Buletin Ilmiah
Ilmu Komputer dan Multimedia

BIIKMA : Buletin Ilmiah Ilmu Komputer dan Multimedia
Volume 2, No. 5, Januari 2025
ISSN 3024-8248 (media online)
Hal 899-903

Perbandingan Dan Penerapan Penggunaan Dataset IRIS

Fathih Sulthoni Nabhan[1*], Muhammad Farhan[2], Raihan Akbar Syamputra[3], Syelvi Naeska Fahira[4]

[1-4]Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspiptek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia
Email: [1*]sayanabhannn@gmail.com, [2]mmdfarhann@gmail.com,
[3]raihanakbarsyamputra@gmail.com, [4]syelvifahira@gmail.com
(* : coressponding author)

Abstrak Penelitian ini mengkaji berbagai metode analisis untuk dataset Iris menggunakan beberapa pendekatan yang berbeda. Perbandingan pertama dilakukan antara algoritma K-Nearest Neighbor (KNN) dan Random Forest (RF) dengan pembagian data 80:20, di mana RF menunjukkan performa superior dengan akurasi 100% dibandingkan KNN yang mencapai 97%. Selanjutnya, implementasi metode Fuzzy C-Means berhasil melakukan pengklusteran dataset berdasarkan karakteristik fisik bunga Iris dengan mempertimbangkan derajat keanggotaan untuk menangani ketidakpastian data. Studi juga membandingkan performa tools data mining, yaitu Rapid Miner dan Orange, dalam mengklasifikasi dataset Iris menggunakan decision tree, dengan Rapid Miner mencapai akurasi 86.67% dan nilai kappa 0.800. Hasil-hasil ini memberikan wawasan komprehensif tentang efektivitas berbagai metode dan tools dalam analisis dataset Iris, yang dapat membantu peneliti dan praktisi dalam memilih pendekatan yang tepat untuk analisis data mereka.

Kata Kunci: Akurasi, Klasifikasi, Pohon Keputusan, Rapid Miner, Orange, Dataset Bunga Iris

Abstract− This research examines various analysis methods for the Iris dataset using several different approaches. The first comparison was made between the K-Nearest Neighbor (KNN) and Random Forest (RF) algorithms with an 80:20 data split, where RF showed superior performance with 100% accuracy compared to KNN which reached 97%. Furthermore, the implementation of Fuzzy C-Means method successfully clustered the dataset based on the physical characteristics of Iris flowers by considering the membership degree to handle data uncertainty. The study also compared the performance of data mining tools, namely Rapid Miner and Orange, in classifying the Iris dataset using decision tree, with Rapid Miner achieving 86.67% accuracy and a kappa value of 0.800. These results provide a comprehensive insight into the effectiveness of various methods and tools in the analysis of Iris datasets, which can help researchers and practitioners in choosing the right approach for their data analysis.

Keywords: Accuracy, Classification, Decision Tree, Rapid Miner, Orange, Envious Flower Dataset

1. PENDAHULUAN

Ketiga jurnal yang direview membahas pemanfaatan dataset Iris sebagai alat untuk menguji berbagai metode dalam data mining, khususnya klasifikasi dan clustering. Dataset Iris dipilih karena sederhana, mudah diakses, dan memiliki atribut yang relevan seperti panjang sepal, lebar sepal, panjang kelopak, dan lebar kelopak, dengan target klasifikasi berupa tiga kelas bunga: iris setosa, iris versicolour, dan iris virginica. Jurnal pertama membahas perbandingan algoritma Decision Tree menggunakan aplikasi RapidMiner dan Orange, menunjukkan bahwa Decision Tree adalah metode yang efektif untuk klasifikasi dataset Iris dengan hasil visualisasi dan akurasi yang bervariasi pada masing-masing aplikasi.

Jurnal kedua membahas pengelompokan data Iris menggunakan metode Fuzzy C-Means, yang menawarkan fleksibilitas melalui logika fuzzy dalam menangani data dengan ketidakpastian, menghasilkan kluster yang lebih akurat dibandingkan metode konvensional. Jurnal ketiga membandingkan performa algoritma K-Nearest Neighbor (KNN) dan Random Forest (RF) dalam klasifikasi dataset Iris, menunjukkan bahwa RF memiliki keunggulan dalam akurasi sempurna (100%) dibandingkan KNN yang memiliki akurasi 97%. Ketiga jurnal ini menyoroti pentingnya penggunaan data mining untuk menangani data berskala besar dengan performa yang baik, serta relevansi dataset Iris sebagai alat untuk menguji keandalan berbagai algoritma dalam analisis data.

