



Clustering and Prediction of the Distribution of Civil Servant Teachers of Papuan and Non-Papuan Origin Using a Data Mining Approach

Pengelompokan dan Prediksi Distribusi Guru ASN Orang Asli Papua dan Non-Orang Asli Papua dengan Pendekatan Data Mining

Usman Arfan^{1*}, Fauzia Fajriyati²

^{1,2}Program Studi Informatika, STMIK Pesat Nabire, Indonesia

E-Mail: ¹usmanarfanpesat@gmail.com, ²fauziafajriyati@gmail.com

Received Apr 29th 2026; Revised Jun 29th 2026; Accepted Jul 20th 2026; Available Online Jul 29th 2026

Corresponding Author: Usman Arfan

Copyright © 2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

The equitable distribution of civil servant teachers (ASN) is a critical factor in ensuring fair educational services, particularly in Nabire Regency, which is characterized by diverse geographical conditions, including urban, coastal, and remote areas. This study aims to analyze the distribution of Indigenous Papuan (OAP) and non-Indigenous Papuan (non-OAP) ASN teachers across educational institutions in Nabire Regency using a data mining approach. The study employed secondary data collected from the Nabire Regency Education Office in 2025, covering 256 educational institutions and 2,147 ASN teachers, including 364 OAP teachers (16.95%) and 1,783 non-OAP teachers (83.05%). Individual teacher data were aggregated at the institutional level using attribute selection, data cleaning, missing value handling, duplicate removal, aggregation, and normalization. K-Means was applied to cluster educational institutions, while Naive Bayes, k-Nearest Neighbors, and Decision Tree were employed to classify distribution categories. The clustering process produced three main clusters. Cluster 1 represented institutions with a high proportion of OAP teachers but a relatively small number of teachers. Cluster 2 was dominated by non-OAP teachers, with a large number of teachers, whereas Cluster 3 comprised 179 educational institutions (69.92%) and reflected the general distribution pattern, which was also dominated by non-OAP teachers. Decision Tree achieved the best performance with an AUC of 0.994, accuracy of 0.992, precision of 0.992, recall of 0.992, and F1-score of 0.992. The findings indicate that the distribution of OAP ASN teachers remains disproportionate, particularly in the "No OAP" and "Low" categories.

Keywords: Civil Servant Teachers, Data Mining, Nabire Regency, Papuan Indigenous People, Teacher Distribution.

Abstrak

Pemerataan distribusi guru ASN menjadi isu penting dalam penyediaan layanan pendidikan yang berkeadilan, terutama di Kabupaten Nabire yang memiliki karakteristik wilayah yang beragam, termasuk perkotaan, pesisir, dan wilayah dengan aksesibilitas terbatas. Penelitian ini bertujuan untuk menganalisis distribusi guru ASN Orang Asli Papua (OAP) dan non-OAP pada satuan pendidikan di Kabupaten Nabire menggunakan pendekatan data mining. Data penelitian berupa data sekunder dari Dinas Pendidikan Kabupaten Nabire tahun 2025 yang mencakup 256 satuan pendidikan dengan total 2.147 guru ASN, terdiri atas 364 guru OAP (16,95%) dan 1.783 guru non-OAP (83,05%). Data ditransformasikan menjadi data agregat per satuan pendidikan melalui seleksi atribut, pembersihan data, penanganan nilai kosong, penghapusan duplikasi, agregasi, dan normalisasi. K-Means digunakan untuk mengelompokkan satuan pendidikan, sedangkan Naive Bayes, k-Nearest Neighbors, dan Decision Tree digunakan untuk mengklasifikasikan kategori distribusi. Hasil clustering menghasilkan tiga cluster utama. Cluster 1 menunjukkan keterwakilan OAP tinggi dengan jumlah guru relatif kecil, Cluster 2 didominasi non-OAP dengan jumlah guru besar, sedangkan Cluster 3 mencakup 179 satuan pendidikan (69,92%) dan menunjukkan dominasi guru non-OAP. Decision Tree memberikan performa terbaik dengan AUC 0,994, accuracy 0,992, precision 0,992, recall 0,992, dan F1-score 0,992. Temuan menunjukkan distribusi guru ASN OAP belum proporsional, terutama pada kategori Tidak ada OAP dan Rendah.

Kata Kunci: Data Mining, Distribusi Guru, Guru ASN, Kabupaten Nabire, Orang Asli Papua.

1. PENDAHULUAN

Pemerataan pendidikan merupakan salah satu isu penting dalam pembangunan sumber daya manusia, khususnya di daerah yang memiliki karakteristik geografis dan sosial yang beragam [1]. Dalam upaya mewujudkan pemerataan tersebut, keberadaan guru tidak hanya dinilai dari jumlahnya, tetapi juga dari bagaimana guru didistribusikan secara proporsional antarwilayah dan antar satuan pendidikan [2]. Sekolah yang memiliki tenaga pendidik cukup dan sesuai kebutuhan cenderung lebih mampu menjalankan proses pembelajaran secara stabil dibandingkan sekolah yang mengalami kekurangan atau ketimpangan penempatan guru [3]. Oleh karena itu, distribusi guru menjadi salah satu indikator penting dalam melihat kualitas layanan pendidikan secara lebih menyeluruh [4].

Dalam konteks Kabupaten Nabire, persoalan distribusi guru tidak hanya berkaitan dengan jumlah dan lokasi penempatan, tetapi juga dapat dilihat dari komposisi latar belakang guru ASN, khususnya antara guru ASN Orang Asli Papua dan non-Orang Asli Papua. Berdasarkan data Dinas Pendidikan Kabupaten Nabire tahun 2025, jumlah guru ASN di Kabupaten Nabire tercatat sebanyak 2.147 orang, terdiri atas 364 guru ASN OAP (16,95%) dan 1.783 guru ASN non-OAP (83,05%). Data ini menunjukkan bahwa secara komposisi, guru ASN non-OAP masih mendominasi struktur tenaga pendidik ASN di Kabupaten Nabire. Oleh karena itu, analisis distribusi guru ASN OAP dan non-OAP menjadi penting untuk melihat apakah komposisi tersebut tersebar secara merata pada satuan pendidikan atau justru terkonsentrasi pada sekolah dan wilayah tertentu [5]. Aspek keterwakilan guru ASN OAP dalam penelitian ini tidak dimaknai sebagai faktor yang secara langsung menentukan motivasi atau hasil belajar siswa, melainkan sebagai indikator representasi sosial dalam pemerataan tenaga pendidik di wilayah Papua. Dalam konteks pendidikan di Papua, kedekatan sosial, budaya, dan pemahaman terhadap karakteristik masyarakat lokal dapat dipandang sebagai salah satu aspek yang relevan dalam penyelenggaraan layanan pendidikan yang lebih kontekstual [6]. Namun demikian, penelitian ini tetap membatasi analisis pada aspek distribusi dan komposisi guru ASN berdasarkan kategori OAP dan non-OAP, bukan pada pengukuran langsung terhadap dampak keberadaan guru OAP terhadap motivasi belajar siswa.

Distribusi guru pada dasarnya merupakan proses penyebaran tenaga pendidik ke sekolah atau wilayah tertentu sesuai kebutuhan pendidikan [7]. Distribusi yang baik tidak hanya berarti seluruh sekolah memiliki guru, tetapi juga menunjukkan adanya keseimbangan penempatan berdasarkan kondisi nyata lapangan [8]. Dalam penelitian penulis, distribusi guru dianalisis berdasarkan latar belakang Orang Asli Papua dan non-Orang Asli Papua untuk melihat pola keterwakilan keduanya pada satuan pendidikan di Kabupaten Nabire. Untuk membaca pola tersebut secara lebih mendalam, penelitian penulis memanfaatkan pendekatan data mining. Pendekatan ini dipilih karena mampu mengolah data dalam jumlah besar dan menemukan pola yang tidak selalu terlihat melalui analisis statistik deskriptif biasa.

Permasalahan utama dalam penelitian ini adalah masih adanya indikasi ketimpangan distribusi guru ASN Orang Asli Papua dan non-Orang Asli Papua pada satuan pendidikan di Kabupaten Nabire. Penelitian ini mencakup satuan pendidikan pada jenjang TK, SD, SMP, SMA/ sederajat, serta PKBM yang tercatat dalam data Dinas Pendidikan Kabupaten Nabire. Dengan cakupan tersebut, analisis tidak hanya diarahkan pada satu jenjang pendidikan tertentu, tetapi pada pola distribusi guru ASN OAP dan non-OAP secara lintas jenjang satuan pendidikan. Hasil pengolahan data menunjukkan bahwa tidak semua satuan pendidikan memiliki komposisi keterwakilan guru ASN OAP yang seimbang. Sebagian satuan pendidikan berada pada kategori tanpa guru ASN OAP, sebagian lainnya berada pada kategori rendah, dan hanya sebagian satuan pendidikan yang menunjukkan keterwakilan sedang hingga tinggi. Kondisi ini menunjukkan bahwa keberadaan guru ASN OAP belum tersebar secara proporsional pada seluruh satuan pendidikan yang dianalisis. Jika distribusi seperti ini tidak dievaluasi secara lebih rinci, maka pemerataan penempatan guru akan sulit dinilai hanya dengan melihat jumlah total guru ASN secara agregat, tanpa memperhatikan komposisi OAP dan non-OAP pada masing-masing sekolah atau satuan pendidikan [9].

Salah satu pendekatan yang dapat digunakan untuk membaca permasalahan tersebut secara lebih sistematis adalah data mining melalui kombinasi clustering dan prediksi [10]. Nilai tambah tahap klasifikasi dalam penelitian ini terletak pada kemampuannya untuk mengubah hasil pemetaan pola menjadi alat bantu pengambilan keputusan. Clustering digunakan untuk menemukan pola kelompok distribusi guru ASN OAP dan non-OAP, sedangkan klasifikasi digunakan untuk memperkirakan kategori distribusi satuan pendidikan berdasarkan pola yang telah terbentuk [11]. Dengan demikian, Dinas Pendidikan tidak hanya memperoleh informasi mengenai kelompok sekolah yang memiliki karakteristik serupa, tetapi juga memperoleh dasar awal untuk mengidentifikasi satuan pendidikan yang perlu diprioritaskan dalam evaluasi dan perencanaan redistribusi guru pada periode berikutnya.

