

**ANALISIS PROYEK DATA GAJI PEGAWAI BIAYA OPERASIONAL
DI SWAN CAFFEE & RESTO RANTAUPRAPAT
MENGUNAKAN METODE KNN**

Hadron Sofyan¹⁾

Fakultas Sains dan Teknologi, Universitas Labuhanbatu

e-mail: hadronsofyan@gmail.com

Masrizal²⁾

Fakultas Sains dan Teknologi, Universitas Labuhanbatu

e-mail: masrizal120405@gmail.com

Sudi Suryadi³⁾

Fakultas Sains dan Teknologi, Universitas Labuhanbatu

e-mail: sudisuryadi28@gmail.com

Budianto Bangun⁴⁾

Fakultas Sains dan Teknologi, Universitas Labuhanbatu

e-mail: budiantobangun44@gmail.com

Correspondence Author: hadronsofyan@gmail.com

ABSTRAK

Pemanfaatan data operasional sebagai dasar pengambilan keputusan penting dalam pengelolaan usaha kuliner. Data gaji pegawai dan biaya operasional yang tercatat secara rutin memiliki potensi untuk dianalisis guna mengidentifikasi pola pengeluaran dan mendukung efisiensi operasional. Namun, data tersebut umumnya masih digunakan sebatas laporan administratif. Penelitian ini bertujuan untuk menganalisis dan mengklasifikasikan data gaji pegawai serta biaya operasional pada Swan Caffee & Resto Rantauprapat menggunakan pendekatan *Knowledge Discovery in Databases* (KDD) dengan algoritma *K-Nearest Neighbor* (KNN). Penelitian menggunakan pendekatan kuantitatif dengan desain studi kasus. Data dianalisis melalui tahapan pembersihan, transformasi, dan klasifikasi menggunakan KNN. Evaluasi model dilakukan dengan perangkat lunak Orange menggunakan skema *random sampling* dengan data pelatihan 70%. Hasil penelitian menunjukkan nilai akurasi sebesar 53,3% dan AUC sebesar 0,407, yang mengindikasikan keterbatasan model akibat ketidakseimbangan kelas. Meskipun demikian, KNN dapat dimanfaatkan sebagai alat eksploratif dalam mendukung pengambilan keputusan berbasis data pada usaha kuliner.

Kata kunci: Data Mining, KNN, KDD, Gaji Pegawai, Biaya Operasional, usaha kuliner.

ABSTRACT

The use of operational data as a basis for decision making is essential in culinary business management. Employee salary and operational cost data, which are routinely recorded, have the potential to be analyzed to identify expenditure patterns and improve operational efficiency. However, such data are often limited to administrative reporting. This study aims to analyze and classify employee salary and operational cost data at Swan Caffe & Resto Rantauprapat using the Knowledge Discovery in Databases (KDD) approach with the K-Nearest Neighbor (KNN) algorithm. This research employs a quantitative approach with a case study design. Data were processed through data cleaning, transformation, and classification stages using the KNN algorithm. Model evaluation was conducted using the Orange software with a random sampling scheme and a 70% training data proportion. The results show an accuracy of 53.3% and an AUC value of 0.407, indicating limitations due to class imbalance. Nevertheless, KNN can be utilized as an exploratory tool to support data-driven decision making in culinary businesses.

Keywords: *Data Mining, K-Nearest Neighbor, KDD, Employee Salary, Operational Cost*

1. PENDAHULUAN

Pemanfaatan data sebagai dasar pengambilan keputusan menjadi aspek penting dalam pengelolaan usaha di era digital. Data operasional yang dihasilkan secara rutin tidak hanya berfungsi sebagai catatan administratif, tetapi juga dapat diolah untuk menghasilkan informasi yang bernilai strategis. Pada usaha kuliner, data gaji pegawai dan biaya operasional merupakan komponen utama yang memengaruhi efisiensi operasional dan stabilitas keuangan usaha.