Data mining menjadi metode yang efektif untuk menganalisis dataset berskala besar, dengan K-Nearest Neighbor (KNN) dan Random Forest (RF) sebagai dua algoritma yang menunjukkan performa unggul. Dataset Iris, yang berisi informasi karakteristik bunga seperti panjang sepal, lebar sepal, panjang kelopak, dan lebar kelopak, sering digunakan sebagai benchmark dalam pengujian

Buletin Ilmiah
Ilmu Komputer dan Multimedia
BIIKMA : Buletin Ilmiah Ilmu Komputer dan Multimedia
Volume 2, No. 5, Januari 2025
ISSN 3024-8248 (media online)
Hal 899-903

metode klasifikasi dan clustering karena kesederhanaan dan aksesibilitasnya. Penelitian menunjukkan bahwa RF mencapai akurasi 100% dibandingkan dengan KNN yang mencapai 97% dalam klasifikasi dataset Iris dengan pembagian data 80:20.

Dalam konteks Indonesia yang kaya akan keanekaragaman hayati, namun baru sekitar 20% dari total tanaman yang teridentifikasi, teknik clustering menjadi sangat relevan untuk mengklasifikasikan spesies tanaman yang belum teridentifikasi. Metode Fuzzy C-Means yang menggunakan konsep logika fuzzy terbukti efektif dalam pengelompokan dataset Iris, memberikan pendekatan yang lebih fleksibel dibandingkan dengan logika klasik karena menggunakan rentang keanggotaan antara 0 dan 1.

Perbandingan performa tools data mining seperti Rapid Miner dan Orange dalam klasifikasi dataset Iris menggunakan decision tree menunjukkan hasil yang menjanjikan, dengan Rapid Miner mencapai akurasi 86.67% dan nilai kappa 0.800. Hal ini membuktikan bahwa pemilihan tools dan metode yang tepat sangat penting dalam proses klasifikasi dan clustering data. Hasil-hasil ini memberikan wawasan berharga bagi peneliti dan praktisi dalam memilih pendekatan yang sesuai untuk analisis data tanaman, khususnya dalam konteks identifikasi dan klasifikasi spesies tanaman di Indonesia.

2. METODE PENELITIAN

Ketiga jurnal yang direview membahas pemanfaatan dataset Iris untuk menguji berbagai metode data mining, baik klasifikasi maupun clustering, dengan fokus pada algoritma dan tools yang digunakan. Jurnal pertama membandingkan performa algoritma Decision Tree pada aplikasi Rapid Miner dan Orange, menunjukkan bahwa Rapid Miner menghasilkan akurasi sebesar 86,67% dengan nilai kappa 0,800, sedangkan Orange memberikan hasil klasifikasi yang terperinci berdasarkan atribut data. Jurnal kedua mengulas penggunaan metode Fuzzy C-Means dalam pengelompokan dataset Iris, yang memanfaatkan logika fuzzy untuk menghasilkan kluster lebih fleksibel dan akurat, terutama dalam menangani ketidakpastian dan ambiguitas data. Sementara itu, jurnal ketiga membandingkan algoritma K-Nearest Neighbor (KNN) dan Random Forest (RF), dengan hasil menunjukkan RF lebih unggul dengan akurasi 100%, dibandingkan KNN yang mencapai 97% pada pembagian data 80:20. Ketiga jurnal ini memberikan gambaran penting tentang relevansi dataset Iris sebagai benchmark, sekaligus menggarisbawahi pentingnya pemilihan algoritma dan tools yang tepat untuk meningkatkan akurasi dan fleksibilitas dalam proses analisis data, terutama dalam konteks identifikasi spesies tanaman seperti yang relevan di Indonesia.

