung menggunakan persamaan sebagai berikut [7]:

$$D_{L_2}(x_2, x_1) = \|x_2 - x_1\|_2 = \sqrt{\sum_j^p (x_{2j} - x_{1j})^2} \quad (2)$$

2.3. K-Medoid

Algoritma *K-Medoid* atau dengan nama lainnya *Partitioning Around Medoid* (PAM) menggunakan medoid (objek pada kelompok objek yang mewakili cluster). Kedekatan *medoid* dengan non-*medoid* yang dihitung akan menghasilkan *cluster* [12]. Medoid adalah objek cluster yang paling terpusat, dengan jumlah minimum jarak ke titik lain. Algoritma *k-medoid* juga lebih tahan terhadap *noise* dan *outlier* [13]. Penentuan pusat *cluster* atau *k* terlebih dahulu sehingga pengelompokan objek pada *cluster* dilakukan dengan menggunakan persamaan jarak *Euclidean Distance* menggunakan persamaan :

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - x_{kj})^2} \quad (3)$$

2.4. Hold-Out

Hold-out merupakan metode pemecahan data sederhana yang membagi data menjadi dua bagian berupa data latih dan data uji. *Dataset* yang dipecah telah diidentifikasi label kelasnya. Pada teknik *hold-out*, salah satu bagian digunakan untuk melatih pengklasifikasi (*classifier*) dan bagian lainnya untuk uji pengklasifikasi [14].

2.5. K-Nearest Neighbor (KNN)

K-Nearest Neighbor atau KNN adalah salah satu algoritma dalam data mining yang menerapkan pemilihan nilai yang sesuai untuk *k* (jumlah data yang terdekat dengan suatu objek), sehingga klasifikasi dengan KNN sangat bergantung pada nilai *k* tersebut. Dalam memilih nilai *k*, cara yang sederhana adalah dengan menjalankan algoritma berulang kali sehingga menghasilkan nilai *k* yang berbeda-beda untuk kemudian diambil salah satu nilai *k* dengan kinerja terbaik [15]. Adapun persamaan yang memenuhi *K-Nearest Neighbor* sebagai berikut:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2} \quad (4)$$

Dimana hasil dari $d(x_i, x_j)$ merupakan pengurangan pada masing – masing atribut dikuadratkan serta dijumlahkan berdasarkan nilai yang paling kecil dengan data uji [16].

2.6. Naive Bayes Classifier (NBC)

Naive Bayes Classifier (NBC) didasarkan pada teorema bayes adalah algoritma yang melakukan klasifikasi dengan menggunakan metode statistik dan peluang dalam mengklasifikasikan suatu variabel tertentu dengan persamaan sebagai berikut:

$$eP(R|S) = \frac{P(R) P(S|R)}{P(S)} \quad (5)$$

Dimana R merupakan data yang kelasnya belum diketahui dan S merupakan hipotesis pada data R yang merupakan kelas khusus [17].

2.7. Confusion Matrix

Confusion matrix adalah metode dalam evaluasi yang dapat menentukan kinerja yang didasarkan pada benar dan salah pada sebuah klasifikasi. Dalam confusion matrix terdapat

akurasi, *precision*, dan *recall*. Akurasi menjelaskan tingkat efektifitas algoritma yang digunakan [18]. Precision adalah padanan jumlah data yang berpotensi mendapatkan hasil positif dengan nilai positif pada data asli. Recall adalah padanan jumlah data yang aslinya positif dan dengan menggunakan model algoritma menghasilkan prediksi positif secara benar [19].

3. Hasil dan Pembahasan

3.1. Pengumpulan Data

Data yang digunakan untuk penelitian ini adalah data Sanitasi Sekolah berupa komponen Akses Air pada situs web Sarana dan Prasarana pada Kementerian Pendidikan dan Kebudayaan (<https://sarpras.dikdasmen.kemdikbud.go.id>) dengan mengambil batasan Sekolah Menengah Atas (SMA) dan Sekolah Menengah Kejuruan (SMK) di Provinsi Kepulauan Riau dengan jumlah 215 data.

