

Perbandingan Algoritma Klasifikasi K-Nearest Neighbors dan Random Forest Classifier dalam Klasifikasi Tren Penjualan Berdasarkan Waktu dan Kota

Mary Elizabeth Tjang¹

¹Sistem Informasi Fakultas Teknik Informatika,

¹mary.elizabeth@student.umn.ac.id

Abstrak – Tren penjualan yang terus berubah terjadi karena adanya perubahan perilaku konsumen seiring berjalan waktu. Teknologi analisis data, seperti algoritma K-Nearest Neighbors (K-NN) dan Random Forest Classifier, menjadi kunci dalam memahami tren penjualan. Penelitian ini mengeksplorasi kedua algoritma ini untuk menganalisis tren penjualan berdasarkan waktu dan lokasi, serta membandingkan akurasi pada K-NN dan Random Forest Classifier. Berdasarkan dari akurasi model, dapat disimpulkan bahwa akurasi Random Forest lebih bagus daripada K-NN

Kata Kunci: Penjualan, K-NN, Random Forest

PENDAHULUAN

1. Latar Belakang dan Tujuan

Tren penjualan yang kompleks merupakan gambaran dari adanya perubahan perilaku konsumen yang terus berubah seiring waktu berjalan dan adanya perbedaan antara lokasi secara geografis. [12] Perubahan yang terus menerus berubah dalam perilaku konsumen menjadi indikator yang signifikan dalam mengevaluasi dinamika pasar [19]. Kehadiran tren penjualan yang dinamis dan

kompleks menggambarkan perubahan terkini dalam preferensi, kebutuhan, dan pola pembelian konsumen.

Perilaku konsumen yang terus berubah menjadi alasan dengan perubahan tren penjualan. Faktor-faktor seperti perubahan gaya hidup, perubahan kebutuhan, tren model pakaian, dan inovasi produk dapat secara signifikan memengaruhi cara konsumen memilih dan membeli produk. Tren yang terus berubah ini memunculkan tantangan bagi perusahaan untuk terus beradaptasi agar tetap relevan dan memenuhi kebutuhan konsumen yang berubah.

Selain perubahan perilaku konsumen, variabilitas antar lokasi geografis menjadi faktor penting yang mempengaruhi tren penjualan [11]. Preferensi dan kebutuhan konsumen di satu daerah mungkin berbeda dengan yang lainnya, dipengaruhi oleh faktor demografis, budaya, dan ekonomi setempat. Hal ini menyebabkan perbedaan signifikan dalam permintaan produk di berbagai lokasi, menggambarkan kompleksitas dalam memahami pasar secara keseluruhan[22].

Teknologi dan inovasi dalam analisis data juga menjadi kunci dalam memahami tren penjualan yang dinamis ini [5]. Alat analisis data canggih memungkinkan perusahaan untuk melacak perubahan tren secara real-time, mengidentifikasi pola pembelian yang muncul, dan merespons perubahan pasar dengan lebih cepat dan tepat [2].

Dengan menggunakan algoritma klasifikasi K-Nearest Neighbors (K-NN) dan

Random Forest Classifier dalam analisis tren penjualan berdasarkan waktu dan lokasi memberikan wawasan yang penting dalam memahami cara yang efektif untuk memodelkan dan menyusun data penjualan. Dalam penelitian ini, kedua algoritma memberikan pendekatan yang unik dan memiliki kelebihan serta kelemahan yang perlu dipertimbangkan dalam analisis data penjualan.

STUDI LITERATUR

a. Algoritma K-Nearest Neighbors

K-Nearest Neighbors merupakan salah satu algoritma *supervised learning* atau algoritma terbimbing untuk masalah klasifikasi dan regresi [3]. Algoritma KNN memilih K data terdekat dalam dataset sebagai tetangga terdekat (nearest neighbors) dan menggunakan mayoritas kelas dari tetangga terdekat untuk memprediksi kelas dari data yang akan diprediksi [1].

Dalam jurnal yang berjudul "Implementasi Metode K-Means Clustering pada Sistem Rekomendasi Dosen Tetap Berdasarkan Penilaian Dosen" oleh Santoso et al. menjelaskan mengenai langkah-langkah dalam melakukan clustering dengan metode KNN:

- Menentukan jumlah *cluster* k
- Memberi nilai k pusat *cluster* (dapat dilakukan dengan cara random)
- Mengalokasikan semua data pada *cluster* terdekat. Kedekatan dua data ditentukan oleh jarak kedua data tersebut.