Namun demikian, ketersediaan data operasional tersebut belum selalu diikuti dengan pemanfaatan analisis yang optimal. Data gaji pegawai dan biaya operasional umumnya disimpan dalam bentuk laporan periodik tanpa dilakukan pengolahan lebih lanjut untuk mengidentifikasi pola, hubungan, atau kecenderungan tertentu antarperiode. Akibatnya, informasi yang dihasilkan dari data tersebut belum sepenuhnya mendukung

proses evaluasi dan perencanaan keuangan secara objektif.

Pendekatan Data Mining melalui kerangka *Knowledge Discovery in Databases* (KDD) dapat digunakan untuk menjawab permasalahan tersebut. KDD memungkinkan proses pengolahan data dilakukan secara sistematis melalui tahapan seleksi, pembersihan, transformasi, dan analisis data. Salah satu algoritma yang dapat diterapkan dalam proses ini adalah *K-Nearest Neighbor* (KNN), yang mengelompokkan data berdasarkan tingkat kemiripan dengan menghitung jarak antar data.

Penelitian ini bertujuan untuk menerapkan metode KNN dalam menganalisis dan mengklasifikasikan data gaji pegawai serta biaya operasional di Swan Caffe & Resto Rantauprapat. Melalui penerapan tahapan KDD, penelitian ini diharapkan mampu mengidentifikasi pola hubungan antar data operasional yang dapat

dimanfaatkan sebagai dasar pendukung pengambilan keputusan manajerial.

Kontribusi penelitian ini terletak pada penerapan metode KNN dalam konteks analisis data operasional usaha kuliner, yang masih relatif terbatas pada penelitian sebelumnya. Hasil penelitian diharapkan dapat memberikan kontribusi akademik dalam pengembangan penerapan Data Mining, serta manfaat praktis bagi pengelolaan keuangan usaha kecil dan menengah.

2. LANDASAN TEORI

Pemanfaatan data operasional sebagai dasar pengambilan keputusan berbasis sistem informasi berlandaskan pada konsep Knowledge Discovery in Databases (KDD). KDD merupakan kerangka kerja sistematis yang digunakan untuk mengekstraksi pengetahuan dari basis data melalui tahapan seleksi data, data cleaning, transformasi, penerapan algoritma data mining, serta evaluasi hasil (Hidayat et al., 2021); (Alfandi et al., 2024). Dalam konteks sistem informasi bisnis, KDD berfungsi sebagai pendekatan konseptual untuk mengubah data mentah menjadi informasi yang bernilai guna mendukung keputusan manajerial.

Pada tahap inti KDD, Data Mining berperan sebagai mekanisme analisis untuk menemukan pola, hubungan, dan struktur tersembunyi dalam dataset berukuran besar. Data Mining memanfaatkan teknik statistik, komputasi, dan machine learning untuk melakukan analisis deskriptif maupun prediktif (Nirwana & Hayati, 2021); (Fau, 2024). Salah satu fungsi utama Data Mining adalah klasifikasi, yaitu proses pengelompokan data ke dalam kelas tertentu berdasarkan karakteristik dan atribut yang

dimiliki, sehingga relevan untuk analisis data keuangan dan operasional.

Algoritma K-Nearest Neighbor (KNN) merupakan metode klasifikasi berbasis instance-based learning yang menentukan kelas suatu data berdasarkan mayoritas kelas dari K data terdekat menggunakan ukuran jarak, seperti Euclidean Distance (Putri, 2021); (Yolanda & Fahmi, 2021). KNN tidak memerlukan proses pelatihan model secara eksplisit, sehingga relatif sederhana dan mudah diimplementasikan pada sistem analisis data. Namun, kompleksitas komputasi KNN meningkat seiring bertambahnya jumlah data karena perhitungan jarak dilakukan terhadap seluruh data latih (Prasetyo et al., 2025)

Sejumlah penelitian terdahulu membuktikan efektivitas algoritma KNN dalam berbagai domain. Mulyati et al. (2020) dan Sari (2023) menunjukkan bahwa KNN mampu menghasilkan klasifikasi yang akurat dan stabil pada data pendidikan. Pada domain bisnis, Kasus dan Karawang (2023) menerapkan KNN untuk prediksi penjualan produk dan menunjukkan potensi metode ini dalam analisis pola data transaksi. Penerapan KNN juga meluas ke bidang kesehatan dan sosial ekonomi, seperti klasifikasi citra medis (Penerapan et al., 2025) serta klasifikasi status ketahanan pangan (Sinaga & Ginting, 2025) yang menegaskan fleksibilitas algoritma ini dalam menangani berbagai jenis data.