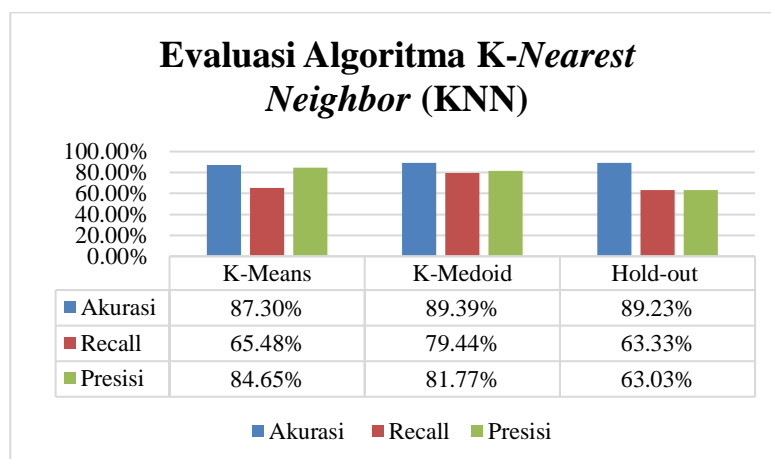
Atribut yang digunakan adalah Status Sekolah (ST), Kabupaten/Kota (KK), Ketersediaan Air di Lingkungan Sekolah (KAL), Sumber Air Sanitasi (SAT), Kecukupan Air (KA), Mayoritas Membawa Air Minum (MAM), Air Minum Untuk Siswa (AMS), Sekolah Memproses Air (SMA), dan Akses Air yang merupakan indikator yang ditetapkan Kementerian Pendidikan dan Kebudayaan. Berikut hasil normalisasi data pada Tabel 1. Sebagai berikut :

Tabel 1. Data Normalisasi

No	ST	KK	KAL	SAT	KA	MAM	AMS	SMA	Akses Air
1	1	7	2	2	1	2	2	2	Dasar
2	1	7	1	10	2	1	1	2	Tidak Ada
3	1	7	1	7	1	2	1	2	Dasar
4	1	7	2	2	1	1	2	2	Dasar
5	1	7	1	7	1	1	1	2	Dasar
6	1	7	2	7	1	2	2	2	Dasar
7	1	7	2	10	1	1	2	2	Terbatas/Kurang
8	2	7	2	10	1	1	2	2	Tidak Ada
9	2	7	2	7	1	1	2	1	Dasar
10	1	7	2	2	1	2	2	2	Terbatas/Kurang
...
213	1	3	2	10	2	2	2	2	Tidak Ada
214	1	3	2	9	1	2	2	2	Dasar
215	1	3	2	2	1	1	2	2	Dasar

3.2. Pembagian Data *K-Nearest Neighbor*

Dalam penelitian ini, akan dilakukan proses klasifikasi menggunakan algoritma *K-Nearest Neighbor* (KNN) menggunakan hasil pembagian data latih dan data uji pada teknik pembagian data *K-Means Clustering*, *K-Medoid*, dan *Hold-out*. Adapun parameter K mengambil nilai K parameter 5, 10, 15 berdasarkan penelitian dahulu oleh Gede pada tahun 2015 [20]. Berikut adalah hasil dari perbandingan evaluasi pada algoritma *K-Nearest Neighbor* (KNN) pada Gambar 2. :

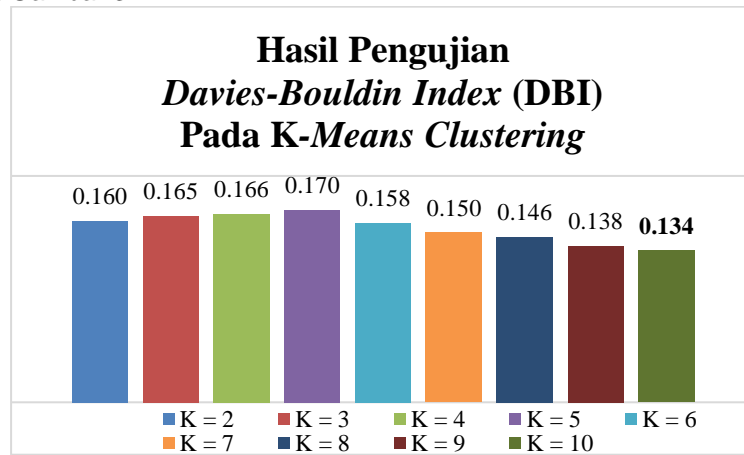


Gambar 2. Perbandingan Evaluasi Algoritma KNN

Evaluasi Algoritma *K-Nearest Neighbor* (KNN) terhadap tiga teknik pembagian data menghasilkan *K-Medoid* sebagai teknik pembagian data terbaik sebagai teknik pembagian data dengan nilai akurasi 89,39%, nilai *recall* sebesar 79,44% dan nilai presisi sebesar 81,77%. Sedangkan teknik pembagian data *K-Means* menduduki terendah dengan nilai akurasi 87,30%, nilai *recall* sebesar 65,48% dan nilai presisi sebesar 84,65%. Sehingga, teknik pembagian data *K-Medoid* dapat digunakan sebagai perbandingan algoritma pada penelitian ini.