Demikian juga kedekatan suatu data ke *cluster* tertentu ditentukan jarak antar data dengan pusat *cluster*. Dalam tahap ini perlu menghitung jarak tiap data ke setiap pusat *cluster*. [6] Jarak antara satu data dengan satu *cluster* tertentu dapat menentukan suatu data akan tergabung ke dalam *cluster* yang mana. Rumus menghitung jarak data ke setiap titik pusat *cluster* dapat menggunakan rumus jarak Euclidian Distance yaitu [7]:

$$d(x_i, x_j) = \sqrt{\sum_{i=0}^n (x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{in} - x_{jn})^2}$$

Gambar 2. 1. Rumus K-Nearest Neighbor

Dengan keterangan:

- $d(X_i, X_j)$: Jarak data ke i ke pusat *cluster* j

X_{i2} : Data ke i pada atribut data ke k

Y_{i2} : Titik pusat ke j pada atribut ke k

- Melakukan perhitungan ulang untuk pusat *cluster* dengan anggota *cluster* yang telah ditentukan. Pusat *cluster* adalah nilai rata-rata dari semua data yang ada di dalam *cluster* dan dapat juga dihitung menggunakan nilai tengah dari *cluster* itu sendiri.
- Mengulangi perintah bagi setiap data untuk menggunakan pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses *clustering* selesai. Jika masih berubah, maka akan kembali ke proses 3 dan terus berulang hingga pusat *cluster* tidak berubah lagi [21].

b. Random Forest Classifier

Algoritma Random Classifier merupakan salah satu algoritma yang digunakan melakukan klasifikasi data berdasarkan dari beberapa pohon keputusan (*Decision Tree*). Algoritma Random Forest memanfaatkan banyak pohon keputusan (decision trees) yang dibangun secara acak selama proses pelatihan [15]. Setiap pohon keputusan dalam "hutan" (forest) ini dibangun dengan menggunakan sampel data yang diambil secara acak dari set data pelatihan. Selain itu, setiap pohon memilih sejumlah fitur secara acak dari dataset tersebut untuk membuat keputusan [4].

Algoritma menggunakan rata-rata untuk meningkatkan akurasi prediksi dan mengontrol *overfitting* (suatu kondisi dimana model terlalu menyesuaikan diri dengan data *training* sehingga menyebabkan model menjadi kurang mampu saat dihadapkan dengan data baru yang belum pernah dilihat sebelumnya) [8]. Pohon keputusan dimulai dengan cara menghitung nilai *entropy* dan *Information gain* sebagai penentu tingkat ketidakmurnian atribut dan nilai *information gain*. Untuk menghitung nilai *entropy* digunakan rumus dibawah

$$Entropy(Y) = - \sum_i p(c|Y) \log_2 p(c|Y)$$

Gambar 2. 2 Rumus Entropy

Keterangan:

- Y: Himpunan kasus
 - $P(c|Y)$: proporsi nilai Y terhadap kelas c
- Nilai information gain menggunakan persamaan

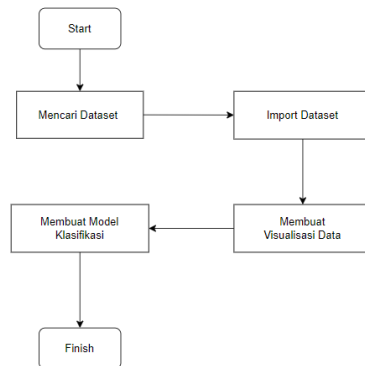
$$Gain(S,A) = Entropy(S) - \sum_{v \in \text{Value}(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Keterangan:

- Entropy(S): nilai Entropy
- S_i : subkelas dari S
- S: Nilai Entropy S

Algoritma ini sering digunakan karena kinerjanya yang baik dalam klasifikasi serta kemampuannya dalam menggeneralisasi dari dataset pelatihan ke data yang belum pernah dilihat sebelumnya[20].