Meskipun demikian, sebagian besar penelitian sebelumnya menitikberatkan pada pengukuran akurasi klasifikasi dan diterapkan pada data pendidikan, kesehatan, dan penjualan. Penerapan KNN untuk menganalisis keterkaitan antara data gaji pegawai dan biaya operasional pada usaha kuliner masih jarang dilakukan. Selain itu, pemanfaatan hasil klasifikasi sebagai dasar

pendukung pengambilan keputusan manajerial berbasis data belum banyak dibahas secara eksplisit dalam konteks sistem informasi operasional.

Oleh karena itu, penelitian ini mengisi kesenjangan tersebut dengan menerapkan algoritma KNN dalam analisis data gaji pegawai dan biaya operasional melalui kerangka KDD. Pendekatan ini diharapkan dapat memperluas penerapan Data Mining pada sistem informasi usaha kecil dan menengah serta memberikan kontribusi terhadap pengembangan analisis data operasional berbasis keputusan.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan desain studi kasus, yang berfokus pada analisis data numerik untuk mengidentifikasi pola dan hubungan antarvariabel melalui teknik data mining. Pendekatan ini dipilih karena sesuai untuk pengolahan data keuangan dan operasional secara objektif serta mendukung pengambilan keputusan berbasis data.

Penelitian dilaksanakan di Swan Caffee & Resto Rantauprapat, Kabupaten Labuhanbatu. Data yang digunakan merupakan data gaji pegawai, biaya operasional, dan data pendukung kinerja pegawai yang terdokumentasi pada periode pengamatan penelitian. Pemilihan lokasi didasarkan pada ketersediaan dan kelengkapan data operasional yang relevan dengan tujuan penelitian.

Populasi penelitian mencakup seluruh data operasional yang tersedia pada periode pengamatan. Teknik sampling menggunakan total sampling, di mana seluruh data yang memenuhi kriteria kelengkapan digunakan

sebagai sampel. Teknik ini dipilih karena jumlah data relatif terbatas dan seluruh data dianggap representatif untuk menggambarkan kondisi operasional usaha.

Instrumen penelitian berupa dataset operasional dan kuesioner kinerja pegawai dengan indikator disiplin, tanggung jawab, keterlambatan, dan bonus. Data kategorikal ditransformasikan ke dalam bentuk numerik untuk mendukung proses klasifikasi. Validitas data dijaga melalui proses verifikasi internal serta data preprocessing yang meliputi pembersihan dan standarisasi data.

Analisis data dilakukan menggunakan kerangka Knowledge Discovery in Databases (KDD) dengan algoritma K-Nearest Neighbor (KNN) sebagai metode klasifikasi. Penghitungan jarak antar data dilakukan menggunakan Euclidean Distance, kemudian klasifikasi ditentukan berdasarkan mayoritas kelas dari K tetangga terdekat. Data dibagi menjadi training data dan testing data untuk mengevaluasi kinerja model secara sistematis sebagai dasar pendukung pengambilan keputusan manajerial berbasis data.

Kinerja model KNN dievaluasi dengan membandingkan hasil klasifikasi pada data pengujian terhadap data aktual menggunakan metrik akurasi. Evaluasi ini bertujuan untuk memastikan bahwa model mampu mengklasifikasikan data operasional secara konsisten dan layak digunakan sebagai alat bantu analisis dalam pengambilan keputusan manajerial.