Penelitian Kuwil dan Atila (2024) menunjukkan bahwa pendekatan hibrida melalui integrasi K-Means dan Critical Distance Clustering (CDC-2) mampu meningkatkan kualitas proses pengelompokan data [12]. Namun, penelitian tersebut berfokus pada optimalisasi teknik *clustering* dan belum melanjutkan hasil pengelompokan ke tahap prediksi menggunakan algoritma klasifikasi. Selain itu, konteks penelitian tersebut juga belum menyentuh isu distribusi sumber daya manusia pendidikan, khususnya distribusi guru berdasarkan latar belakang sosial atau kewilayahan.

Dalam konteks Papua, isu distribusi guru ASN Orang Asli Papua dan non-Orang Asli Papua memiliki relevansi kebijakan yang kuat dengan pelaksanaan Otonomi Khusus Papua. Undang-Undang Otonomi Khusus Papua menegaskan bahwa Orang Asli Papua berhak memperoleh kesempatan dan diutamakan untuk mendapatkan pekerjaan dalam semua bidang pekerjaan di wilayah Papua berdasarkan pendidikan dan keahliannya [13][14]. Selain itu, Peraturan Pemerintah Nomor 106 Tahun 2021 mengatur kewenangan Pemerintah Daerah Provinsi Papua dan Pemerintah Daerah Kabupaten/Kota dalam pelaksanaan kebijakan Otonomi Khusus Papua [15]. Pada bidang pendidikan, Peraturan Daerah Khusus Provinsi Papua Nomor 3 Tahun 2013 juga menekankan perlunya perlakuan afirmatif yang tepat sasaran untuk mengatasi masalah mendasar di bidang pendidikan yang dialami Orang Asli Papua [16]. Oleh karena itu, penelitian ini tidak hanya memandang distribusi guru sebagai persoalan teknis penempatan tenaga pendidik, tetapi juga sebagai bagian dari evaluasi pemerataan dan representasi sosial dalam layanan pendidikan di Kabupaten Nabire.

Selanjutnya, Schonlau (2023) menjelaskan bahwa k-Nearest Neighbors (kNN) merupakan metode klasifikasi yang efektif, terutama untuk data yang memiliki pola nonlinier, dan dapat menjadi alternatif yang kompetitif dalam tugas klasifikasi [17]. Meskipun demikian, pembahasan tersebut lebih menitikberatkan pada karakteristik teoritis metode klasifikasi dan belum diarahkan pada perbandingan aplikatif dalam konteks distribusi tenaga pendidik. Kajian tersebut juga belum mengintegrasikan pendekatan klasifikasi dengan hasil *clustering* dalam satu kerangka analisis yang utuh. Berbeda dengan posisi tersebut, penelitian penulis membandingkan beberapa algoritma prediksi setelah tahap *clustering*, sehingga tidak hanya berhenti pada pemahaman teoritis mengenai keunggulan algoritma, tetapi juga menguji secara empiris model yang paling sesuai untuk memprediksi kategori distribusi guru pada satuan pendidikan di Kabupaten Nabire.

Penelitian Arfan dan Paraga (2024) telah menunjukkan bahwa kombinasi K-Means dengan algoritma klasifikasi seperti Naïve Bayes dan Decision Tree dapat digunakan untuk membaca pola data dan mendukung prediksi dalam konteks penjualan BBM [18]. Hasil penelitian tersebut memperlihatkan bahwa integrasi *clustering* dan prediksi cukup efektif untuk membangun dasar pengambilan keputusan berbasis data. Akan tetapi, penelitian tersebut masih berada pada domain bisnis dan distribusi komoditas, sehingga belum menjawab persoalan distribusi tenaga pendidik dalam konteks pendidikan daerah. Oleh karena itu, penelitian ini mengambil posisi pengembangan dengan menerapkan pendekatan serupa pada domain yang berbeda, yaitu distribusi guru ASN Orang Asli Papua dan non-Orang Asli Papua, sehingga berkontribusi pada pemanfaatan data mining dalam evaluasi pemerataan penempatan guru di Kabupaten Nabire.

Berbeda dengan penelitian sebelumnya, penelitian ini tidak hanya melakukan *clustering*, tetapi juga melanjutkannya ke tahap prediksi untuk mengidentifikasi kategori distribusi guru ASN Orang Asli Papua dan non-Orang Asli Papua di Kabupaten Nabire. Dengan demikian, hasil penelitian ini diharapkan dapat memberikan dasar analitis bagi Dinas Pendidikan dalam mengevaluasi pola penempatan guru ASN, khususnya dalam melihat apakah keterwakilan guru OAP telah tersebar secara proporsional pada satuan pendidikan atau masih terkonsentrasi pada sekolah dan jenjang tertentu.

Berdasarkan hal tersebut, penelitian ini memiliki kebaruan pada tiga aspek utama. Pertama, penelitian ini secara khusus menelaah distribusi guru ASN berdasarkan latar belakang Orang Asli Papua dan non-Orang Asli Papua, yang masih jarang diangkat dalam studi pendidikan berbasis data, khususnya dalam konteks Kabupaten Nabire. Kedua, penelitian ini tidak hanya memetakan distribusi guru secara deskriptif, tetapi juga mengelompokkan satuan pendidikan berdasarkan kemiripan pola distribusi guru ASN OAP dan non-OAP melalui *clustering*. Ketiga, penelitian ini melanjutkan hasil pengelompokan tersebut ke tahap klasifikasi untuk menguji apakah kategori distribusi yang terbentuk dapat dikenali kembali secara konsisten oleh model pembelajaran mesin.

Dalam penelitian ini, istilah prediksi tidak dimaknai semata-mata sebagai peramalan kondisi masa depan, tetapi sebagai proses klasifikasi kategori distribusi berdasarkan pola data yang telah terbentuk. Tahap ini diperlukan karena hasil *clustering* bersifat eksploratif, sedangkan klasifikasi digunakan untuk menilai sejauh mana pola kelompok tersebut dapat direpresentasikan oleh model yang terukur melalui metrik evaluasi seperti akurasi, *precision*, *recall*, dan *F1-score*. Dengan demikian, model klasifikasi berfungsi sebagai validasi aplikatif terhadap pola hasil *clustering* sekaligus sebagai alat bantu apabila Dinas Pendidikan melakukan pembaruan data pada periode berikutnya, menambahkan data sekolah baru, atau mengevaluasi perubahan komposisi guru ASN pada satuan pendidikan tertentu. Dengan pendekatan tersebut, penelitian ini tidak hanya menghasilkan informasi tentang pola distribusi guru ASN OAP dan non-OAP, tetapi juga menawarkan kerangka analitis yang dapat digunakan sebagai dasar evaluasi awal terhadap pemerataan penempatan guru. Satuan pendidikan yang diklasifikasikan dalam kategori tanpa keterwakilan atau rendah dapat menjadi prioritas perhatian dalam proses evaluasi dan perencanaan redistribusi guru, sedangkan satuan pendidikan dengan kategori sedang atau tinggi dapat digunakan sebagai pembanding untuk memahami pola distribusi yang lebih proporsional di Kabupaten Nabire. Dengan demikian, penelitian ini memiliki urgensi akademik dan praktis, karena dapat menjadi dasar awal bagi pemangku kepentingan dalam merumuskan langkah-langkah yang lebih tepat untuk mendukung pemerataan penempatan tenaga pendidik di Kabupaten Nabire.

2. BAHAN DAN METODE PENELITIAN

Penelitian penulis menggunakan pendekatan kuantitatif dengan metode data *mining* untuk menganalisis pola distribusi guru ASN Orang Asli Papua dan non-Orang Asli Papua pada satuan pendidikan di Kabupaten Nabire. Pendekatan ini dipilih karena penelitian penulis berorientasi pada pengolahan data terstruktur, pengelompokan pola distribusi, serta prediksi kategori distribusi guru berdasarkan variabel yang tersedia. Secara umum, penelitian penulis dilaksanakan melalui dua tahapan utama, yaitu *clustering* untuk mengidentifikasi pola distribusi sekolah dan prediksi untuk mengklasifikasikan kategori distribusi guru.

Bahan penelitian ini berupa data sekunder guru ASN pada satuan pendidikan di Kabupaten Nabire tahun 2025 yang diperoleh dari Dinas Pendidikan Kabupaten Nabire. Data tersebut mencakup satuan pendidikan pada jenjang TK, SD, SMP, SMA/ sederajat, serta PKBM. Berdasarkan data yang digunakan dalam penelitian ini, jumlah guru ASN yang dianalisis sebanyak 2.147 orang, terdiri atas 364 guru ASN Orang Asli Papua dan 1.783 guru ASN non-Orang Asli Papua. Dari sisi status kepegawaian, guru ASN tersebut terdiri atas 1292 guru PNS dan 855 guru PPPK. Adapun jumlah satuan pendidikan yang menjadi subjek penelitian adalah sebanyak 256 satuan pendidikan.

Perangkat lunak yang digunakan dalam penelitian ini adalah Microsoft Excel dan *Orange Data Mining*. Microsoft Excel digunakan pada tahap awal untuk menyusun, memeriksa, membersihkan, dan mentransformasi data guru ASN sebelum dianalisis lebih lanjut. *Orange Data Mining* digunakan sebagai alat utama untuk pemodelan data, meliputi proses *clustering*, klasifikasi, dan evaluasi model. Penggunaan *Orange Data Mining* dipilih karena mendukung pemrosesan data secara visual, memudahkan eksplorasi pola, serta menyediakan berbagai algoritma analisis yang sesuai dengan kebutuhan penelitian [19].