METODOLOGI PENELITIAN



Gambar 3. 1 Flowchart Penelitian

Berikut merupakan tahap-tahap dilakukan oleh peneliti :

Peneliti melakukan beberapa tahap yang dimulai dari mencari dataset lewat data sekunder yaitu Kaggle, kemudian melakukan *import dataset* ke dalam Python, kemudian, membuat visualisasi data yang terdiri dari Histogram, Heatmap, Line Chart, Density Plot dan Pie Chart. Setelah itu, peneliti juga membuat model klasifikasi dengan algoritma K-Nearest Neighbors dan Random Forest Classifier serta membandingkan akurasi pada kedua algoritma tersebut.

3.1 Mencari Dataset

order_id	order_date	status	item_id	sku	qty_order	price	value	discount_total	category	payment	st	cut	id	year	
1E+08	10/1/2020	received	574772	oasis_Oas	21	89.9	1798	0	1798	Men's Fast	cod	Valid	60124	2	
1E+08	10/1/2020	received	574774	Fantastic_	11	19	190	0	190	Men's Fast	cod	Valid	60124	2	
1E+08	10/1/2020	complete	574777	mdeal_Oh	9	149.9	1199.2	0	1199.2	Men's Fast	cod	Net	60124	2	
1E+08	10/1/2020	complete	574779	oasis_Oas	9	79.9	639.2	0	639.2	Men's Fast	cod	Net	60124	2	
1E+08	11/13/2020	received	595185	MEFNAR5	2	99.9	99.9	0	99.9	Men's Fast	cod	Valid	60124	2	
1E+08	11/13/2020	received	595186	MEFBL151	2	39.9	39.9	0	39.9	Men's Fast	cod	Valid	60124	2	
1E+08	11/13/2020	order_ref	595192	MATDANG	2	47.6	47.6	0	47.6	Modles & cod		Valid	60124	2	
1E+08	10/1/2020	canceled	574789	091_01_U	2	49	49	0	49	Modles & Payax		Gross	42485	2	
1E+08	10/1/2020	canceled	574770	oasis_King	2	135	135	0	135	Computing Payax		Gross	42485	2	
1E+08	10/1/2020	canceled	571771	Geepas_G	2	549.0	549.0	0	549.0	Appliances Payax		Gross	42485	2	
1E+08	10/9/2020	order_ref	577687	Geepas_G	1	549.9	0	0	0	Appliances Payax		Valid	42485	2	
1E+08	10/11/2020	order_ref	581862	APPOVESS	1	549.9	0	0	0	Appliances cod		Valid	42485	2	
1E+08	12/24/2020	canceled	656937	APPOVESS	2	254.8	30.80628	214.9937	254.8	Appliances Easy	pay	Gross	42485	2	
1E+08	12/24/2020	canceled	656938	APPOVESS	2	315.5	315.5	49.28917	266.2108	Appliances Easy	pay	Gross	42485	2	
1E+08	12/24/2020	canceled	656939	HALSHO2E	2	69.8	69.8	10.90455	58.89545	Home & Li	Easy	pay	Gross	42485	2
1E+08	12/24/2020	canceled	656939	APPOVESS	2	254.8	254.8	30.80628	214.9937	Appliances Easy	pay	Gross	42485	2	

Gambar 3. 2. Isi Dataset SALES

Pada tahap pertama, peneliti melakukan pencarian dataset yang sesuai dengan kebutuhan peneliti. Peneliti menemukan dataset, tentang penjualan di suatu perusahaan. Dataset ini merupakan dataset penjualan dari tahun 2020 sampai tahun 2021.

3.2 Melakukan Import Dataset

Pada tahap kedua, setelah menemukan dataset yang diperlukan, peneliti melakukan

```

sales = pd.read_csv("sales.csv")
sales.head()

C:\Users\Asus\AppData\Local\Temp\ipykernel_7840\3320443873.py:1: DtypeWarning:
Columns (0) have mixed types. Specify dtype option on import or set low_m
emory=False.
sales = pd.read_csv("sales.csv")
  
```

	order_id	order_date	status	item_id	sku	qty_ordered	price	val
0	100354678	2020-10-01	received	574772.0	oasis_Oasis-064-36	21.0	89.9	1798
1	100354678	2020-10-01	received	574774.0	Fantastic_FT-48	11.0	19.0	190
2	100354680	2020-10-01	complete	574777.0	mdeal_DMC-610-8	9.0	149.9	1199
3	100354680	2020-10-01	complete	574779.0	oasis_Oasis-061-36	9.0	79.9	639
4	100367357	2020-11-13	received	595185.0	MEFNAR59C38B6CA08CD	2.0	99.9	99

5 rows × 9 columns

Gambar 3. 3 Import Dataset import dataset ke dalam Jupyter Notebook.