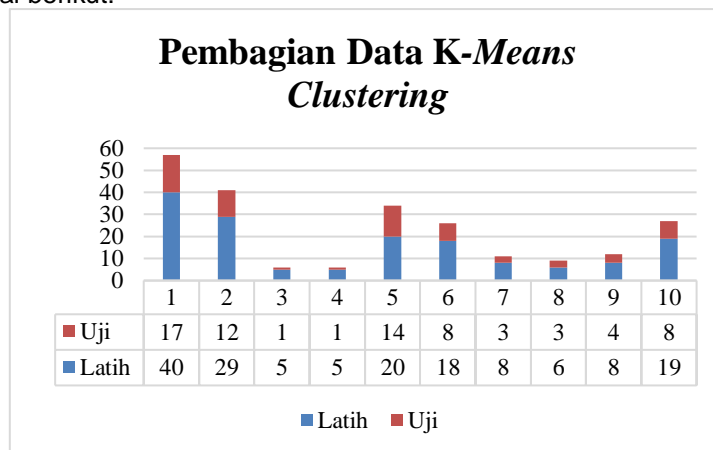
3.3. Pembagian Data *Naïve Bayes Classifier*

Proses klasifikasi dilakukan proses klasifikasi menggunakan algoritma *Naïve Bayes Classifier* dengan menggunakan hasil pembagian data latih dan data uji pada teknik pembagian data *K-Means Clustering*, *K-Medoid*, dan *Hold-out*. Berikut adalah salah satu percobaan pembagian data dengan menggunakan *K-Means Clustering* dimana penentuan nilai *k* terbaik yakni berdasarkan pengujian nilai *Davis Bouldin Index* (DBI) dengan nilai uji terendah sebagai *cluster* terbaik pada Gambar 3.



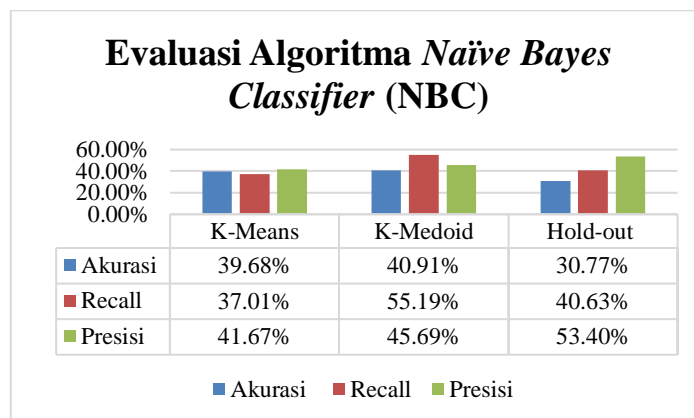
Gambar 3. Hasil Pengujian DBI pada *K-Means Clustering*

Adapun *cluster* 10 atau *K* = 10 terpilih sebagai *cluster* terbaik dengan nilai 0,134 dimana hasil *clustering* tersebut terbagi menjadi 70% data latih dan 30% data uji dengan komposisi pada Gambar 3 sebagai berikut:



Gambar 4. Pembagian Data *K-Means Clustering*

Berikut adalah hasil komposisi pembagian data dengan menggunakan *K-Means Clustering* yang selanjutnya diterapkan juga pada *K-Medoid* dan *Hold-out*. Berikut adalah hasil dari perbandingan evaluasi pada algoritma *Naïve Bayes Classifier* (NBC) pada Gambar 5. :