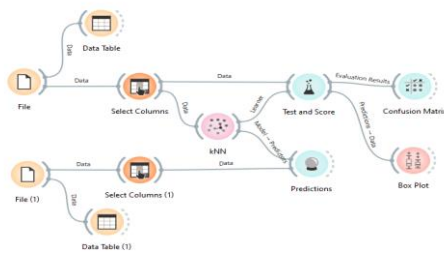
4. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk mengevaluasi kinerja algoritma K-Nearest Neighbor (KNN) dalam mengklasifikasikan data operasional berdasarkan atribut yang

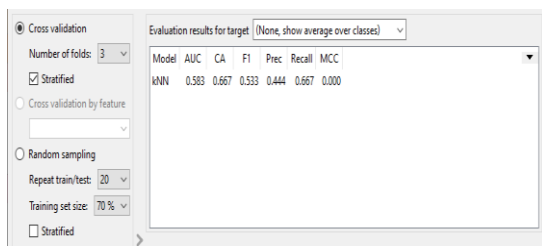
telah ditentukan. Proses pengujian model dilakukan menggunakan perangkat lunak Orange dengan skema random sampling dan proporsi data pelatihan sebesar 70%. Evaluasi dilakukan untuk menilai kemampuan model dalam mengidentifikasi pola klasifikasi serta keterbatasan yang muncul dari karakteristik data yang digunakan.

4.1 Pengujian Data Prediksi

Pada tahap ini data yang telah dimasukkan akan diprediksi oleh sistem orange dengan menyambungkan file prediksi ke predicitons dan load kNN ke predicitons, maka metode yang digunakan akan memperhitungkan hasil prediksi buah yang baik atau tidak seperti gambar di bawah ini:



Gambar 1. Tampilan Prediksi Setelah menghubungkan keduanya ke predictions, maka dapat dilihat hasil prediksinya dengan mengklik dua kali pada bagian predictions. Akan menampilkan seperti gambar berikut ini:



Gambar 2. Hasil Evaluasi Kinerja Model KNN

Berdasarkan Gambar 2, model KNN menghasilkan nilai akurasi sebesar 53,3%, yang menunjukkan bahwa model mampu melakukan klasifikasi sedikit di atas tebakan acak. Nilai Area Under Curve (AUC) sebesar 0,407 mengindikasikan bahwa kemampuan model dalam membedakan kelas masih rendah. Selain itu, nilai precision dan F1-score yang relatif rendah mencerminkan adanya ketidakseimbangan dalam hasil prediksi antar kelas. Nilai Matthews Correlation Coefficient (MCC) yang bernilai negatif menunjukkan korelasi yang lemah antara hasil prediksi dan kelas aktual, sehingga performa model belum optimal secara keseluruhan.

4.2 Analisis Confusion Matrix

		Predicted		Σ
		0	1	
Actual	0	32	6	38
	1	22	0	22
Σ		54	6	60

Gambar 3. Confusion Matrix Model KNN Berdasarkan Jumlah Instance

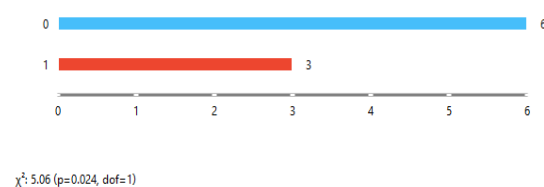
Hasil confusion matrix pada Gambar 3 menunjukkan bahwa dari total 60 data yang diuji, sebanyak 32 data dengan kelas aktual 0 berhasil diklasifikasikan dengan benar, sementara 6 data kelas 0 salah diklasifikasikan sebagai kelas 1. Sebaliknya, seluruh 22 data dengan kelas aktual 1 tidak berhasil diprediksi dengan benar dan diklasifikasikan sebagai kelas 0. Temuan ini menunjukkan bahwa model KNN memiliki kecenderungan bias terhadap kelas mayoritas dan kurang mampu mengenali karakteristik kelas minoritas.