Sebelum data diolah dalam *Orange Data Mining*, dilakukan tahap pra-pemrosesan data [20]. Penanganan data hilang dilakukan dengan memeriksa kelengkapan atribut utama, yaitu nama satuan pendidikan, distrik/kecamatan, status Orang Asli Papua dan non-Orang Asli Papua, jenis PTK, dan status kepegawaian. Data yang tidak memiliki informasi nama satuan pendidikan atau status OAP/non-OAP tidak digunakan dalam proses agregasi karena atribut tersebut menjadi dasar pembentukan variabel penelitian. Untuk atribut pendukung yang tidak memengaruhi perhitungan utama, data diperiksa kembali dan disesuaikan berdasarkan format yang tersedia dalam rekapitulasi Dinas Pendidikan Kabupaten Nabire.

Penanganan data pencilan dilakukan melalui pemeriksaan nilai jumlah guru ASN pada setiap satuan pendidikan. Nilai yang sangat tinggi atau sangat rendah tidak langsung dihapus, karena dalam konteks distribusi guru, perbedaan jumlah guru antar satuan pendidikan dapat mencerminkan kondisi faktual seperti perbedaan jenjang, ukuran sekolah, jumlah rombongan belajar, atau status satuan pendidikan. Oleh karena itu, pencilan hanya diperiksa untuk memastikan tidak terjadi kesalahan input, duplikasi data, atau ketidaksesuaian kategori. Apabila nilai tersebut sesuai dengan data sumber, maka data tetap dipertahankan dalam analisis. *Outlier* tidak langsung dihapus karena dalam konteks distribusi guru, nilai ekstrem dapat mencerminkan kondisi faktual satuan pendidikan. Oleh karena itu, *outlier* hanya diverifikasi untuk memastikan tidak terjadi kesalahan input atau duplikasi data.

Transformasi data dilakukan dengan mengubah data guru individual menjadi data agregat per satuan pendidikan. Setiap guru ASN dikelompokkan berdasarkan nama satuan pendidikan, kemudian dihitung jumlah guru ASN OAP, jumlah guru ASN non-OAP, total guru ASN, jumlah guru PNS, dan jumlah guru PPPK. Selanjutnya, dihitung persentase guru ASN OAP dan non-OAP pada setiap satuan pendidikan. Persentase guru ASN OAP dihitung dengan membagi jumlah guru ASN OAP dengan total guru ASN pada satuan pendidikan, kemudian dikalikan 100%. Hasil persentase tersebut digunakan sebagai dasar pembentukan kategori distribusi, yaitu tidak ada keterwakilan, rendah, sedang, dan tinggi.

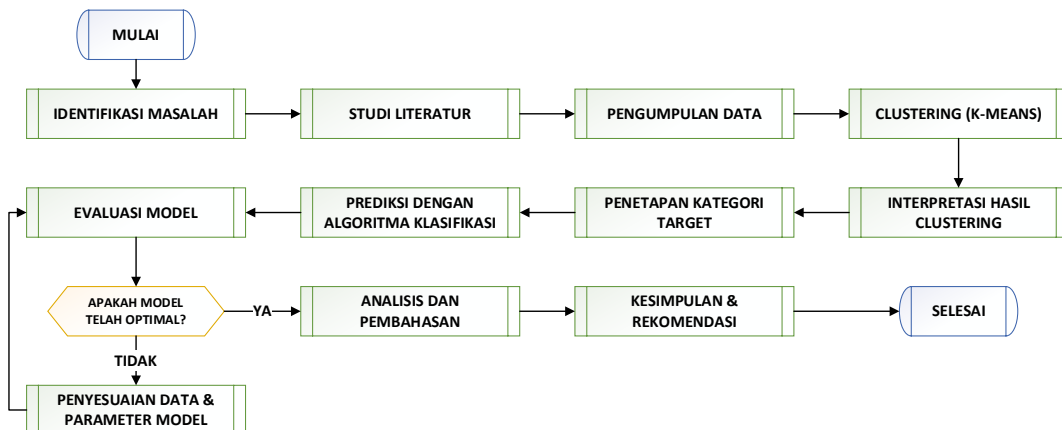
2.1. Tahapan Penelitian

Tahapan penelitian disusun mengikuti alur penelitian yang telah dirancang, yaitu mulai dari identifikasi masalah hingga penyusunan rekomendasi. Tahapan penelitian ini ditunjukkan pada Gambar 1.

Tahap awal penelitian dilakukan melalui identifikasi masalah utama, yaitu belum meratanya distribusi guru ASN Orang Asli Papua dan non-Orang Asli Papua antar satuan pendidikan di Kabupaten Nabire. Permasalahan ini ditandai oleh adanya sekolah yang tidak memiliki guru Orang Asli Papua, sekolah dengan keterwakilan yang rendah, serta sekolah dengan distribusi yang relatif lebih seimbang. Pada tahap ini juga dilakukan studi literatur untuk memperkuat dasar teoritis penelitian. Literatur yang dikaji meliputi konsep distribusi guru, pemerataan pendidikan, data mining dalam bidang pendidikan, algoritma K-Means, algoritma Naive Bayes, algoritma K-Nearest Neighbors (KNN), algoritma Decision Tree, serta penelitian terdahulu yang relevan. Studi literatur dilakukan untuk menemukan *research gap*, menyusun kerangka berpikir, dan menempatkan posisi penelitian penulis dalam konteks penelitian sebelumnya.

Tahap berikutnya adalah pengumpulan data sekunder dari guru ASN di Kabupaten Nabire. Data diperoleh dari kantor dinas Pendidikan kabupaten nabire yang memuat informasi guru berdasarkan satuan pendidikan. Pada tahap ini, peneliti melakukan inventarisasi seluruh atribut yang tersedia, lalu memilih atribut yang relevan dengan fokus penelitian. Unit analisis yang digunakan adalah satuan pendidikan/sekolah, sehingga data guru individu perlu disusun kembali menjadi data agregat per sekolah. Pra-pemrosesan dilakukan

melalui seleksi atribut, pembersihan data, penghapusan duplikasi, penanganan data kosong, transformasi data individu menjadi agregat per sekolah, dan normalisasi data numerik. Tahap ini bertujuan untuk memastikan bahwa data yang digunakan berkualitas baik dan siap diolah pada tahap pemodelan. Unit analisis dalam penelitian ini adalah sekolah/satuan pendidikan, sehingga setiap baris data akhir merepresentasikan satu sekolah.



Gambar 1. Tahapan Penelitian

Pada tahap clustering, penelitian menggunakan algoritma K-Means untuk mengelompokkan sekolah berdasarkan kemiripan pola distribusi guru ASN Orang Asli Papua dan non-Orang Asli Papua. Variabel yang digunakan dalam proses clustering adalah atribut numerik yang merepresentasikan komposisi guru per sekolah. Hasil *clustering* kemudian dievaluasi menggunakan *Silhouette Coefficient* untuk menilai kualitas pemisahan antar-cluster. Selanjutnya, setiap *cluster* diinterpretasikan untuk mengidentifikasi kelompok sekolah yang menunjukkan pola distribusi timpang, sedang, atau relatif lebih merata.

Setelah tahap clustering, penelitian dilanjutkan dengan penetapan kategori distribusi guru sebagai target klasifikasi. Kategori yang digunakan terdiri atas tidak ada OAP, Rendah, Sedang, dan Tinggi. Kategori ini ditentukan berdasarkan tingkat keterwakilan guru ASN Orang Asli Papua pada masing-masing sekolah. Penetapan kategori target bertujuan untuk membangun model prediksi yang mampu memperkirakan posisi distribusi setiap satuan pendidikan secara lebih sistematis.

Tahap prediksi dilakukan dengan membandingkan tiga algoritma klasifikasi, yaitu Naive Bayes, K-NN, dan Decision Tree. Ketiga algoritma tersebut dipilih karena mewakili pendekatan klasifikasi yang berbeda dan banyak digunakan dalam penelitian data *mining* [21]. Model dibangun dan diuji menggunakan skema validasi pada *Orange Data Mining* untuk mengetahui performa masing-masing algoritma dalam memprediksi kategori distribusi guru [22].

Evaluasi model dilakukan menggunakan beberapa metrik, yaitu *AUC (Area Under Curve)*, *Accuracy*, *Precision*, *Recall*, *F1-Score*, dan *Confusion Matrix*. *Accuracy* digunakan untuk melihat tingkat ketepatan prediksi secara umum, *Precision* untuk menilai ketepatan hasil klasifikasi, *Recall* untuk melihat kemampuan model mengenali kelas aktual, dan *F1-Score* untuk menunjukkan keseimbangan antara *Precision* dan *Recall*. Sementara itu, *Confusion Matrix* digunakan untuk mengidentifikasi distribusi prediksi benar dan salah pada masing-masing kategori. Model dengan performa terbaik selanjutnya dijadikan dasar dalam analisis prediktif penelitian.

Tahap akhir adalah analisis dan interpretasi hasil. Hasil clustering digunakan untuk menjelaskan pola kelompok sekolah berdasarkan distribusi guru ASN Orang Asli Papua dan non-Orang Asli Papua, sedangkan hasil prediksi digunakan untuk menilai kemampuan model dalam mengenali kategori distribusi guru. Analisis dilakukan dengan menafsirkan cluster yang menunjukkan ketimpangan, cluster yang relatif lebih merata, serta mengidentifikasi sekolah atau wilayah yang memerlukan perhatian dalam evaluasi pemerataan penempatan guru. Berdasarkan keseluruhan hasil tersebut, penelitian ini menghasilkan kesimpulan dan rekomendasi sebagai dasar evaluasi distribusi guru ASN di Kabupaten Nabire.

