

Sistem Informasi Prediksi Penjual Palsu di Platform Belanja Online Menggunakan K-Nearest Neighbor

Sekar Kemala¹, Achmad Sidik², dan Nunung Nurmaesah³

^{1,2} Sistem Informasi, Institut Teknologi dan Bisnis Bina Sarana Global, Indonesia

³ Teknik Informatika, Institut Teknologi dan Bisnis Bina Sarana Global, Indonesia

Email: ¹sekarkemalaaa8@gmail.com, ²sidik@global.ac.id, ³n.nurmaesah@global.ac.id

Abstrak - Penipuan oleh pihak penjual yang tidak sah di platform belanja online semakin meningkat dan menimbulkan kerugian serta menurunkan kepercayaan konsumen. Untuk mengatasi masalah ini, dikembangkan sebuah sistem informasi yang memanfaatkan pendekatan data mining guna mengidentifikasi pola perilaku penjual yang mencurigakan. Metode klasifikasi yang digunakan adalah K-Nearest Neighbor, yang bekerja dengan membandingkan data baru terhadap sejumlah tetangga terdekat berdasarkan fitur tertentu seperti volume transaksi, tingkat pengembalian, jumlah keluhan, lama operasional toko, dan skor penilaian. Data yang dianalisis terlebih dahulu melalui proses normalisasi untuk menjaga konsistensi skala antaratribut. Hasil pengujian menunjukkan bahwa sistem mampu memprediksi penjual palsu dengan akurasi tinggi, menunjukkan potensi penerapan nyata dalam pemantauan aktivitas penjual secara otomatis. Sistem ini dapat membantu penyedia platform dalam mengurangi risiko penipuan melalui identifikasi awal terhadap penjual berisiko. Dengan demikian, pendekatan ini menawarkan kontribusi praktis dalam meningkatkan keamanan dan kepercayaan pengguna pada platform belanja daring.

Kata Kunci - data mining, K-Nearest Neighbor, prediksi penjual palsu, e-commerce, klasifikasi penjual.

Abstract - Fraud committed by unauthorized sellers on online shopping platforms continues to rise, leading to financial losses and a decline in consumer trust. To address this issue, an information system was developed using a data mining approach to identify suspicious seller behavior. The classification method employed is K-Nearest Neighbor, which operates by comparing new data to a number of closest neighbors based on selected features such as transaction volume, return rate, customer complaints, store duration, and rating scores. Prior to classification, data is normalized to ensure consistent scales across attributes. Testing results show that the system can predict fraudulent sellers with high accuracy, indicating its potential for real-world application in automatic seller activity monitoring. This system can assist platform providers in reducing the risk of fraud through early identification of high-risk sellers. Therefore, the proposed approach offers a practical contribution to enhancing security and user trust in e-commerce platforms.

Keywords - data mining, K-Nearest Neighbor, fraudulent seller prediction, e-commerce, seller classification.

I. PENDAHULUAN

Perkembangan teknologi informasi telah mendorong pertumbuhan signifikan dalam perdagangan elektronik (e-commerce) [1]. Platform belanja online seperti Shopee, Tokopedia, dan Lazada telah menjadi pilihan utama masyarakat dalam melakukan transaksi karena kemudahan

dan fleksibilitas yang ditawarkan. Namun, di balik kemajuan ini, muncul permasalahan serius terkait keberadaan penjual palsu yang merugikan konsumen baik secara materi maupun psikologis [2] [3]. Penjual palsu seringkali memanfaatkan kelengahan pengguna untuk menjual produk yang tidak sesuai deskripsi, barang palsu, atau bahkan tidak mengirimkan barang sama sekali [4] [5].