Pada Dataset ini terdapat 36 kolom dan terdiri dari 8950 baris yang terdiri dari:

- order_id : Nomor ID unik untuk setiap pemesanan.
- order_date : Tanggal pemesanan oleh pelanggan
- status : Status pemesanan
- item_id : Nomor ID setiap item
- sku: Surat keterangan usaha
- qty_ordered: kuantitas order
- price: harga barang
- value: value barang
- total: total barang

- category: Kategori barang
- bi_st:
- cust_id Nomor unik pada setiap pelanggan
- year: tahun pemesanan
- month: tanggal pemesanan
- ref_number: referensi nomor
- Name Preference: Panggilan nama
- First name: Nama depan
- Middle Initial: Inisial tengah
- Last Name: Name belakang
- Gender: Jenis Kelamin
- age: umur
- full_name: nama panjang
- E-Mail: email pelanggan
- Customer Since: kapan menjadi pelanggan tetap
- SSN: Nomor jaminan sosial
- Phone Number: nomor telepon pelanggan
- Place Name: tempat tinggal pelanggan
- County: daerah pelanggan
- City: kota pelanggan
- State: negara pelanggan
- Zip: zip negara
- Region: wilayah pelanggan
- username: nama yang digunakan untuk
- Discount_percent: diskon yang digunakan oleh pelanggan

3.4 Membuat Visualisasi Data

Visualisasi data yang terdiri dari Histogram, Heatmap, Line Chart, Density Plot dan Pie Chart yang dilakukan pada Jupyter Notebook. Pembahasan ini akan dilanjutkan kepada “HASIL DAN PEMBAHASAN”

3.5 Membuat Model Klasifikasi

Algoritma Klasifikasi yang digunakan adalah K-Nearest Neighbors dan Random Forest Classifier yang digunakan untuk melakukan perbandingan hasil akurasi dari kedua algoritma tersebut. Hasil dari akurasi algoritma akan dibahas di “HASIL DAN PEMBAHASAN”.

HASIL DAN PEMBAHASAN

Hasil visualisasi pada Dataset Sales

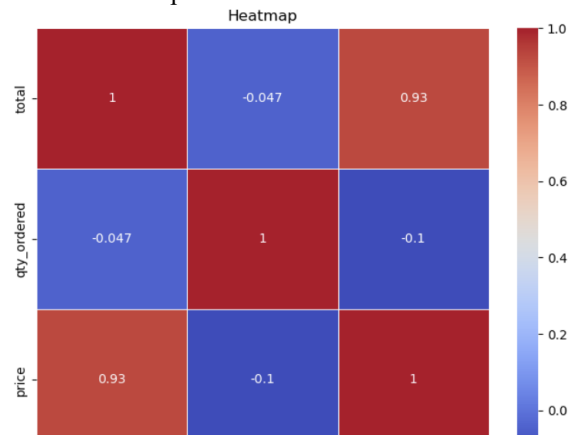
- Histogram



Gambar 4. 1 Visualisasi data Histogram

Berdasarkan Histogram ini dapat disimpulkan bahwa total dari pemesan banyak sekali mencapai diatas 250.000 dengan frekuensi atau terlalu sering membeli pakaian tersebut sebanyak 200.000

- Heatmap

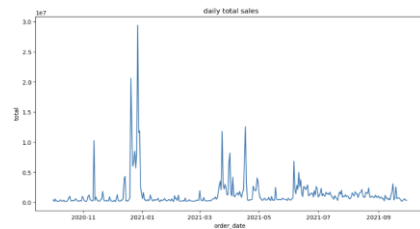


Gambar 4. 2 Visualisasi data Heatmap

Berdasarkan Heatmap ini dapat disimpulkan bahwa heatmap ini menampilkan hubungan antara kolom total dan “price” yang positif dan kuat karena mendekati 1, sementara hubungan antara “qty_ordered” dan “price” serta “total” yang mendekat 0.0

atau bahkan mines (-) yang artinya tidak ada korelasi sama sekali.