		Predicted		Σ
		0	1	
Actual	0	100.0 %	0.0 %	6
	1	100.0 %	0.0 %	3
Σ		9	0	9

Gambar 4. Confusion Matrix Model KNN Berdasarkan Proporsi Aktual

Berdasarkan proporsi aktual pada Gambar 4, terlihat bahwa tingkat pengenalan kelas 1 oleh model adalah 0%, sedangkan kelas 0 memiliki tingkat pengenalan yang jauh lebih tinggi. Ketidakseimbangan ini menunjukkan bahwa distribusi data yang tidak seimbang antar kelas berdampak signifikan terhadap kinerja algoritma KNN yang berbasis perhitungan jarak. Kondisi ini menyebabkan model lebih dominan memprediksi kelas dengan jumlah data yang lebih besar.

4.3 Analisis Box Plot



Gambar 5. Distribusi Klasifikasi Data Menggunakan Metode KNN (Box Plot]

Gambar 5 menampilkan visualisasi distribusi hasil klasifikasi data menggunakan metode K-Nearest Neighbor (KNN) dalam bentuk box plot. Visualisasi ini menunjukkan perbandingan proporsi hasil klasifikasi antara kelas 0 dan kelas 1 berdasarkan label aktual dan hasil prediksi model. Terlihat bahwa mayoritas data terklasifikasi pada kelas 0,

sementara proporsi data pada kelas 1 relatif lebih kecil, yang mengindikasikan adanya ketidakseimbangan distribusi kelas dalam dataset.

Hasil uji statistik yang ditampilkan pada box plot menunjukkan nilai χ^2 sebesar 2,30 dengan nilai signifikansi $p = 0,129$ ($p > 0,05$). Hal ini mengindikasikan bahwa perbedaan distribusi hasil klasifikasi antar kelas tidak signifikan secara statistik. Temuan ini memperkuat hasil evaluasi sebelumnya bahwa algoritma KNN belum mampu membedakan kelas secara optimal, terutama dalam kondisi distribusi data yang tidak seimbang dan jumlah sampel yang terbatas.

Secara analitis, visualisasi box plot memberikan gambaran tambahan mengenai kecenderungan model dalam mengklasifikasikan data ke kelas mayoritas. Kondisi ini umum terjadi pada algoritma berbasis jarak seperti KNN, di mana kedekatan data dengan kelompok dominan lebih memengaruhi hasil klasifikasi dibandingkan karakteristik kelas minoritas. Dengan demikian, box plot berperan sebagai alat eksploratif yang memperjelas keterbatasan performa model serta karakteristik data yang digunakan dalam penelitian ini.

4.4 Analisis Hasil Prediksi Data

Show probabilities for (None)		Restore Original Order			
KNN	x1	x2	x3	x4	Label
1	1	1	1	1	1
2	0	0	0	0	0
3	1	0	0	0	0

Gambar 6. Contoh Hasil Prediksi Data Menggunakan KNN

Gambar 6 menunjukkan contoh hasil prediksi yang dihasilkan oleh algoritma KNN. Prediksi dilakukan berdasarkan tingkat kemiripan antar atribut numerik. Hasil ini mengindikasikan bahwa KNN mampu mengenali pola data yang serupa, namun masih mengalami kesalahan klasifikasi pada kelas tertentu akibat keterbatasan jumlah data dan dominasi satu kelas dalam dataset.

4.4 Pembahasan

Hasil penelitian menunjukkan bahwa kinerja algoritma KNN sangat dipengaruhi oleh karakteristik dataset yang digunakan. Ukuran data yang relatif kecil serta distribusi kelas yang tidak seimbang menyebabkan algoritma berbasis jarak seperti KNN mengalami kesulitan dalam membedakan kelas minoritas. Meskipun demikian, penerapan KNN tetap memberikan gambaran awal mengenai pola kemiripan data operasional dan potensi pemanfaatannya dalam analisis berbasis sistem pendukung keputusan.