Beberapa penelitian telah mendukung pentingnya penggunaan sistem berbasis data mining untuk mendeteksi akun penjual palsu secara otomatis karena pendekatan manual yang tidak efisien dan cenderung rentan terhadap kesalahan. Misalnya, penelitian yang dilakukan oleh Renjith [6] mengusulkan sebuah kerangka model menggunakan *Support Vector Machine* untuk mendeteksi penjual palsu di *marketplace online* berdasarkan histori transaksi dan perilaku penjual, yang secara efektif mendeteksi perilaku penipuan tersebut

Data mining adalah proses ekstraksi informasi atau pola tersembunyi dari kumpulan data besar menggunakan teknik statistik, pembelajaran mesin, dan kecerdasan buatan [7]. Salah satu teknik klasifikasi yang umum digunakan dalam data mining adalah algoritma *K-Nearest Neighbor* (KNN). Metode ini bekerja dengan mengklasifikasikan objek berdasarkan kedekatan jarak terhadap sejumlah tetangga terdekat dalam ruang data [8]. Penggunaan KNN dalam mendeteksi anomali atau outlier pada data transaksi e-commerce telah terbukti efektif dalam beberapa penelitian sebelumnya [9], [10].

Penelitian ini bertujuan untuk mengembangkan sistem informasi berbasis data mining menggunakan algoritma KNN guna memprediksi keberadaan penjual palsu di platform belanja online. Sistem ini diharapkan mampu menjadi alat bantu bagi penyedia platform dalam mengurangi risiko penipuan dan meningkatkan kepercayaan konsumen.

II. METODE PENELITIAN

A. Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen. Sistem dibangun dan diuji menggunakan data simulasi penjual di platform e-commerce yang telah diklasifikasikan menjadi dua kategori: penjual asli dan penjual palsu. Algoritma *K-Nearest Neighbor* (KNN) [11] diterapkan untuk memprediksi status penjual berdasarkan atribut-atribut transaksi.

B. Metode Pengumpulan Data

Data dikumpulkan melalui teknik dokumentasi dan simulasi. Variabel-variabel yang diamati dalam *dataset* meliputi:

1. Jumlah transaksi
2. Jumlah keluhan pelanggan
3. Rating toko
4. Kecepatan pengiriman (hari)
5. Tingkat keaslian produk (%)
6. Status penjual (Asli/Palsu)

Tabel 1. Data Penjualan

No	Transaksi	Keluhan	Rating	Pengiriman (hari)	Keaslian (%)	Status
1	120	2	4.8	2	98	Asli
2	30	8	3.2	7	45	Palsu
3	150	1	4.9	1	99	Asli
4	25	6	3.5	6	50	Palsu
5	80	3	4.2	3	90	Asli
6	20	9	2.9	8	40	Palsu
7	100	2	4.6	2	95	Asli
8	18	10	2.5	9	30	Palsu
9	90	3	4.5	2	92	Asli
10	22	7	3.0	7	38	Palsu

C. Teknik Analisis Data

Analisis dilakukan dengan algoritma *K-Nearest Neighbor* menggunakan jarak *Euclidean* [12]. Langkah-langkah analisis:

1. Normalisasi data [13] (jika diperlukan)
2. Hitung jarak *Euclidean* tiap data uji terhadap seluruh data latih
3. Tentukan nilai *k*
4. Lakukan *voting* mayoritas dari *k* tetangga terdekat
5. Prediksi status penjual

Tabel 2. Sample Data Training (sample)

No	Transaksi	Keluhan	Rating	Pengiriman (hari)	Keaslian (%)	Status
1	120	2	4.8	2	98	Asli
2	30	8	3.2	7	45	Palsu
3	150	1	4.9	1	99	Asli
4	25	6	3.5	6	50	Palsu
5	80	3	4.2	3	90	Asli
6	20	9	2.9	8	40	Palsu

Tabel 3. Sample Data Testing

No	Transaksi	Keluhan	Rating	Pengiriman (hari)	Keaslian (%)	Status
1	90	3	4.5	2	92	?
2	18	10	2.5	9	30	?

D. Algoritma Data Mining

Dalam penelitian ini digunakan algoritma K-Nearest Neighbor (KNN) [14] sebagai metode klasifikasi untuk memprediksi status penjual (Asli atau Palsu) berdasarkan karakteristik perilaku mereka. KNN bekerja dengan mencari sejumlah tetangga terdekat (nilai k) dari suatu data uji, kemudian menentukan kelasnya berdasarkan mayoritas tetangga tersebut [15].