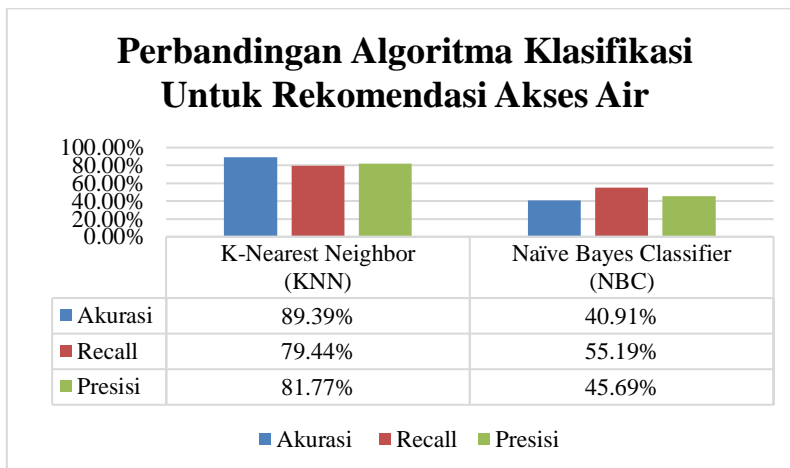


Gambar 6. Perbandingan Evaluasi Algoritma NBC

Evaluasi Algoritma *Naïve Bayes Classifier* (NBC) terhadap tiga teknik pembagian data menghasilkan *K-Medoid* sebagai teknik pembagian data terbaik sebagai teknik pembagian dengan nilai akurasi 40,91%, nilai *recall* sebesar 55,19% dan nilai presisi sebesar 45,69%. Sedangkan teknik pembagian data *Hold-out* menduduki terendah dengan nilai akurasi 30,77%, nilai *recall* sebesar 40,63% dan nilai presisi sebesar 53,40%. Sehingga, teknik pembagian data *K-Medoid* dapat digunakan sebagai perbandingan algoritma pada penelitian ini.

3.4. Perbandingan Algoritma KNN dan NBC

Perbandingan algoritma Data Mining dengan menggunakan metode klasifikasi antara *K-Nearest Neighbor* (KNN) dan *Naïve Bayes Classifier* (NBC) ditujukan pada Gambar 7. sebagai berikut.



Gambar 7. Perbandingan Algoritma KNN dan NBC

Evaluasi Algoritma Klasifikasi untuk menentukan Rekomendasi Akses Air menghasilkan *K-Nearest Neighbor* (KNN) sebagai algoritma terbaik pada penelitian ini, dengan nilai akurasi sebesar 83,39% sedangkan nilai *recall* dan presisi masing-masing 79,44% dan 81,77%, sedangkan *Naïve Bayes Classifier* (NBC) menghasilkan nilai akurasi 40,91% dengan nilai *recall* 55,19% dan Presisi 45,69%. Berdasarkan hasil penelitian ini, teridentifikasi bahwa akurasi KNN memiliki perbedaan secara signifikan terhadap NBC. Beberapa penelitian perbandingan algoritma KNN dan NBC menghasilkan KNN sebagai algoritma dengan akurasi terbaik dibandingkan dengan NBC [21][22][23], dengan jenis *dataset* yang berbeda.

4. Kesimpulan

Berdasarkan hasil penelitian dan pemaparan diatas dapat disimpulkan bahwa perbandingan teknik pembagian data terbaik menghasil *K-Medoid* secara berurutan pada algoritma *K-Nearest Neighbor* (KNN) pada parameter $K=10$ dengan nilai akurasi 89,39%, nilai *recall* sebesar 79,44% dan nilai presisi sebesar 81,77% dan *Naïve Bayes Classifier* dengan nilai akurasi 40,91%, nilai *recall* sebesar 55,19% dan nilai presisi sebesar 45,69%. Berdasarkan hasil penelitian ini maka percobaan yang dilakukan menggunakan data Sarana Akses Air dapat

digunakan *K-Medoid* sebagai teknik pembagian data untuk klasifikasi Sarana Akses air pada algoritma klasifikasi.