2.2. Metode Analisis Data

Metode pertama yang digunakan dalam penelitian penulis adalah K-Means clustering. Algoritma ini digunakan untuk mengelompokkan sekolah berdasarkan kemiripan distribusi guru ASN Orang Asli Papua dan non-Orang Asli Papua. Pada algoritma K-Means, kedekatan antara data dengan pusat cluster dihitung menggunakan jarak *Euclidean*, pada persamaan 1.

$$d(x_i, c_j) = \sqrt{\sum_{k=1}^n (x_{ik} - c_{jk})^2} \quad (1)$$

Pada rumus jarak *Euclidean*, $d(x_i, c_j)$ menyatakan jarak antara data ke- i dengan *centroid* atau pusat *cluster* ke- j . Simbol x_i menunjukkan data ke- i yang akan dihitung kedekatannya terhadap suatu *cluster*, sedangkan c_j menunjukkan *centroid* dari *cluster* ke- j . Nilai x_{ik} merupakan nilai atribut ke- k pada data ke- i , sementara c_{jk} adalah nilai atribut ke- k pada *centroid* ke- j . Simbol n menyatakan jumlah atribut yang digunakan dalam proses perhitungan. Dengan demikian, rumus ini digunakan untuk mengukur seberapa dekat suatu data dengan pusat *cluster* berdasarkan seluruh atribut yang dianalisis. Semakin kecil nilai jarak *Euclidean*, maka semakin besar kedekatan data tersebut terhadap *cluster* yang bersangkutan. Pusat *cluster* atau *centroid* dihitung dari rata-rata seluruh data dalam satu *cluster* [23], dinyatakan sebagai Persamaan 2.

$$c_j = \frac{1}{n_j} \sum_{x_i \in C_j} x_i \quad (2)$$

Pada rumus *centroid*, c_j menyatakan pusat *cluster* ke- j , yaitu titik rata-rata yang mewakili seluruh data dalam *cluster* tersebut. Simbol n_j menunjukkan jumlah data yang menjadi anggota pada *cluster* ke- j , sedangkan C_j merupakan himpunan seluruh data yang berada dalam *cluster* tersebut. Simbol x_i menyatakan data ke- i yang termasuk dalam *cluster* ke- j . Dengan demikian, *centroid* diperoleh dengan menjumlahkan seluruh nilai data pada satu *cluster*, kemudian membaginya dengan jumlah anggota *cluster* tersebut. Nilai *centroid* ini berfungsi sebagai pusat acuan dalam proses K-Means untuk menentukan kedekatan data terhadap suatu *cluster* dan memperbarui pembentukan *cluster* secara iteratif hingga diperoleh hasil pengelompokan yang stabil [24]. Tujuan algoritma K-Means adalah meminimalkan jumlah kuadrat jarak antara data dan *centroid* dalam *cluster*, yang dikenal sebagai *Within Cluster Sum of Squares* (WCSS), dinyatakan sebagai Persamaan 3.

$$J = \sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - c_j\|^2 \quad (3)$$

Pada rumus *Within Cluster Sum of Squares* (WCSS), simbol J menyatakan fungsi objektif pada algoritma K-Means yang digunakan untuk mengukur total variasi data di dalam *cluster*. Simbol k menunjukkan jumlah *cluster* yang dibentuk, sedangkan C_j merupakan himpunan data yang termasuk dalam *cluster* ke- j . Simbol x_i menyatakan data ke- i yang berada pada *cluster* tertentu, dan c_j adalah *centroid* atau pusat *cluster* ke- j . Notasi $\|x_i - c_j\|^2$ menunjukkan kuadrat jarak antara data ke- i dengan *centroid cluster* ke- j . Dengan demikian, WCSS menghitung total kuadrat jarak seluruh data terhadap pusat *cluster*nya masing-masing. Semakin kecil nilai WCSS, maka semakin baik kualitas pengelompokan karena menunjukkan bahwa data dalam satu *cluster* semakin dekat dengan *centroid* dan memiliki tingkat kemiripan yang lebih tinggi [25].

Pada tahap prediksi, penelitian ini menggunakan tiga algoritma klasifikasi, yaitu Naive Bayes, K-NN, dan Decision Tree. Naive Bayes digunakan untuk menghitung probabilitas suatu data termasuk ke dalam kelas tertentu berdasarkan *Teorema Bayes*. KNN digunakan untuk menentukan kelas data berdasarkan kedekatan terhadap sejumlah tetangga terdekat menggunakan ukuran jarak. Sementara itu, Decision Tree membentuk aturan keputusan melalui pemilihan atribut terbaik berdasarkan nilai entropi dan *information gain*. Algoritma Naive Bayes didasarkan pada *Teorema Bayes*, karena Naive Bayes mengasumsikan setiap atribut saling bebas sehingga persamaan Naive Bayes dinyatakan pada persamaan 4.

$$P(C_k | X) = \frac{P(C_k) \prod_{i=1}^n P(x_i | C_k)}{P(X)} \quad (4)$$

Pada rumus Naive Bayes, $P(C_k | X)$ menyatakan probabilitas bahwa data X termasuk ke dalam kelas C_k setelah mempertimbangkan atribut yang dimiliki data tersebut. Simbol $P(X | C_k)$ menunjukkan probabilitas kemunculan data X jika diketahui data tersebut berada pada kelas C_k , sedangkan $P(C_k)$ merupakan probabilitas awal atau prior dari kelas C_k . Simbol $P(X)$ menyatakan probabilitas total dari data X yang diamati. Adapun C_k menunjukkan kelas ke- k , sedangkan X merupakan kumpulan atribut atau fitur yang digunakan dalam proses klasifikasi. Pada asumsi independensi yang digunakan dalam Naive Bayes, setiap atribut x_i dianggap saling bebas satu sama lain, sehingga probabilitas gabungan dapat dihitung dari hasil perkalian probabilitas tiap atribut terhadap kelas tertentu. Dengan demikian, algoritma Naive Bayes digunakan untuk menentukan kelas dengan probabilitas posterior tertinggi berdasarkan data yang diamati. Algoritma KNN menentukan kelas suatu data berdasarkan kedekatan jarak terhadap sejumlah tetangga terdekat [26]. Rumus jarak yang umum digunakan adalah *Euclidean distance* yang dinyatakan pada persamaan 5.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (5)$$

Pada rumus KNN, $d(x, y)$ menyatakan jarak antara data uji x dan data latih y . Simbol x_i menunjukkan nilai atribut ke- i pada data uji, sedangkan y_i menyatakan nilai atribut ke- i pada data latih. Simbol n menunjukkan jumlah atribut yang digunakan dalam perhitungan jarak. Rumus ini digunakan untuk menghitung kedekatan

antara data baru dengan data yang sudah diketahui kelasnya. Semakin kecil nilai jarak *Euclidean*, maka semakin dekat hubungan antara data uji dan data latih tersebut. Dalam algoritma KNN, kelas suatu data ditentukan berdasarkan mayoritas kelas dari sejumlah k tetangga terdekat. Oleh karena itu, metode ini sangat bergantung pada ukuran jarak dan kedekatan antar data dalam ruang atribut [27]. Pada Decision Tree, salah satu ukuran yang digunakan untuk memilih atribut terbaik adalah *Entropy* yang dijelaskan pada persamaan 6.

$$Entropy(S) = -\sum_{i=1}^c P_i \log 2P_i \quad (6)$$

Pada rumus *Entropy*, $Entropy(S)$ menyatakan tingkat ketidakpastian atau ketidakmurnian dari suatu himpunan data S . Simbol p_i menunjukkan proporsi data yang termasuk ke dalam kelas ke- i , sedangkan c menyatakan jumlah seluruh kelas yang terdapat dalam data. Nilai entropi digunakan untuk mengukur seberapa bercampurnya data dalam suatu himpunan. Jika semua data berada pada satu kelas yang sama, maka nilai entropi akan rendah, sedangkan jika data tersebar ke beberapa kelas secara lebih merata, maka nilai entropinya akan lebih tinggi [28]. Dalam algoritma Decision Tree, entropi digunakan sebagai dasar untuk menentukan seberapa baik suatu atribut dapat memisahkan data ke dalam kelas-kelas yang lebih homogen. Selanjutnya, untuk menentukan atribut terbaik digunakan Information Gain yang dijelaskan pada persamaan 7.

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (7)$$

Pada rumus *Information Gain*, $Gain(S, A)$ menyatakan besarnya informasi yang diperoleh dari atribut A dalam membagi data S ke dalam kelompok yang lebih terstruktur. Simbol $Entropy(S)$ menunjukkan entropi total dari himpunan data sebelum dilakukan pemisahan, sedangkan S_v merupakan subset data hasil pembagian berdasarkan nilai v dari atribut A . Simbol $|S_v|$ menunjukkan jumlah data pada subset S_v , sedangkan $|S|$ adalah jumlah total data pada himpunan awal. Notasi $Values(A)$ menunjukkan seluruh nilai yang dimiliki oleh atribut A . *Information Gain* dihitung dengan mengurangi entropi total data dengan jumlah entropi subset yang telah diberi bobot berdasarkan proporsi datanya. Semakin besar nilai information gain, maka semakin baik atribut tersebut dalam memisahkan data ke dalam kelas yang lebih homogen [29]. Oleh karena itu, pada Decision Tree, atribut dengan nilai information gain tertinggi akan dipilih sebagai *node* atau percabangan terbaik.

3. HASIL DAN PEMBAHASAN

Penelitian penulis menggunakan data guru Aparatur Sipil Negara (ASN) pada satuan pendidikan di Kabupaten Nabire yang telah melalui tahap seleksi dan penyesuaian sesuai kebutuhan analisis. Data akhir disusun pada level sekolah, sehingga setiap baris data merepresentasikan satu satuan pendidikan. Variabel utama yang digunakan meliputi kecamatan/distrik, jumlah sekolah, jumlah guru ASN Orang Asli Papua, jumlah guru ASN non-Orang Asli Papua, total guru ASN. Penyusunan data pada level sekolah dilakukan agar proses clustering dan prediksi dapat menggambarkan pola distribusi guru secara lebih terstruktur.