2.1 Desain Penelitian

Desain penelitian ini mengadopsi pendekatan deskriptif kualitatif dan kuantitatif , yang bertujuan untuk menggali lebih dalam penerapan machine learning dalam industri transportasi Indonesia. Pendekatan deskriptif kualitatif digunakan untuk menganalisis persepsi, pengalaman, dan feedback dari pengguna layanan transportasi berbasis online yang dikumpulkan dari media sosial seperti Twitter dan Instagram. Sementara itu, pendekatan kuantitatif digunakan untuk menganalisis data numerik yang terkait dengan operasional perusahaan transportasi seperti data perjalanan, waktu kedatangan, keterlambatan, dan permintaan layanan. Dengan menggabungkan kedua pendekatan ini, penelitian ini bertujuan untuk memberikan gambaran yang lebih komprehensif mengenai penerapan machine learning dalam meningkatkan kinerja dan efisiensi dalam industri transportasi.

a. **Pendekatan Kualitatif:** Pada bagian ini, penelitian berfokus pada analisis teks dan umpan balik yang diterima dari pengguna melalui platform media sosial dan portal berita. Analisis sentimen dilakukan untuk memahami persepsi pengguna terhadap layanan yang diberikan oleh perusahaan transportasi, seperti Gojek, Grab, dan Blue Bird. Data yang diperoleh digunakan untuk mengkategorikan sentimen menjadi tiga kelas utama: positif, negatif, dan netral.
b. **Pendekatan Kuantitatif:** Penelitian ini juga menggunakan data operasional seperti waktu kedatangan kendaraan, durasi perjalanan, dan tingkat keterlambatan untuk melakukan prediksi permintaan dan mengoptimalkan rute perjalanan. Teknik pemodelan statistik dan machine learning diterapkan untuk memproyeksikan

Buletin Ilmiah
Ilmu Komputer dan Multimedia
BIIKMA : Buletin Ilmiah Ilmu Komputer dan Multimedia
Volume 2, No. 5, Januari 2025
ISSN 3024-8248 (media online)
Hal 899-903

permintaan masa depan dan menganalisis faktor-faktor yang berpengaruh terhadap keterlambatan serta efisiensi operasional.

2.2 Pengumpulan data

Pengumpulan data merupakan tahap krusial dalam penelitian ini. Data yang digunakan dalam penelitian ini bersumber dari berbagai platform yang relevan dengan topik penelitian, yaitu perusahaan transportasi, media sosial, dan portal berita online. Pengumpulan data dilakukan dengan menggunakan data sekunder yang diperoleh melalui beberapa teknik sebagai berikut:

a. Analisis Literatur Melibatkan tinjauan literatur untuk memahami kerangka konseptual dan temuan penelitian sebelumnya terkait dampak kenaikan BBM pada biaya transportasi.
b. Analisis SWOT: Melibatkan Analisis SWOT (Strenght, Weakness, Opportunity, Threat) untuk merumuskan strategi yang perlu diimplementasikan oleh pelaku bisnis yang diharapkan memberikan wawasan dalam penentuan strategi yang efektif

Proses pengumpulan data ini bertujuan untuk mendapatkan data yang lengkap dan beragam, yang akan digunakan dalam analisis lebih lanjut.

Ketiga jurnal yang direview mengkaji pemanfaatan dataset Iris, salah satu dataset standar yang sering digunakan dalam pengujian algoritma data mining untuk klasifikasi dan clustering. Dataset ini terdiri dari 150 sampel bunga dengan empat atribut utama, yaitu panjang sepal, lebar sepal, panjang kelopak, dan lebar kelopak, serta tiga kelas target: Iris Setosa, Iris Versicolor, dan Iris Virginica. Jurnal pertama membandingkan performa algoritma Decision Tree menggunakan dua aplikasi, RapidMiner dan Orange, untuk klasifikasi dataset Iris. Hasilnya, RapidMiner menunjukkan akurasi sebesar 86,67% dengan nilai kappa 0,800, sementara Orange menampilkan hasil visualisasi yang lebih terperinci berdasarkan atribut data. Jurnal kedua membahas pengelompokan dataset Iris menggunakan metode Fuzzy C-Means, yang berbasis pada logika fuzzy. Logika fuzzy memberikan fleksibilitas dalam menangani ketidakpastian melalui derajat keanggotaan yang bernilai antara 0 hingga 1, sehingga metode ini mampu menghasilkan kluster yang lebih akurat dibandingkan dengan metode konvensional. Sementara itu, jurnal ketiga membandingkan algoritma K-Nearest Neighbor (KNN) dan Random Forest (RF) untuk klasifikasi dataset Iris. Hasil menunjukkan bahwa RF unggul dengan akurasi sempurna (100%) dan F1-Score 1,00, sedangkan KNN mencapai akurasi 97% dengan F1-Score 0,98. RF menunjukkan kemampuan yang lebih baik dalam membedakan ketiga spesies bunga tanpa kesalahan, sementara KNN masih menunjukkan kesalahan kecil pada beberapa data. Secara keseluruhan, ketiga jurnal ini menyoroti relevansi dataset Iris sebagai alat pengujian standar dalam data mining dan pentingnya pemilihan algoritma atau tools yang sesuai. Hasil penelitian memberikan wawasan tentang performa algoritma dalam mengklasifikasikan atau mengelompokkan data, serta menegaskan nilai lebih metode seperti logika fuzzy dalam menangani ketidakpastian data, yang relevan untuk aplikasi luas, termasuk identifikasi dan klasifikasi spesies tanaman di Indonesia.