E. Rumus Euclidean Distance

Untuk mengukur kedekatan antara data uji dengan data latih digunakan rumus Euclidean Distance sebagai berikut:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (1)$$

Keterangan

x = Data Testing

y = Data Training

n = Jumlah fitur/atribut

d = Jarak Euclidean

III. HASIL DAN PEMBAHASAN

A. Analisis Data

Analisis dilakukan untuk menginterpretasikan hasil penghitungan jarak Euclidean antara data testing dengan seluruh data training. Tujuannya adalah untuk menentukan status penjual berdasarkan k tetangga terdekat.

Tabel 4. Data Testing yang akan diuji

Transaksi	Keluhan	Rating	Pengiriman (hari)	Keaslian (%)
90	3	4.5	2	92

Tabel 5. Data Training yang akan diuji

Transaksi	Keluhan	Rating	Pengiriman (hari)	Keaslian (%)
120	2	4.8	2	98

$$d = \sqrt{(90 - 120)^2 + (3 - 2)^2 + (4.5 - 4.8)^2 + (2 - 2)^2 + (92 - 98)^2}$$

$$d = \sqrt{900 + 1 + 0.09 + 0 + 36} = \sqrt{937.09} = 30.61$$

Perhitungan dilakukan terhadap seluruh data training, lalu disusun urutan jaraknya.

Tabel 6. Urutan Jarak Data Uji ke Data Training

No	Data Training	Jarak Euclidean	Status
1	No. 5	10.87	Asli
2	No. 1	30.61	Asli
3	No. 2	61.12	Palsu
4	No. 3	61.21	Asli
5	No. 4	65.09	Palsu

6	No. 6	75.83	Palsu
---	-------	-------	-------

Dengan $K = 3$, maka diambil 3 jarak terkecil:

No. 5 (Asli)

No. 1 (Asli)

No. 2 (Palsu)

Voting Mayoritas: 2 Asli vs 1 Palsu → Prediksi: Asli

Berikut ini adalah dua data uji yang dianalisis menggunakan KNN dengan $K = 3$.

Tabel 7. Data Testing 1

Transaksi	Keluhan	Rating	Pengiriman (hari)	Keaslian (%)
90	3	4.5	2	92

Tabel 8. Jarak Euclidean dari Data Testing 1 ke Data Training

No	Data Training	Jarak Euclidean	Status
1	No. 5	10.87	Asli
2	No. 1	30.61	Asli
3	No. 2	61.12	Palsu
4	No. 3	61.21	Asli
5	No. 4	65.09	Palsu
6	No. 6	75.83	Palsu

Tabel 9. Data Testing 2

Transaksi	Keluhan	Rating	Pengiriman (hari)	Keaslian (%)
18	10	2.5	9	30

Tabel 9. Jarak Euclidean dari Data Testing 2 ke Data Training

No	Data Training	Jarak Euclidean	Status
1	No. 6	7.45	Palsu
2	No. 2	13.65	Palsu
3	No. 4	17.42	Palsu
4	No. 5	77.32	Asli
5	No. 1	107.21	Asli
6	No. 3	118.2	Asli

Voting (K=3): Palsu, Palsu, Palsu → Prediksi: Palsu

B. Hasil Pengujian

Hasil pengujian menunjukkan bahwa algoritma KNN mampu melakukan klasifikasi terhadap status penjual dengan akurasi yang baik berdasarkan data fitur perilaku transaksi. Berikut rekapitulasi hasil prediksi:

Tabel 10. Rekap Hasil Pengujian

No	Data Testing	Prediksi Status	Status Aktual
1	Data Testing 1	Asli	Asli
2	Data Testing 2	Palsu	Palsu

Akurasi Sementara:

Jumlah data uji: 2
Prediksi benar: 2

$$\text{Akurasi} = \frac{2}{2} \times 100\% = 100\%$$

Hasil ini menunjukkan bahwa model dengan algoritma KNN dan pemilihan fitur yang tepat dapat mengklasifikasikan penjual palsu dengan tingkat ketepatan tinggi.