Tabel 1. Ringkasan Data Penelitian

Kecamatan	Jumlah Sekolah	Jumlah PNS	Jumlah PPPK	Jumlah Guru OAP	Jumlah Guru Non OAP	Total Guru ASN
Dipa	4	8	0	7	1	8
Makimi	12	74	58	25	107	132
Menou	2	3	2	5	0	5
Moora	6	9	3	8	4	12
Nabire	113	791	404	160	1035	1195
Nabire Barat	22	110	105	48	167	215
Napan	7	12	17	6	23	29
Siriwo	9	15	10	7	18	25
Teluk Kimi	21	75	59	28	106	134
Teluk Umar	5	9	6	3	12	15
Uwapa	11	40	25	19	46	65
Wanggar	13	58	54	12	100	112
Wapoga	5	3	16	5	14	19
Yaro	20	75	77	25	127	152
Yaur	6	10	19	6	23	29

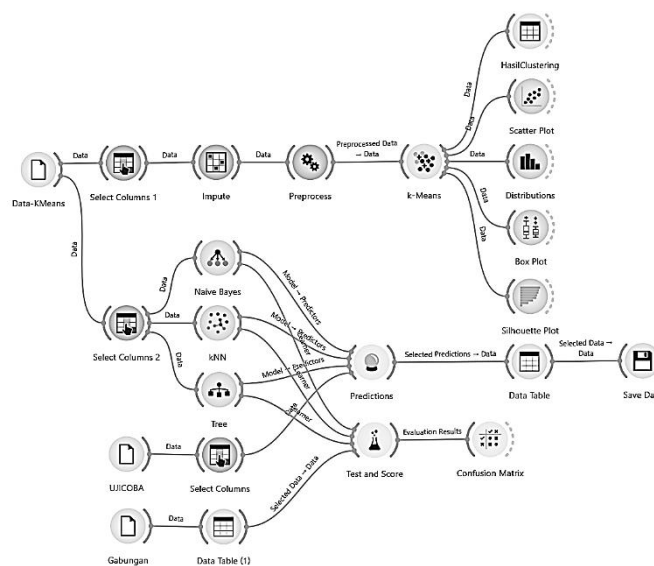
(Sumber: Data Dinas Pendidikan Kabupaten Nabire, 2025)

3.1. Hasil Pra-pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk memastikan data berada dalam kondisi siap analisis. Proses ini mencakup seleksi atribut, pembersihan data, penghapusan duplikasi, penanganan nilai kosong, transformasi data individu guru menjadi data agregat per satuan pendidikan, serta normalisasi data numerik. Hasil tahap ini

adalah dataset akhir yang lebih konsisten, terstandar, dan sesuai untuk digunakan pada proses clustering dan klasifikasi. Dengan demikian, kualitas data masukan dapat mendukung ketepatan hasil analisis.

Setelah dataset akhir diperoleh, proses analisis dilakukan melalui dua tahapan utama, yaitu *clustering* dan klasifikasi. Pada tahap clustering, data agregat per satuan pendidikan digunakan untuk mengelompokkan sekolah berdasarkan kemiripan pola distribusi guru ASN Orang Asli Papua dan non-Orang Asli Papua. Hasil clustering kemudian digunakan sebagai dasar pembentukan label atau kategori distribusi, seperti tidak ada keterwakilan, rendah, sedang, dan tinggi. Pada tahap klasifikasi, dataset yang telah memiliki label kategori distribusi dievaluasi menggunakan pendekatan K-Fold Cross Validation pada *Orange Data Mining* melalui *widget Test & Score*. Dalam penelitian ini digunakan skema 10-Fold Cross Validation, yaitu dataset dibagi menjadi sepuluh bagian secara bergantian. Sebanyak sembilan bagian digunakan sebagai data latih, sedangkan satu bagian digunakan sebagai data uji. Proses ini diulang sebanyak sepuluh kali hingga seluruh bagian data memperoleh kesempatan menjadi data uji. Pendekatan ini dipilih karena mampu memberikan evaluasi model yang lebih stabil dibandingkan pembagian data latih dan data uji secara tunggal, terutama pada dataset dengan jumlah satuan pendidikan yang terbatas. Evaluasi model klasifikasi dilakukan dengan membandingkan beberapa algoritma berdasarkan nilai *Accuracy*, *Precision*, *Recall*, *F1-Score*, dan *AUC*. Hasil evaluasi tersebut digunakan untuk menentukan algoritma yang paling sesuai dalam mengenali kategori distribusi guru ASN OAP dan non-OAP pada satuan pendidikan di Kabupaten Nabire.



Gambar 2. *Workflow Clustering dan Prediksi Distribusi Guru ASN (Aplikasi Orange)*

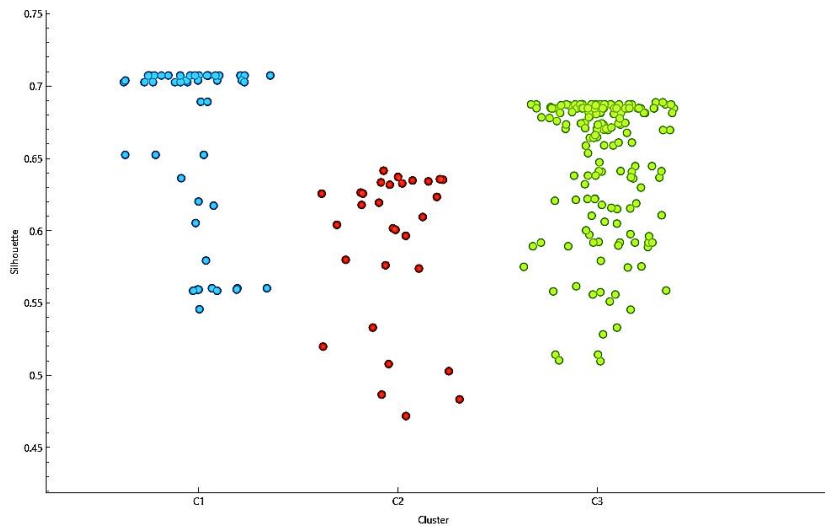
Workflow pada gambar 2 menunjukkan alur analisis data di *Orange Data Mining* yang terdiri atas dua bagian utama, yaitu clustering dan prediksi. Pada bagian pertama, data dimasukkan melalui *Data-KMeans*, lalu dipilih atribut yang relevan pada *Select Columns 1*. Setelah itu data diperbaiki melalui *Impute* untuk menangani nilai kosong dan *Preprocess* untuk tahap pembersihan atau penyesuaian data. Data yang telah siap kemudian dianalisis menggunakan *K-Means* untuk membentuk kelompok sekolah berdasarkan pola distribusi guru. Hasil clustering ini divisualisasikan dan diinterpretasikan melalui *Hasil Clustering*, *Scatter Plot*, *Distributions*, *Box Plot*, dan *Silhouette Plot*. Bagian ini bertujuan untuk melihat pola distribusi guru ASN Orang Asli Papua dan non-Orang Asli Papua serta kualitas pembentukan *cluster*.

Pada bagian kedua, data diproses kembali melalui *Select Columns 2* untuk menyesuaikan atribut yang digunakan pada tahap klasifikasi. Tiga algoritma prediksi yang dipakai adalah *Naive Bayes*, *KNN*, dan *Tree*. Ketiga model tersebut dihubungkan ke *widget Predictions* untuk melihat hasil prediksi, dan ke *Test and Score* untuk membandingkan performa masing-masing model. Hasil evaluasi model selanjutnya dianalisis melalui *Confusion Matrix* untuk melihat ketepatan klasifikasi tiap kategori distribusi guru. Output prediksi kemudian ditampilkan dalam *Data Table* dan disimpan melalui *Save Data*. *workflow* pada gambar 2 menunjukkan bahwa penelitian dilakukan melalui pengelompokan data sekolah terlebih dahulu dengan *K-Means*, kemudian dilanjutkan dengan prediksi kategori distribusi guru menggunakan tiga algoritma klasifikasi, sehingga hasil akhirnya dapat digunakan untuk membaca pola distribusi sekaligus mengevaluasi model prediksi terbaik.

3.2. Hasil Clustering Menggunakan K-Means

Algoritma *K-Means* digunakan untuk mengelompokkan sekolah berdasarkan kemiripan distribusi guru ASN Orang Asli Papua dan non-Orang Asli Papua. Hasil analisis menunjukkan bahwa data sekolah di Kabupaten Nabire terbagi ke dalam tiga *cluster* utama, yang menandakan bahwa distribusi guru tidak bersifat

homogen. Pembentukan tiga cluster ini menunjukkan adanya perbedaan karakteristik distribusi antar sekolah, baik dari sisi jumlah guru, proporsi keterwakilan guru Orang Asli Papua, maupun komposisi guru secara umum. Dengan demikian, hasil *clustering* menjadi dasar awal untuk membaca pola ketimpangan dan pemerataan distribusi guru.



Gambar 3. Hasil *Clustering K-Means (Scatter Plot)*

Scatter plot tersebut menampilkan hasil pengelompokan data sekolah ke dalam tiga *cluster*, yaitu C1, C2, dan C3. Sumbu horizontal menunjukkan cluster, sedangkan sumbu vertikal menunjukkan nilai *silhouette* dari setiap anggota data. Nilai *silhouette* digunakan untuk melihat seberapa baik suatu data cocok dengan clusternya sendiri dibandingkan dengan *cluster* lain. Semakin tinggi nilainya, maka semakin baik posisi data tersebut di dalam cluster.