Secara praktis, hasil ini menunjukkan bahwa KNN dapat digunakan sebagai alat eksploratif untuk memahami pola data gaji pegawai dan biaya operasional. Namun, untuk meningkatkan kinerja model, diperlukan pengembangan lebih lanjut seperti penambahan atribut, penyeimbangan data, atau kombinasi dengan metode klasifikasi lain. Dengan demikian, penelitian ini memberikan dasar empiris bagi pengembangan analisis data operasional yang lebih komprehensif pada usaha kecil dan menengah.

5. KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini menunjukkan bahwa algoritma K-Nearest Neighbor (KNN) dapat digunakan untuk mengklasifikasikan data gaji pegawai dan biaya operasional pada usaha kuliner. Hasil evaluasi menunjukkan nilai akurasi sebesar 53,3% dan AUC sebesar 0,407, yang menandakan bahwa kemampuan model dalam membedakan kelas masih terbatas. Analisis confusion matrix dan box plot mengungkap adanya ketidakseimbangan distribusi kelas, sehingga model cenderung memprediksi kelas mayoritas. Meskipun demikian, penerapan KNN tetap memberikan gambaran awal mengenai pola data operasional dan dapat dimanfaatkan sebagai alat eksploratif untuk mendukung pengambilan keputusan berbasis data.

Saran

Penelitian selanjutnya disarankan untuk menambah jumlah data dan menyeimbangkan distribusi kelas agar kinerja klasifikasi meningkat. Selain itu, penggunaan atribut yang lebih beragam serta penerapan atau perbandingan dengan algoritma klasifikasi lain diharapkan dapat menghasilkan model yang lebih optimal dan akurat dalam mendukung pengelolaan data operasional usaha kuliner.

DAFTAR PUSTAKA

- Alfandi, M. S., Fatah, Z., Informasi, S., Ibrahimy, U., Timur, S. J., Informasi, S., Ibrahimy, U., & Timur, S. J. (2024). *PENERAPAN DATA MINING MENGGUNAKAN METODE K-MEANS CLUSTERING UNTUK*. 1(4), 94–102.
- Fau, I. (2024). *Penerapan Data Mining Dengan Metode Support Vector*

Machine Untuk Prediksi Cuaca. 4(1), 19–27.

Hidayat, I. S., Defit, S., & Nurcahyo, G. W. (2021). *Jurnal Sistim Informasi dan Teknologi.* 3, 281–286. <https://doi.org/10.37034/jsisfotek.v3i3.79>

Nirwana, N., & Hayati, M. N. (2021). *CATEGORIC DATA GROUPING BY ALGORITHM QUICK ROBUST CLUSTERING USING LINKS (QROCK) (Case Study : Status of Value Addrd Tax Payments at the Samarinda Ulu Primary Tax Office in 2018).* 9(1).

Penerapan, J., Informasi, T., Ujianto, N. T., Fadilah, H., Fanti, A. P., & Saputra, A. D. (2025). *IT-EXPLORE Penerapan algoritma K-Nearest Neighbors (KNN) untuk klasifikasi citra medis.* 02, 33–43.

Prasetyo, M. B., Oktarina, D., Informasi, S. S., Komputer, F. I., & Indonesia, P. (2025). *Penerapan Data Mining dalam Klasifikasi Penjualan Top Up Game Menggunakan Metode K-Nearest Neighbor.* 3(2024), 232–243.

Putri, A. A. (2021). *Penerapan Data Mining Untuk Memprediksi Penjualan Buah Dan Sayur Menggunakan Metode K-Nearest Neighbor (Studi Kasus : PT . Central Brastagi Utama).* 1(6), 354–361.

Sinaga, C. A., & Ginting, A. K. (2025). *Implementasi Algoritma K-Nearest Neighbors Dengan Pendekatan Elbow Method Untuk Klasifikasi Status Ketahanan Pangan Provinsi Di Indonesia.* 07(01), 27–34.

Yolanda, I., & Fahmi, H. (2021). *Penerapan Data Mining Untuk Prediksi Penjualan Produk Roti*

Terlaris Pada PT . Nippon Indosari Corpindo Tbk Menggunakan Metode K-Nearest Neighbor. 3(3), 9–15.