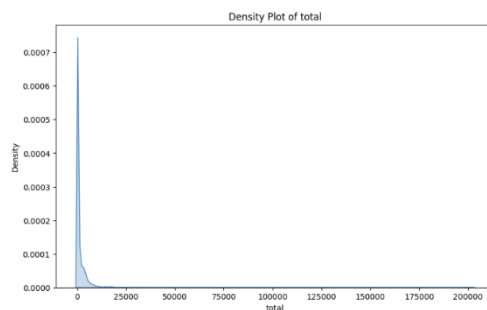
- Line Chart



Gambar 4. 3 Visualisasi data Line Chart

Berdasarkan dari Line Chart ini, dapat disimpulkan bahwa pada bulan Januari tahun 2021, menampilkan total penjualan yang tinggi.

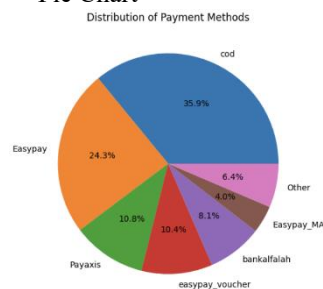
- Density Plot



Gambar 4. 5 Visualisasi data Density Plot

Berdasarkan dari Density Plot ini, dapat dilihat bahwa nilai-nilai yang terdapat dalam kolom “total” menyebar dalam Density Plot ini.

- Pie Chart



Gambar 4. 7 Visualisasi data Pie Chart

Berdasarkan dari Pie Chart, dapat disimpulkan bahwa sebanyak 35.9% melakukan COD (Cash On Delivery), sebanyak 24.3% melakukan transfer

melalui Easypay, setelah itu sebanyak 10.8% melakukan transfer melalui Payaxis, kemudian sebanyak 10.4% membayar dengan easypay_voucher, kemudian sebanyak 8.1% membayar dengan bankalfalah, setelah itu sebanyak 4.0% melakukan pembayaran melalui Easypay_MA, dan 6.4% melakukan pembayaran dengan metode lain.

Hasil akurasi model klasifikasi antara K-Nearest Neighbors dan Random Forest Classifier.

- Hasil model akurasi pada algoritma K-Nearest Neighbors.

```
from sklearn.metrics import classification_report, accuracy_score

# Predicting the test set results
y_pred = knn_model.predict(X_test)
```

```
# Generating the classification report and accuracy
print(classification_report(y_test, y_pred))
print("Accuracy:", accuracy_score(y_test, y_pred))
```

	precision	recall	f1-score	support
0	0.82	0.87	0.84	24485
1	0.80	0.81	0.80	15202
2	0.85	0.85	0.85	30875
3	0.84	0.76	0.80	15356
accuracy			0.83	85918
macro avg	0.83	0.82	0.82	85918
weighted avg	0.83	0.83	0.83	85918

Gambar 4. 4 Code dan Hasil dari KNN

Berdasarkan dari hasil KNN yang telah dibuat, menunjukkan akurasi sebesar 0.83 / 83% yang artinya cocok untuk dilakukan klasifikasi pada dataset ini.