Tabel 2. Cluster Distribusi Guru ASN OAP dan non-OAP di Kabupaten Nabire

Cluster	Jumlah Satuan Pendidikan	Jumlah Guru ASN OAP	Jumlah Guru ASN non-OAP	% OAP	% Non-OAP	Rata-rata Guru ASN per Satuan Pendidikan	Distrik Dominan	Interpretasi Profil Cluster
C1	47	82	26	75,93 %	24,07 %	2,30	Nabire, Moora, Teluk Kimi, Dipa, Uwapa	Cluster dengan keterwakilan guru ASN OAP tinggi, tetapi umumnya berada pada satuan pendidikan dengan jumlah guru ASN relatif kecil. Cluster dengan jumlah guru ASN besar, tetapi didominasi guru ASN non-OAP. Cluster mayoritas yang tersebar di hampir semua distrik dengan komposisi guru ASN non-OAP yang dominan.
C2	30	123	830	12,91 %	87,09 %	31,77	Nabire, Nabire Barat, Makimi, Wanggar, Yaro	
C3	179	159	927	14,64 %	85,36 %	6,07	Nabire, Nabire Barat, Teluk Kimi, Yaro, Wanggar, Makimi	
Total	256	364	1.783	16,95	83,05	8,39	-	-

Berdasarkan Tabel 2, hasil *clustering* menunjukkan bahwa setiap *cluster* memiliki karakteristik distribusi guru ASN OAP dan non-OAP yang berbeda. *Cluster* 1 (C1) terdiri atas 47 satuan pendidikan dengan total 108 guru ASN. Dari jumlah tersebut, 82 guru atau 75,93% merupakan guru ASN OAP, sedangkan 26 guru atau 24,07% merupakan guru ASN non-OAP. Secara substantif, C1 dapat diinterpretasikan sebagai *cluster* dengan tingkat keterwakilan guru ASN OAP yang tinggi. Namun demikian, rata-rata jumlah guru ASN pada C1 hanya sebesar 2,30 guru per satuan pendidikan. Hal ini menunjukkan bahwa meskipun proporsi OAP tinggi, *cluster* ini umumnya terdiri atas satuan pendidikan dengan jumlah guru ASN relatif kecil. Secara administratif, C1 tersebar pada beberapa distrik, dengan jumlah anggota yang relatif menonjol di Distrik Nabire, Moora, Teluk Kimi, Dipa, dan Uwapa.

Cluster 2 (C2) terdiri atas 30 satuan pendidikan dengan total 953 guru ASN. Dari jumlah tersebut, 123 guru atau 12,91% merupakan guru ASN OAP, sedangkan 830 guru atau 87,09% merupakan guru ASN non-OAP. C2 memiliki rata-rata jumlah guru ASN tertinggi, yaitu 31,77 guru per satuan pendidikan. Dengan demikian, C2 dapat dipahami sebagai *cluster* yang merepresentasikan satuan pendidikan dengan jumlah guru ASN besar, tetapi komposisinya masih sangat didominasi oleh guru ASN non-OAP. Secara geografis-administratif, C2 paling banyak terdapat di Distrik Nabire, yaitu 24 satuan pendidikan, sementara sisanya tersebar di Nabire Barat, Makimi, Wanggar, dan Yaro. Kondisi ini menunjukkan bahwa dominasi guru ASN non-OAP tidak hanya terjadi pada distrik pinggiran, tetapi juga tampak kuat pada wilayah pusat layanan pendidikan.

Sementara itu, *Cluster* 3 (C3) merupakan *cluster* terbesar dengan 179 satuan pendidikan atau sekitar 69,92% dari seluruh satuan pendidikan yang dianalisis. *Cluster* ini memiliki total 1.086 guru ASN, terdiri atas 159 guru ASN OAP atau 14,64% dan 927 guru ASN non-OAP atau 85,36%. C3 memiliki rata-rata 6,07 guru ASN per satuan pendidikan, sehingga dapat dikategorikan sebagai *cluster* mayoritas dengan ukuran satuan pendidikan sedang hingga kecil. Secara administratif, C3 mendominasi hampir semua distrik, terutama Distrik Nabire sebanyak 77 satuan pendidikan, Nabire Barat sebanyak 17 satuan pendidikan, Teluk Kimi dan Yaro masing-masing sebanyak 16 satuan pendidikan, Wanggar sebanyak 11 satuan pendidikan, serta Makimi sebanyak 9 satuan pendidikan. Dominasi C3 menunjukkan bahwa pola umum distribusi guru ASN di Kabupaten Nabire masih didominasi oleh guru ASN non-OAP.

Dengan demikian, temuan *clustering* tidak hanya menunjukkan kualitas pengelompokan secara teknis melalui nilai *silhouette*, tetapi juga memperlihatkan pola substantif distribusi guru ASN OAP dan non-OAP. C1 menunjukkan kelompok dengan keterwakilan OAP tinggi, tetapi jumlah guru ASN per satuan pendidikan relatif kecil. C2 menunjukkan kelompok dengan jumlah guru ASN besar, tetapi sangat didominasi oleh non-OAP. Sementara itu, C3 menjadi *cluster* mayoritas yang tersebar luas di hampir seluruh distrik dengan proporsi OAP yang masih rendah. Temuan ini penting bagi Dinas Pendidikan Kabupaten Nabire karena menunjukkan bahwa persoalan keterwakilan guru ASN OAP tidak cukup dibaca dari total kabupaten, tetapi perlu dilihat berdasarkan profil *cluster*, ukuran sekolah, dan sebaran distrik.

3.3. Hasil Penilaian Algoritma Klasifikasi

Tahap prediksi dilakukan untuk menguji sejauh mana kategori distribusi guru dapat dikenali secara konsisten oleh model klasifikasi. Tiga algoritma yang digunakan adalah Naive Bayes, KNN, dan Decision Tree. Proses ini bertujuan untuk membandingkan performa masing-masing algoritma dalam memprediksi kategori distribusi guru pada satuan pendidikan. Hasil prediksi kemudian digunakan untuk menentukan model terbaik yang dapat mendukung evaluasi distribusi guru secara lebih akurat.

Tabel 3. Hasil Test and Score Algoritma Klasifikasi

Model	AUC	CA	F1	Precision	Recall
Naive Bayes	0,993	0,945	0,945	0,947	0,945
k-Nearest Neighbors (kNN)	0,968	0,871	0,875	0,886	0,871
Decision Tree	0,994	0,992	0,992	0,992	0,992

Berdasarkan Tabel 3, algoritma Decision Tree menunjukkan performa terbaik dibandingkan Naive Bayes dan KNN. Decision Tree memperoleh nilai tertinggi pada seluruh metrik evaluasi, yaitu AUC sebesar 0,994, CA sebesar 0,992, F1 sebesar 0,992, *Precision* sebesar 0,992, dan *Recall* sebesar 0,992. Hasil ini menunjukkan bahwa Decision Tree merupakan model yang paling baik dalam mengklasifikasikan kategori distribusi guru ASN OAP dan non-OAP pada satuan pendidikan di Kabupaten Nabire.

Keunggulan Decision Tree dalam penelitian ini dapat dijelaskan dari karakteristik dataset distribusi guru ASN di Kabupaten Nabire. Dataset yang digunakan memiliki pola kategorisasi yang sangat berkaitan dengan variabel numerik terstruktur, seperti jumlah guru ASN OAP, jumlah guru ASN non-OAP, total guru ASN, persentase guru ASN OAP, persentase guru ASN non-OAP, jumlah PNS, dan jumlah PPPK. Kategori distribusi seperti “tidak ada OAP”, “rendah”, “sedang”, dan “tinggi” pada dasarnya dapat dibedakan melalui batas-batas nilai tertentu, terutama pada variabel jumlah dan persentase guru ASN OAP. Karakter seperti ini sangat sesuai dengan cara kerja Decision Tree, karena algoritma ini membangun aturan keputusan berbasis pemisahan nilai

atribut. Dengan kata lain, Decision Tree mampu menangkap pola berbentuk aturan, misalnya ketika jumlah guru ASN OAP sama dengan nol maka satuan pendidikan cenderung masuk kategori “tidak ada OAP”, atau ketika persentase guru ASN OAP berada pada rentang tertentu, maka satuan pendidikan dapat diklasifikasikan ke kategori rendah, sedang, atau tinggi.

Decision Tree bekerja lebih baik daripada kNN karena kNN sangat bergantung pada jarak antardata. Pada dataset distribusi guru, satuan pendidikan dapat memiliki nilai yang sangat beragam, terutama antara sekolah besar di wilayah pusat seperti Distrik Nabire dan sekolah kecil di distrik lain. Perbedaan jumlah total guru ASN antarsatuan pendidikan dapat memengaruhi perhitungan jarak, sehingga data yang sebenarnya berbeda secara kategori distribusi dapat terlihat berdekatan dalam ruang fitur, atau sebaliknya. Hal ini menyebabkan KNN lebih rentan mengalami kesalahan klasifikasi, khususnya pada kategori yang memiliki batas transisi, seperti rendah dan sedang. Kondisi tersebut tercermin pada *confusion matrix* KNN, di mana masih terdapat kesalahan klasifikasi yang cukup besar antara kategori rendah, sedang, dan tidak ada OAP.

Decision Tree juga lebih unggul dibandingkan dengan Naive Bayes karena Naive Bayes bekerja dengan asumsi bahwa setiap atribut bersifat independen satu sama lain. Dalam konteks dataset distribusi guru, asumsi ini sulit dipenuhi karena beberapa atribut memiliki hubungan yang sangat kuat. Misalnya, jumlah guru ASN OAP, jumlah guru ASN non-OAP, total guru ASN, persentase guru ASN OAP, dan persentase guru ASN non-OAP saling berkaitan secara matematis. Ketika atribut-atribut tersebut saling bergantung, asumsi independensi pada Naive Bayes dapat menyebabkan probabilitas kelas menjadi kurang optimal. Meskipun Naive Bayes tetap menunjukkan performa yang baik, Decision Tree lebih mampu menangani hubungan antarvariabel tersebut karena tidak mensyaratkan independensi antaratribut. Decision Tree dapat langsung memilih atribut yang paling informatif dan membentuk aturan klasifikasi berdasarkan struktur data yang ada.