3. ANALISA DAN PEMBAHASAN

Dalam penelitian ini, penerapan machine learning dalam industri transportasi di Indonesia telah memberikan wawasan yang signifikan tentang bagaimana teknologi ini dapat membantu meningkatkan efisiensi operasional dan memberikan pengalaman pengguna yang lebih baik. Berdasarkan hasil analisis data yang diperoleh dari berbagai sumber, beberapa temuan utama terkait penerapan machine learning di industri transportasi online di Indonesia adalah sebagai berikut. Ketiga jurnal yang direview membahas performa algoritma dalam klasifikasi dan clustering dataset Iris, yang terdiri dari 150 sampel bunga dengan atribut utama berupa panjang sepal, lebar sepal, panjang kelopak, dan lebar kelopak.

a. **Jurnal Pertama** menunjukkan bahwa algoritma Decision Tree merupakan metode yang efektif untuk klasifikasi, dengan hasil akurasi menggunakan RapidMiner sebesar 86,67%, nilai kappa 0,800, dan weighted mean precision 90,48%. RapidMiner unggul dalam efisiensi, sementara Orange memberikan visualisasi hasil klasifikasi yang lebih terperinci.

Buletin Ilmiah
Ilmu Komputer dan Multimedia
BIIKMA : Buletin Ilmiah Ilmu Komputer dan Multimedia
Volume 2, No. 5, Januari 2025
ISSN 3024-8248 (media online)
Hal 899-903

b. **Jurnal Kedua** membahas penerapan metode Fuzzy C-Means untuk clustering data Iris, memanfaatkan logika fuzzy yang memungkinkan adanya derajat keanggotaan dalam rentang 0 hingga 1. Metode ini menghasilkan kluster yang lebih fleksibel dan akurat dibandingkan pendekatan konvensional, menjadikannya ideal untuk menangani dataset dengan ambiguitas tinggi.
c. **Jurnal Ketiga** membandingkan algoritma K-Nearest Neighbor (KNN) dan Random Forest (RF) dalam klasifikasi dataset Iris, di mana RF menunjukkan performa unggul dengan akurasi 100% dan F1-Score sempurna 1,00, sedangkan KNN mencatat akurasi 97% dan F1-Score 0,98. Perbedaan ini menegaskan keunggulan RF dalam menangani data yang lebih kompleks, terutama dalam membedakan spesies seperti Iris Versicolor dan Iris Virginica.

Secara keseluruhan, ketiga jurnal ini memberikan wawasan penting mengenai keandalan algoritma dan tools dalam mengolah dataset Iris, dengan fokus pada akurasi, fleksibilitas, dan visualisasi hasil yang relevan untuk berbagai kebutuhan penelitian dan aplikasi.

3.1 Analisis Sentimen untuk Menilai Pengalaman Pengguna

Jurnal-jurnal yang direview menyoroti pentingnya analisis sentimen sebagai alat untuk memahami pengalaman pengguna melalui pengolahan data berbasis algoritma data mining. Dalam berbagai penelitian, algoritma seperti **Support Vector Machine (SVM)**, **Random Forest (RF)**, dan **Fuzzy C-Means** telah diaplikasikan pada dataset Iris sebagai benchmark. Algoritma ini menunjukkan kemampuan yang baik dalam menganalisis pola data dan memberikan hasil klasifikasi atau clustering yang akurat.

3.1.1 Tantangan dan Peluang dalam Penerapan analisis data iris

Tantangan utama dalam analisis data adalah kebutuhan untuk menangani data berskala besar dan kompleksitas dalam mengklasifikasikan data dengan atribut yang saling terkait. Namun, jurnal menunjukkan bahwa peluang besar ada dalam pengembangan algoritma yang lebih efisien dan pemanfaatan alat bantu seperti **Rapid Miner**, **Orange**, dan kerangka berbasis logika fuzzy. Tantangan lainnya termasuk optimasi parameter algoritma seperti konstanta K dalam KNN atau hyperparameter pada SVM. Peluang besar terletak pada kemampuan algoritma ini untuk diterapkan pada berbagai domain lain, seperti klasifikasi tanaman lokal di Indonesia, dengan potensi membantu identifikasi spesies yang belum teridentifikasi.