Daftar Pustaka

- [1] D. Duijster, B. Monse, J. Dimaisip-Nabuab, P. Djuharnoko, R. Heinrich-Weltzien, M. Hobdell, K. Kromeyer-Hauschild, Y. Kunthearith, M. C. Mijares-Majini, N. Siegmund, P. Soukhanouvong, dan H. Benzian, "Fit for school' - a school-based water, sanitation and hygiene programme to improve child health: Results from a longitudinal study in Cambodia, Indonesia and Lao PDR," *BMC Public Health*, vol. 17, no. 1, hal. 1–15, 2017.
- [2] E. W. Purwanto, "Pembangunan Akses Air Bersih Pasca Krisis Covid-19," *J. Perenc. Pembang. Indones. J. Dev. Plan.*, vol. 4, no. 2, hal. 207–214, 2020.
- [3] Kemendikbud, *Profil Sanitasi Sekolah*. Kemedikbud, 2017.
- [4] Y. Cheng, K. Chen, H. Sun, Y. Zhang, dan F. Tao, "Data and knowledge mining with big data towards smart production," *J. Ind. Inf. Integr.*, vol. 9, no. October, hal. 1–13, 2018.
- [5] W. I. Sabilla dan T. E. Putri, "Prediksi Ketepatan Waktu Lulus Mahasiswa dengan k- Nearest Neighbor dan Naïve Bayes Classifier (Studi Kasus Prodi D3 Sistem Informasi Universitas Airlangga)," *J. Komput. Terap.*, vol. 3, no. 2, hal. 233–240, 2017.
- [6] F. Tempola, M. Muhammad, dan A. Khairan, "Perbandingan Klasifikasi Antara KNN dan Naive Bayes pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, hal. 577, 2018.
- [7] Mustakim, "Effectiveness of K-means clustering to distribute training data and testing data on K-nearest neighbor classification," *J. Theor. Appl. Inf. Technol.*, vol. 95, no. 21, hal. 5693–5700, 2017.
- [8] Mustakim, S. Syahidatul Helma, U. Ramadhani, G. S. Achmad Daengs, R. Novita, Nuryanti, dan S. R. Fitriati, "Data Sharing Technique Modeling for Naive Bayes Classifier for Eligibility Classification of Recipient Students in the Smart Indonesia Program," *J. Phys. Conf. Ser.*, vol. 1424, no. 1, 2019.
- [9] A. P. U. Sembiring dan M. Ginting, "Analisis faktor-faktor yang mempengaruhi pengunduran diri mahasiswa dengan aplikasi data mining add-ins – studi kasus pada STMIK Mikroskil," *JSM STMIK Mikroskil*, vol. 14, no. 2, hal. 139–146, 2013.
- [10] D. K. Sharma, S. K. Dhurandher, D. Agarwal, dan K. Arora, "kROp: k-Means clustering based routing protocol for opportunistic networks," *J. Ambient Intell. Humaniz. Comput.*, vol. 10, no. 4, hal. 1289–1306, 2019.
- [11] I. Kamila, U. Khairunnisa, dan M. Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, hal. 119, 2019.
- [12] D. Listiyanti, Y. A. Syahbana, dan S. R. Henim, "Perancangan dan Implementasi Aplikasi Android Penentu Salient Area pada Video dengan Algoritma K-Medoids," vol. 2, no. 1, hal. 96–101, 2016.
- [13] X. Jin dan J. Han, "K-Medoids Clustering," *Encycl. Mach. Learn. Data Min.*, hal. 697–700, 2017.
- [14] A. Ghazvini, J. Awwalu, dan A. Abu Bakar, "Comparative Analysis of Algorithms in Supervised Classification: A Case study of Bank Notes Dataset," *Int. J. Comput. Trends Technol.*, vol. 17, no. 1, hal. 39–43, 2014.
- [15] G. Guo, H. Wang, D. Bell, Y. Bi, dan K. Greer, "An kNN model-based approach and its application in text categorization," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2945, hal. 559–570, 2004.
- [16] Okfalisa, I. Gazalba, Mustakim, dan 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, no. October 2019, hal. 294–298, 2018.
- [17] S. D. Jadhav dan H. P. Channe, "Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques," *Int. J. Sci. Res.*, vol. 5, no. 1, hal. 1842–1845, 2016.
- [18] M. S. Pervez dan D. M. Farid, "Feature selection and intrusion classification in NSL-KDD cup 99 dataset employing SVMs," *Ski. 2014 - 8th Int. Conf. Software, Knowledge, Inf. Manag. Appl.*, 2014.
- [19] S. Dewi, "Komparasi 5 Metode Algoritma Klasifikasi Data Mining Pada Prediksi Keberhasilan Pemasaran Produk Layanan Perbankan," *None*, vol. 13, no. 1, hal. 60–66, 2016.
- [20] I. Gede, "Texture Analysis on Image Motif of Endek Bali using K-Nearest Neighbor Classification Method," *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 9, hal. 205–211, 2015.
- [21] K. Chandel, V. Kunwar, S. Sabitha, T. Choudhury, dan S. Mukherjee, "A comparative study on thyroid disease detection using K-nearest neighbor and Naive Bayes classification techniques," *CSI Trans. ICT*, vol. 4, no. 2–4, hal. 313–319, 2016.
- [22] Z. E. Rasjid dan R. Setiawan, "Performance Comparison and Optimization of Text Document Classification using k-NN and Naïve Bayes Classification Techniques," *Procedia Comput. Sci.*, vol. 116, hal. 107–112, 2017.
- [23] A. Singh, M. N., dan R. Lakshminathan, "Impact of Different Data Types on Classifier Performance of Random Forest, Naïve Bayes, and K-Nearest Neighbors Algorithms," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 12, hal. 1–11, 2017.