Tabel 4. Ringkasan Hasil Confusion Matrix

Model	Jumlah Data	Prediksi Benar	Prediksi Salah	Akurasi Berdasarkan CM
k-Nearest Neighbors (kNN)	256	223	33	87,11%
Naive Bayes	256	242	14	94,53%
Decision Tree	256	254	2	99,22%

Berdasarkan *confusion matrix*, Decision Tree menghasilkan jumlah kesalahan klasifikasi paling sedikit, yaitu hanya 2 kesalahan dari 256 data. Model ini berhasil mengklasifikasikan seluruh data pada kategori “Sedang” dan “Tidak ada OAP” dengan tepat, serta hanya mengalami kesalahan pada kategori “Rendah” dan “Tinggi”. Naive Bayes menghasilkan 14 kesalahan klasifikasi, terutama pada kategori “Sedang” dan “Tinggi”. Sementara itu, kNN menghasilkan 33 kesalahan klasifikasi dan menjadi model dengan performa terendah dibandingkan dengan dua algoritma lainnya. Hasil ini memperkuat temuan *pada Test and Score* bahwa Decision Tree merupakan algoritma paling optimal untuk mengklasifikasikan kategori distribusi guru ASN OAP dan non-OAP pada satuan pendidikan di Kabupaten Nabire.

Secara saintifik, performa Decision Tree yang lebih tinggi menunjukkan bahwa pola distribusi kategori guru ASN OAP dan non-OAP di Kabupaten Nabire cenderung bersifat rule-based atau berbasis aturan nilai tertentu. Hal ini logis karena kategori distribusi guru dalam penelitian ini dibentuk dari indikator kuantitatif yang memiliki struktur jelas, seperti jumlah guru ASN OAP dan persentase guru ASN OAP pada setiap satuan pendidikan. Pada kondisi seperti ini, algoritma yang mampu membentuk aturan pemisahan secara eksplisit akan lebih sesuai dibandingkan dengan algoritma yang berbasis jarak atau probabilitas sederhana. Decision Tree dapat mengidentifikasi atribut yang paling menentukan dalam proses klasifikasi dan memisahkan data ke dalam kategori distribusi secara bertahap berdasarkan nilai ambang tertentu.

Keunggulan Decision Tree juga berkaitan dengan karakteristik data pendidikan di Kabupaten Nabire yang tidak sepenuhnya homogen. Jumlah guru ASN pada setiap satuan pendidikan berbeda-beda, baik karena perbedaan jenjang pendidikan, ukuran sekolah, lokasi distrik, maupun status satuan pendidikan. Dalam dataset seperti ini, model berbasis jarak seperti kNN dapat mengalami keterbatasan karena sensitif terhadap skala dan kedekatan numerik antarobjek. Sementara itu, Naive Bayes dapat mengalami keterbatasan karena hubungan antar variabel distribusi guru tidak sepenuhnya independen. Decision Tree lebih adaptif terhadap kondisi tersebut karena mampu menangkap hubungan nonlinier dan interaksi antaratribut tanpa memerlukan asumsi distribusi data tertentu.

Hasil ini memperlihatkan bahwa Decision Tree tidak hanya unggul secara statistik, tetapi juga lebih mudah diinterpretasikan untuk kebutuhan kebijakan pendidikan. Model Decision Tree dapat diterjemahkan ke dalam aturan keputusan yang sederhana dan dapat dipahami oleh pengambil kebijakan, misalnya satuan pendidikan dengan jumlah guru ASN OAP nol dapat dikategorikan sebagai tidak ada keterwakilan, sedangkan satuan pendidikan dengan persentase OAP tertentu dapat dikelompokkan ke dalam kategori rendah, sedang, atau tinggi. Dengan demikian, Decision Tree memiliki nilai tambah praktis bagi Dinas Pendidikan Kabupaten Nabire karena tidak hanya menghasilkan akurasi prediksi yang tinggi, tetapi juga menyediakan pola keputusan yang transparan untuk mendukung evaluasi pemerataan penempatan guru ASN OAP dan non-OAP.

Meskipun Decision Tree menunjukkan performa terbaik, hasil ini tetap perlu dipahami dalam konteks dataset penelitian yang digunakan, yaitu data guru ASN pada satuan pendidikan di Kabupaten Nabire tahun 2025. Oleh karena itu, pengujian pada data tahun berikutnya atau pada wilayah lain tetap diperlukan untuk menilai konsistensi performa model secara lebih luas.

3.4. Ketimpangan dan Pemerataan Distribusi

Hasil *clustering* dan klasifikasi menunjukkan bahwa distribusi guru ASN Orang Asli Papua dan non-Orang Asli Papua pada satuan pendidikan di Kabupaten Nabire belum menunjukkan pola yang proporsional. Ketimpangan terutama terlihat pada satuan pendidikan dalam kategori “Tidak ada OAP” dan “Rendah”, karena kedua kategori tersebut menunjukkan rendahnya keterwakilan guru ASN OAP. Namun, kategori tersebut tidak dapat langsung dimaknai sebagai kekurangan guru secara absolut, sebab penentuan kebutuhan guru tetap memerlukan indikator tambahan seperti rasio guru-siswa, jumlah rombongan belajar, jenjang pendidikan, kebutuhan guru mata pelajaran, dan beban mengajar. Hasil *clustering* juga memperlihatkan bahwa C1 memiliki keterwakilan OAP tinggi tetapi jumlah guru relatif kecil, C2 memiliki jumlah guru besar tetapi didominasi non-OAP, sedangkan C3 menjadi *cluster* mayoritas yang tersebar hampir di seluruh distrik dengan komposisi non-OAP yang masih dominan.

Jika dikaitkan dengan konteks wilayah, temuan ini perlu dibaca dalam karakteristik Kabupaten Nabire yang memiliki keragaman geografis, mulai dari wilayah perkotaan, pesisir, hingga wilayah dengan aksesibilitas terbatas. Distribusi guru tidak selalu dapat dilepaskan dari faktor lokasi, jarak tempuh, transportasi, sarana pendukung pendidikan, serta kondisi sosial dan keamanan. Oleh karena itu, ketimpangan distribusi guru ASN OAP dan non-OAP tidak sebaiknya dipahami hanya sebagai persoalan administratif, tetapi juga sebagai persoalan yang berkaitan dengan konteks wilayah dan keberlanjutan penempatan guru. Dengan demikian, hasil *clustering* dan klasifikasi dapat digunakan sebagai dasar awal bagi Dinas Pendidikan Kabupaten Nabire untuk mengidentifikasi satuan pendidikan yang membutuhkan evaluasi lebih lanjut, terutama pada kategori “Tidak ada OAP” dan “Rendah”, dengan tetap mempertimbangkan rasio guru-siswa, kebutuhan guru, aksesibilitas, dan kondisi wilayah.

3.5. Implikasi Penelitian

Implikasi teoretis dari penelitian ini adalah bahwa pendekatan data mining dapat digunakan untuk memperluas cara membaca persoalan distribusi guru, tidak hanya melalui statistik deskriptif, tetapi juga melalui pengelompokan pola dan klasifikasi kategori distribusi. Kombinasi K-Means dan algoritma klasifikasi memberikan kerangka analisis yang lebih sistematis untuk memahami keterwakilan guru ASN Orang Asli Papua dan non-Orang Asli Papua pada satuan pendidikan. Hasil penelitian menunjukkan bahwa pola distribusi guru dapat diidentifikasi melalui karakteristik kuantitatif seperti jumlah guru ASN OAP, jumlah guru ASN non-OAP, total guru ASN, serta persentase keterwakilan OAP. Dengan demikian, penelitian ini memperkuat pemanfaatan data *mining* dalam kajian manajemen pendidikan, khususnya untuk menganalisis pemerataan sumber daya manusia pendidikan di wilayah dengan karakteristik sosial dan geografis yang kompleks.

Implikasi praktis dari penelitian ini adalah bahwa hasil *clustering* dan klasifikasi dapat menjadi alat bantu awal bagi Dinas Pendidikan Kabupaten Nabire dalam mengevaluasi distribusi guru ASN OAP dan non-OAP. Satuan pendidikan yang berada pada kategori “Tidak ada OAP” dan “Rendah” dapat dijadikan prioritas untuk verifikasi lebih lanjut, terutama dengan mempertimbangkan rasio guru-siswa, jumlah rombongan belajar, kebutuhan guru mata pelajaran, aksesibilitas wilayah, dan kondisi keamanan. Namun, hasil kategori dalam penelitian ini tidak boleh dipahami sebagai ukuran tunggal kekurangan guru, melainkan sebagai indikator keterwakilan dan pemerataan sosial tenaga pendidik. Oleh karena itu, temuan ini dapat mendukung penyusunan kebijakan penempatan guru yang lebih berbasis data, kontekstual, dan sesuai dengan karakteristik wilayah Kabupaten Nabire.

4. KESIMPULAN

Berdasarkan hasil penelitian, distribusi guru ASN Orang Asli Papua dan non-Orang Asli Papua pada satuan pendidikan di Kabupaten Nabire menunjukkan pola yang belum proporsional. Dari total 2.147 guru ASN, sebanyak 364 guru atau 16,95% merupakan guru ASN OAP, sedangkan 1.783 guru atau 83,05% merupakan guru ASN non-OAP. Hasil *clustering* menunjukkan tiga pola utama, yaitu C1 sebagai cluster dengan keterwakilan OAP tinggi tetapi jumlah guru relatif kecil, C2 sebagai cluster dengan jumlah guru besar tetapi didominasi non-OAP, dan C3 sebagai cluster mayoritas yang tersebar di hampir seluruh distrik dengan komposisi non-OAP yang masih dominan. Pada tahap klasifikasi, algoritma Decision Tree menjadi model terbaik dengan nilai AUC 0,994, *Classification Accuracy* 0,992, *F1-Score* 0,992, *Precision* 0,992, dan *Recall* 0,992, serta hanya menghasilkan 2 kesalahan klasifikasi dari 256 data. Temuan ini menunjukkan bahwa kombinasi *clustering* dan klasifikasi dapat digunakan sebagai alat bantu analisis awal untuk membaca pola distribusi guru ASN OAP dan non-OAP secara lebih terukur.