3.1.2 Implikasi terhadap Pengembangan Industri Transportasi Di Indonesia

Implikasi dari studi ini mencakup penggunaan metode analisis sentimen untuk meningkatkan pengalaman pengguna dalam berbagai konteks, seperti pengenalan pola, pengelompokan data tanaman, dan klasifikasi dataset dalam skala besar. Sebagai contoh, hasil dari metode SVM dan Random Forest dapat diterapkan untuk mengidentifikasi spesies tanaman atau bunga yang belum teridentifikasi, yang relevan dengan keanekaragaman hayati di Indonesia.

3.1.3 Saran untuk Penerapan Machine Learning di Masa Depan

Implikasi dari studi ini mencakup penggunaan metode analisis sentimen untuk meningkatkan pengalaman pengguna dalam berbagai konteks, seperti pengenalan pola, pengelompokan data tanaman, dan klasifikasi dataset dalam skala besar. Sebagai contoh, hasil dari metode SVM dan Random Forest dapat diterapkan untuk mengidentifikasi spesies tanaman atau bunga yang belum teridentifikasi, yang relevan dengan keanekaragaman hayati di Indonesia.

4. IMPLEMENTASI

Keempat jurnal ini memanfaatkan dataset Iris, yang terdiri dari atribut sepal length, sepal width, petal length, dan petal width, serta tiga kelas (Iris Setosa, Iris Versicolor, dan Iris Virginica), untuk mengevaluasi berbagai algoritma data mining.

1. **Jurnal pertama** mengimplementasikan algoritma *Decision Tree* menggunakan aplikasi Rapid Miner dan Orange. Dengan pembagian data 90% untuk training dan 10%

BIIKMA : Buletin Ilmiah Ilmu Komputer dan Multimedia
Volume 2, No. 5, Januari 2025
ISSN 3024-8248 (media online)
Hal 899-903

untuk testing, Rapid Miner menghasilkan akurasi 86,67% dengan nilai kappa 0,800 dan weighted mean precision 90,48%.

2. **Jurnal kedua** memanfaatkan algoritma *Fuzzy C-Means* yang berbasis logika fuzzy untuk clustering dataset Iris. Dengan logika fuzzy, data dianalisis dengan rentang keanggotaan 0 hingga 1, menghasilkan kluster yang lebih fleksibel dan mampu menangani ambiguitas data.
3. **Jurnal ketiga** membandingkan algoritma *K-Nearest Neighbor (KNN)* dan *Random Forest (RF)*, dengan RF menunjukkan akurasi sempurna (100%) dan F1-Score 1,00, dibandingkan dengan KNN yang mencapai akurasi 97%.
4. **Jurnal keempat** menggunakan metode *Support Vector Machine (SVM)* dengan kernel polynomial. SVM diuji dengan dua metode pelatihan: *percentage split* (80:20) menghasilkan akurasi 96,7%, dan *k-fold cross-validation*(k=10) menghasilkan akurasi 92,6%.

5. KESIMPULAN

Keempat jurnal memberikan wawasan tentang keandalan berbagai metode data mining dalam mengolah dataset Iris:

1. *Decision Tree* efektif untuk klasifikasi, dengan Rapid Miner dan Orange memberikan hasil visualisasi yang berbeda tetapi bermanfaat.
2. *Fuzzy C-Means* menunjukkan keunggulan dalam clustering dengan kemampuan menangani data ambigu menggunakan logika fuzzy.
3. *Random Forest* lebih unggul dibandingkan *KNN* dalam klasifikasi, terutama dalam menangani data yang lebih kompleks.
4. *Support Vector Machine* dengan kernel polynomial dan metode pelatihan *percentage split* menghasilkan akurasi yang sangat tinggi, menjadikannya salah satu metode terbaik untuk klasifikasi dataset Iris.

REFERENCES

JNATIA Volume 2, Nomor 2, Februari 2024 Jurnal Nasional Teknologi Informasi dan Aplikasinya. Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat 2021 Institut Informatika dan Bisnis Darmajaya, 19 Agustus 2021 Vol. 1 No. (1) (Juli 2023) 19-26 Journal of Data Insights http://journalnew.unimus.ac.id/index.php/jodi