- Hasil model akurasi pada algoritma Random Forest Classifier

```
from sklearn.metrics import classification_report, accuracy_score

# Predicting the test set results
y_pred = rf_model.predict(X_test)
```

```
# Generating the classification report and accuracy
print(classification_report(y_test, y_pred))
print("Accuracy:", accuracy_score(y_test, y_pred))
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	24485
1	1.00	1.00	1.00	15202
2	1.00	1.00	1.00	30875
3	1.00	1.00	1.00	15356
accuracy			1.00	85918
macro avg	1.00	1.00	1.00	85918
weighted avg	1.00	1.00	1.00	85918

Accuracy: 1.0

Berdasarkan dari hasil model Random

Gambar 3. 6 Code dan Hasil dari Random Forest Classifier

Forest Classifier yang telah dibuat,

menunjukkan akurasi sebesar 1.00 / 100% yang artinya sangat cocok dalam melakukan klasifikasi pada dataset ini.

KESIMPULAN

Pada penelitian ini dilakukan pengujian menggunakan 2 algoritma yaitu K-Nearest Neighbors (KNN) dan Random Forest Classifier. Berdasarkan hasil dari implementasi algoritma KNN dan Random Forest Classifier untuk klasifikasi tren penjualan berdasarkan waktu dan kota, dapat disimpulkan bahwa akurasi dari algoritma Random Forest Classifier mendapatkan akurasi 100% dan untuk algoritma KNN mendapatkan akurasi 83%. Maka dari itu, berdasarkan hasil penelitian ini menunjukkan bahwa algoritma yang terbaik dalam klasifikasi tren penjualan adalah Random Forest Classifier

REFERENSI

- [1] AJI, S. M. (2022). *KOMPARASI ALGORITMA NAÏVE BAYES DAN K-NEAREST NEIGHBOR UNTUK MENENTUKAN KLASIFIKASI PRODUK TERLARIS (STUDI KASUS: PERUSAHAAN FROZEN FOOD XYZ)* (Doctoral dissertation, Universitas Mercu Buana Jakarta).
- [2] Amiarrahman, M. R., & Handhika, T. (2018). Analisis 1dan Implementasi Algoritma Klasifikasi Random Forest Dalam Pengenalan Bahasa Isyarat Indonesia (BISINDO). Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi), 2(1), 083–088. <https://doi.org/10.29407/inotek.v2i1.461>
- [3] Depari, D. H., Widiastiwi, Y., & Santoni, M. M. (2022). Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung. *Informatik*, 18(3), 239. <https://doi.org/10.52958/iftk.v18i3.4694>
- [4] Erlin, E., Desnelita, Y., Nasution, N., Suryati, L., & Zoromi, F. (2022). Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak Seimbang. *Matrik: Jurnal Manajemen, Teknik Informatika, Dan Rekayasa Komputer*, 21 (3), 677–690. <https://doi.org/10.30812/matrik.v21i3.1726>
- [5] Faisal, F., Dhika, H., & Veris, H. (2021). Penerapan Algoritma Decision Tree Dalam Penjualan Handphone. *JRKT (Jurnal Rekayasa Komputasi Terapan)*, 1(04).
- [6] Harani, N. H., Prianto, C., & Nugraha, F. A. (2020). Segmentasi pelanggan produk digital service Indihome menggunakan algoritma K-Means berbasis Python. *Jamika*, 10(2), 133–146. <https://doi.org/10.34010/jamika.v10i2.2683>
- [7] Herlambang, H. P., Saputra, F., Prasetyo, M. H., Puspitasari, D., & Nurlaela, D. (2023). Perbandingan Klasifikasi Tingkat Penjualan Buah di Supermarket dengan Pendekatan Algoritma Decision Tree, Naive Bayes dan K-Nearest Neighbor. *Jurnal Insan*, 3(1), 21-28.
- [8] Kurniawan, I., Buani, DCP, Abdussomad, A., Apriliah, W., & Saputra, RA (2023). Implementasi algoritma Random Forest untuk menentukan penerima bantuan Raskin. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 10 (2), 421–428. <https://doi.org/10.25126/jtiik.20231026225>
- [9] Rivki, M., & Bachtiar, A. M. (2017). IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR DALAM PENGKLASIFIKASIAN FOLLOWER TWITTER YANG MENGGUNAKAN BAHASA INDONESIA. *Jurnal Sistem Informasi*, 13(1), 31-37. <https://doi.org/10.21609/jsi.v13i1.500>