Berdasarkan temuan tersebut, Dinas Pendidikan Kabupaten Nabire disarankan menjadikan satuan pendidikan pada kategori “Tidak ada OAP” dan “Rendah” sebagai prioritas awal dalam evaluasi keterwakilan

guru ASN OAP. Namun, kategori tersebut tidak boleh langsung dimaknai sebagai kekurangan guru secara absolut, karena penentuan kebutuhan guru tetap perlu divalidasi dengan indikator fundamental seperti rasio guru-siswa, jumlah rombongan belajar, jenjang pendidikan, kebutuhan guru mata pelajaran, serta beban mengajar. Selain itu, evaluasi distribusi guru juga perlu mempertimbangkan aksesibilitas wilayah, jarak tempuh, kondisi keamanan, dan sarana pendidikan. Oleh karena itu, penelitian selanjutnya disarankan mengintegrasikan variabel rasio guru-siswa, rombongan belajar, jumlah siswa, kebutuhan guru per mata pelajaran, dan faktor geografis agar analisis distribusi guru tidak hanya menggambarkan keterwakilan OAP dan non-OAP, tetapi juga kebutuhan riil satuan pendidikan.

REFERENSI

- [1] J. Hodges, R. U. Mun, and A. Rinn, "Disentangling Inequity in Gifted Education: The Need for Nuance in Racial/Ethnic Categories, Socioeconomic Status, and Geography," *Gift. Child Q.*, vol. 66, no. 2, pp. 154–156, 2022, doi: 10.1177/00169862211040533.
- [2] Y. Han, R. Ni, and J. Gao, "Regional Inequality of Higher Education Development in China: Comprehensive Evaluation and Geographical Representation," *Sustainability*, vol. 15, no. 3, p. 1824, 2023, doi: 10.3390/su15031824.
- [3] P. Pholphirul, P. Rukumnuaykit, and S. Teimrad, "Teacher shortages and educational outcomes in developing countries: Empirical evidence from PISA-Thailand," *Cogent Educ.*, vol. 10, 2023, doi: 10.1080/2331186x.2023.2243126.
- [4] "Teacher allocation and learning divides," *Teach. Learn. Int. Surv.*, 2022, doi: 10.1787/1a02af6b-en.
- [5] A. I. Poli, F. Sokoy, A. Kadir, A. Ali, M. Suhun, and A. I. Poli, "Pendidikan Berbasis Budaya Lokal: Pemetaan Kurikulum di Sekolah Dasar Melalui Kolaborasi Guru dan Komunitas Adat," *Abdimas Univers.*, vol. 7, no. 2, pp. 429–436, 2025, doi: 10.36277/abdimasuniversal.v7i2.2711.
- [6] W. Li and P. P. Shein, "Developing sense of place through a place-based Indigenous education for sustainable development curriculum," *Environ. Educ. Res.*, vol. 29, no. 5, pp. 692–714, 2022, doi: 10.1080/13504622.2022.2098933.
- [7] U. Arfan and Y. Pekci, "A Evaluasi Distribusi Guru-Siswa dan Ketersediaan Sekolah untuk Mendukung Pembangunan Pendidikan Menggunakan K-Means Clustering: An Evaluation of Teacher–Student Distribution and School Availability in Supporting Educational Development Using the K-Means," *Inst. Ris. dan Publ. Indones.*, vol. Vol. 5 No., 2025.
- [8] A. Mbabazi, T. Muhammad, Z. L. Bagiwa, and L. Aja, "Enhancing Teacher Performance and Motivation through Effective Appraisal Policies: Evidence-Based Recommendations for Improving Educational Outcomes in Secondary Schools in Uganda," *F1000Research*, vol. 14, p. 826, 2025, doi: 10.12688/f1000research.168834.1.
- [9] Y. Song and C.-C. Hsieh, "Distributed leadership, teacher autonomy, and well-being: A multilevel analysis of novice teachers' teaching practice quality in Taiwan using TALIS 2018 data," *Educ. Manag. Adm. Leadersh.*, 2026, doi: 10.1177/17411432251410038.
- [10] L. Xiao, "Research on the Path of Improving the Quality of School Physical Education Teaching Based on Data Mining Technology," *Appl. Math. nonlinear Sci.*, vol. 9, no. 1, 2024, doi: 10.2478/amns-2024-1040.
- [11] K. Maquiling, "Analyzing Teaching Performance of Instructors Using Data Mining Techniques," *Int. J. Sci. Acad. Res.*, vol. 05, no. 06, pp. 1–9, 2025, doi: 10.54756/ijisar.2025.6.1.
- [12] F. H. Kuwil and Ü. Atila, "A Novel Hybridization Approach to Improve the Critical Distance Clustering Algorithm: Balancing Speed and Quality," *Expert Syst. Appl.*, 2024, doi: 10.1016/j.eswa.2024.123298.
- [13] Republik Indonesia, "Undang-Undang Republik Indonesia Nomor 21 Tahun 2001 tentang Otonomi Khusus bagi Provinsi Papua," JDIH Kementerian Hukum. [Online]. Available: <https://jdih.go.id/pencarian/detail/1325515/index.html>
- [14] Republik Indonesia, "Undang-Undang Republik Indonesia Nomor 2 Tahun 2021 tentang Perubahan Kedua atas Undang-Undang Nomor 21 Tahun 2001 tentang Otonomi Khusus bagi Provinsi Papua," JDIH Kementerian Hukum. [Online]. Available: <https://jdih.go.id/pencarian/detail/1409048/undang-undang-nomor-2-tahun-2021-tentang-perubahan-kedua-atas-undang-undang-nomor-21-tahun-2001-tentang-otonomi-khusus-bagi-provinsi-papua>
- [15] Republik Indonesia, "Peraturan Pemerintah Republik Indonesia Nomor 106 Tahun 2021 tentang Kewenangan dan Kelembagaan Pelaksanaan Kebijakan Otonomi Khusus Provinsi Papua," JDIH Kementerian Hukum. [Online]. Available: <https://jdih.kemendagri.go.id/dokumen/view?id=1136>
- [16] Pemerintah Provinsi Papua, "Peraturan Daerah Khusus Provinsi Papua Nomor 3 Tahun 2013 tentang Pelayanan Pendidikan bagi Komunitas Adat Terpencil," JDIH Kementerian Hukum. [Online]. Available: <https://jdih.papua.go.id/dokumen/view?id=89>
- [17] M. Schonlau, "Nearest Neighbors," *Springer Nature*, 2023, pp. 119–141. doi: 10.1007/978-3-031-33390-3_7.
- [18] U. Arfan and N. Paraga, "Perbandingan Algoritma K-Means, Naïve Bayes dan Decision Tree Dalam

- Memprediksi Penjualan Bahan Bakar Minyak: The Comparison of K-Means, Naïve Bayes and Decision Tree Algorithm in Predicting Fuel Oil Sales,” *Inst. Ris. dan Publ. Indones.*, vol. Vol. 4 No., 2024.
- [19] I. Journal, “Data Science: Data Visualization and Data Analytics in the Process of Data Mining,” *Indian Sci. J. Res. Eng. Manag.*, 2024, doi: 10.55041/ijrem28332.
- [20] B. G. Shinde and S. K. Shivhare, “Impact Of Data Visualization In Data Analysis To Improve The Efficiency Of Machine Learning Models,” *J. Adv. Zool.*, pp. 107–112, 2024, doi: 10.53555/jaz.v45is4.4161.
- [21] K. Karthi, “Heart Attack Prediction using KNN, Naive Bayes and decision tree,” 2025, doi: 10.5281/zenodo.15854384.
- [22] Z. Zulkarnain, R. Mutia, J. A. Ariani, Z. A. Barik, and H. Azmi, “Performance Comparison K-Nearest Neighbor, Naive Bayes, and Decision Tree Algorithms for Netflix Rating Classification,” *IJATIS Indones. J. Appl. Technol. Innov. Sci.*, vol. 1, no. 1, pp. 16–22, 2024, doi: 10.57152/ijatis.v1i1.1104.
- [23] W. Junthopas, C. Wongoutong, W. Junthopas, and C. Wongoutong, “Pre-Determining the Optimal Number of Clusters for k-Means Clustering Using the Parameters Package in R and Distance Metrics,” *Appl. Sci.*, vol. 15, no. 21, p. 11372, 2025, doi: https://10.3390/app152111372.
- [24] 李自刚, “Improvement Methods for Determining the Value of k in the K-Means Clustering Algorithm,” *Shu ju wa jue*, vol. 14, no. 03, pp. 143–148, 2024, doi: 10.12677/hjdm.2024.143013.
- [25] G. Endress, “Optimal Value for Number of Clusters in a Dataset for Clustering Algorithm,” in *Artificial Intelligence and Sustainable Computing: Proceedings of ICSISCET 2021*, 2022, pp. 631–645. doi: 10.1007/978-981-19-1653-3_48.
- [26] L. Luo and T. Liu, “Integrating Advanced Principal Component Analysis Into Naive Bayes For Enhanced Classification Performance,” *Deleted J.*, vol. 3, no. 1, pp. 27–31, 2024, doi: 10.54254/3029-0880/3/2024019.
- [27] D. Ortiz-Villaseñor, G. Trujillo-Hernández, O. Real-Moreno, M. J. Castro-Toscano, L. D. Medina-Madrazo, and D. Barrera-Román, “K-Nearest Neighbors Regression and Applications,” *IGI Glob.*, pp. 295–316, 2025, doi: 10.4018/979-8-3693-6910-4.ch015.
- [28] F. Acito, “k Nearest Neighbors,” in *Predictive Analytics with KNIME: Analytics for Citizen Data Scientists*, 2023, pp. 209–227. doi: 10.1007/978-3-031-45630-5_10.
- [29] M. Sivananda and G. K. Kumar, “Classification and Regression Based on Decision Tree Algorithm for Machine Learning,” *Indian Sci. J. Res. Eng. Manag.*, 2024, doi: 10.55041/ijrem28533.