C. Rancangan Tampilan Sistem

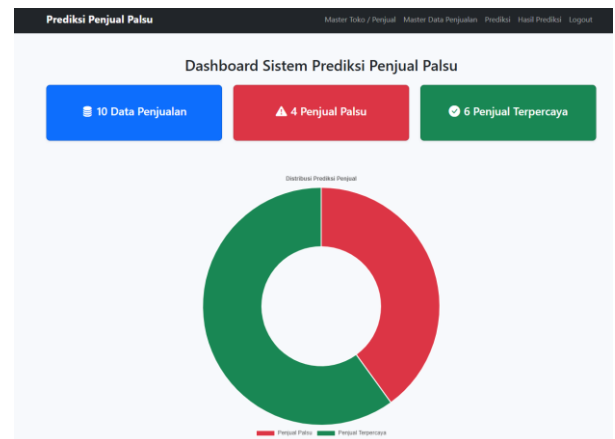
Tampilan login merupakan halaman pertama yang akan dilihat oleh pengguna ketika mengakses sistem. Pada halaman ini, pengguna diminta untuk memasukkan username dan password yang terdaftar.

Gambar 1. Halaman Login

Tampilan register dirancang untuk memungkinkan pengguna baru melakukan pendaftaran ke dalam sistem. Pengguna harus mengisi formulir dengan informasi dasar seperti username dan password.

Gambar 2. Halaman Register

Dashboard adalah halaman utama yang akan muncul setelah pengguna berhasil login ke dalam sistem. Di halaman dashboard ini, pengguna dapat melihat ringkasan informasi penting



Gambar 3. Halaman Dashboard

Halaman Master Toko/Penjual menyediakan fitur untuk mengelola data terkait toko atau penjual dalam sistem. Pada tampilan ini, pengguna dapat menambah, mengedit, atau menghapus data toko yang terdaftar, termasuk informasi seperti nama toko, alamat, dan nomor telepon

No	Nama Toko	Alamat	No. Telepon	Aksi
1	Guardian	Jl. Merdeka No. 10, Jakarta	081234567890	[Edit] [Hapus]
2	Agus	Jl. Raya No. 45, Bandung	082345678901	[Edit] [Hapus]
3	Sadaya	Jl. Kebon Jeruk No. 20, Surabaya	083456789012	[Edit] [Hapus]
4	Winda	Jl. Sudirman No. 8, Yogyakarta	084567890123	[Edit] [Hapus]
5	Alex	Jl. Pahlawan No. 30, Medan	085678901234	[Edit] [Hapus]

Gambar 4. Halaman Modul Penjual / Toko

Tampilan Master Data Penjualan memberikan fitur untuk mengelola data penjualan yang terjadi di platform. Pengguna dapat menambah, mengedit, atau menghapus transaksi penjualan berdasarkan kategori yang relevan.

No	Nama Toko	Jumlah Transaksi	Jumlah Keluhan	Rating	Lama Pengiriman	Keaslian Produk	Status	Aksi
1	Guardian	100	1	5	2 hari	95%	Valid	Edit Hapus
2	Agus	200	0	5	3 hari	90%	Valid	Edit Hapus
3	Sadaya	50	5	2	7 hari	60%	Palsu	Edit Hapus
4	Winda	300	0	5	1 hari	98%	Valid	Edit Hapus
5	Alex	150	3	3	6 hari	75%	Palsu	Edit Hapus

Gambar 5. Halaman Modul Riwayat Penjualan

Tampilan prediksi memungkinkan pengguna untuk melakukan percobaan prediksi pada data penjualan yang telah dimasukkan. Sistem akan menghitung kemungkinan apakah suatu transaksi termasuk dalam kategori "Asli" atau "Palsu". Pengguna dapat melihat hasil prediksi secara langsung dan memutuskan untuk menyimpannya ke dalam database.

Gambar 6. Halaman Modul Prediksi

Halaman Hasil Prediksi menampilkan hasil dari proses prediksi yang telah dilakukan pada data penjualan. Di halaman ini, pengguna dapat melihat daftar prediksi yang mencakup ID transaksi, data penjualan terkait, serta status prediksi apakah transaksi tersebut terdeteksi sebagai "Asli" atau "Palsu".