- [10] Herianto, H., & Syamsiyah, N. (2021). ANALISA TINGKAT PENJUALAN PRODUK MENGGUNAKAN K-NEAREST NEIGHBOR (K-NN) DAN K-MEANS (STUDI KASUS PERUSAHAAN KAYU ELANG PERKASA). *Jurnal Sains Dan Teknologi*, 11(1), 8–18. <https://unsada.e-journal.id/jst/article/download/105/85>
- [11] Lalo, A. K., Batarius, P., & Siki, Y. C. H. (2021). Implementasi Algoritma C4. 5

- Untuk Klasifikasi Penjualan Barang di Swalayan Dutalia. *Jurnal Teknik Informatika UNIKA Santo Thomas*, 6, 1-12.
- [12] Mardiani, E., Rahmansyah, N., Ningsih, S., Lantana, D. A., Wirawan, A. S. P., Wijaya, S. A., & Putri, D. N. (2023). Komparasi Metode Knn, Naive Bayes, Decision Tree, Ensemble, Linear Regression Terhadap Analisis Performa Pelajar Sma. *Innovative: Journal Of Social Science Research*, 3(2), 13880-13892.
- [13] Nasution, M. R. A., & Hayaty, M. (2019). Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter. *Jurnal Informatika*, 6(2), 226–235. <https://doi.org/10.31311/ji.v6i2.5129>
- [14] Nawangsih, I., & Setyaningsih, A. (2019). Penerapan Algoritma Naive Bayes Untuk Menentukan Klasifikasi Produk Terlaris Pada Penjualan Pulsa. *Jurnal SIGMA*, 10(4), 9-15.
- [15] Pahlevi, O., Amrin, A., & Handrianto, Y. (2023). Implementasi Algoritma Klasifikasi Random Forest Untuk Penilaian Kelayakan Kredit. *Jurnal Infortech*, 5(1), 71–76. <https://doi.org/10.31294/infortech.v5i1.15829>
- [16] Permadi, M., Febriani, F., & Panggabean, M. J. (2023, August). Penerapan algoritma K-Nearest Neighbor pada klasifikasi Penjualan produk Indihome daerah Riau. In *SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat* (pp. 105-111).
- [17] Priatmojo, H., Saputra, F., Prasetyo, M. H., Puspitasari, D., & Nurlaela, D. (2023). Perbandingan Klasifikasi Tingkat Penjualan Buah di Supermarket dengan Pendekatan Algoritma Decision Tree, Naive Bayes dan K-Nearest Neighbor. *Jurnal INSAN (Journal of Information Systems Management Innovation)*, 3(1).
- [18] Mardiani, E., Rahmansyah, N., Ningsih, S., Lantana, D. A., Wirawan, A. S. P., Wijaya, S. A., & Putri, D. N. (2023). Komparasi Metode Knn, Naive Bayes, Decision Tree, Ensemble, Linear Regression Terhadap Analisis Performa Pelajar Sma. *Innovative: Journal Of Social Science Research*, 3(2), 13880-13892.
- [19] Rindiyan, R., Primadewi, A., Maimunah, M., & Purwantini, A. H. (n.d.). Klasifikasi Penjualan berdasarkan Platform pada UMKM Omah Branded Menggunakan Random Forest. *JURIKOM (Jurnal Riset Komputer)*, 9(5), 1520. <https://doi.org/10.30865/jurikom.v9i5.4949>
- [20] Samosir, A., Hasibuan, M. S., Justino, W. E., & Hariyono, T. (2021). Komparasi Algoritma Random Forest, Naive Bayes dan K- Nearest Neighbor Dalam klasifikasi Data Penyakit Jantung. *Prosiding Seminar Nasional Darmajaya*, 1, 214–222. <https://jurnal.darmajaya.ac.id/index.php/P SND/article/download/2955/1272>
- [21] Santoso, Y. P., Marlina, & Agung, H. (2018, Desember). Implementasi Metode K-Means Clustering pada Sistem Rekomendasi Dosen Tetap Berdasarkan Penilaian Dosen. *Jurnal Informatika Universitas Pamulang*, 3(4). <http://dx.doi.org/10.32493/informatika.v3i4.2133>
- [22] Sopiyan, I. A., Pudjiantoro, T. H., & Santikarama, I. (2021). Klasifikasi Penjualan Jus Dengan Menggunakan Algoritma K-Nearest Neighbor (K-NN) Untuk Penerapan Konsep Up-Selling. *PROSIDING SNIA*, 5. <https://snia.unjani.ac.id/web/index.php/sn ia/article/download/245/130>
- [23] Suryatman, S. (2017). PENERAPAN METODE CLUSTERING DENGAN ALGORITMA K-MEANS UNTUK MENGANALISIS FAKTOR YANG MEMPENGARUHI NILAI CUMLAUDE MAHASISWA FAKULTAS TEKNIK UNIVERSITAS MUHAMMADIYAH YOGYAKARTA. <http://repository.umy.ac.id/handle/123456789/11085>