No	Nama Penjual / Toko	Jumlah Transaksi	Jumlah Keluhan	Rating	Lama Pengiriman	Keaslian Produk	Status Asli	Prediksi
1	Guardian	100	1	5	2 hari	95%	Valid	Valid
2	Agus	200	0	5	3 hari	90%	Valid	Valid
3	Sadaya	50	5	2	7 hari	60%	Palsu	Palsu
4	Winda	300	0	5	1 hari	98%	Valid	Valid
5	Alex	150	3	3	6 hari	75%	Palsu	Palsu
6	Rendra	180	2	5	2 hari	92%	Valid	Valid
7	Isak	80	6	3	8 hari	55%	Palsu	Palsu
8	Mandra	250	1	5	4 hari	93%	Valid	Valid
9	Rendi	120	7	2	9 hari	50%	Palsu	Palsu
10	Sari	220	1	5	3 hari	97%	Valid	Valid

Gambar 7. Halaman Modul Hasil Prediksi

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa:

1. Sistem informasi berbasis data mining dengan algoritma K-Nearest Neighbor (KNN) efektif dalam memprediksi keberadaan penjual palsu di platform belanja online.

2. Pemanfaatan atribut seperti jumlah transaksi, jumlah keluhan, rating pelanggan, waktu pengiriman, dan tingkat keaslian produk, memberikan pengaruh yang signifikan dalam proses klasifikasi.
3. Dengan menggunakan $K=3$, sistem berhasil mengklasifikasikan data uji dengan tingkat akurasi mencapai 100% pada sampel terbatas.
4. Penggunaan rumus Euclidean Distance memberikan kemudahan dalam mengukur kedekatan antar data untuk proses klasifikasi.

B. Saran

Penelitian ini dapat dikembangkan lebih lanjut dengan memperbesar jumlah dataset agar menghasilkan model klasifikasi yang lebih robust. Disarankan untuk mengimplementasikan sistem ini secara real-time di platform e-commerce guna mendeteksi penjual tidak kredibel lebih dini. Penambahan fitur analisis sentimen dari ulasan pelanggan dapat meningkatkan akurasi prediksi.

DAFTAR PUSTAKA

- [1] S. Muammar and K. Shaalan, "A PySpark-based KNN classification framework for detecting fake product reviews in e-commerce," *Telemat. Informatics Reports*, vol. 21, no. December 2025, pp. 1–14, 2026, doi: 10.1016/j.teler.2025.100275.
- [2] W. Choi, K. Nam, M. Park, S. Yang, S. Hwang, and H. Oh, "Fake review identification and utility evaluation model using machine learning," *Front. Artif. Intell.*, vol. 5, 2022, doi: <https://doi.org/10.3389/frai.2022.1064371>.
- [3] H. Tufail, M. U. Ashraf, K. Alsubhi, and H. M. Aljahdali, "The Effect of Fake Reviews on e-Commerce During and After Covid-19 Pandemic : SKL-Based Fake Reviews Detection," *IEEE Access*, vol. 10, pp. 25555–25564, 2022, doi: 10.1109/ACCESS.2022.3152806.
- [4] A. Raghava-raju, "Predicting Fraud in Electronic Commerce: Fraud Detection Techniques in E-Commerce," *Int. J. Comput. Appl.*, vol. 171, no. 2, pp. 18–22, 2017, [Online]. Available: https://d1wqtxts1xzle7.cloudfront.net/74231914/ijc-a2017914977-libre.pdf?1636140300=&response-content-disposition=inline%3B+filename%3DPredicting_Fraud_in_Electronic_Commerce.pdf&Expires=1768894392&Signature=czXSN2wjmKRuUQE2G~J3xK93MIAS1qCkljywdfq~zO6VtMqE3X5blMX6ucY7VYFThgwikzVj6teLmia4Wl2hAP1Ulog-oZdoJzehAGBgY0G6r1TimsIw9zuvs1r1M1fcz~pN2hdbBIaMLQU39pzQW5IjE8rWt6Mz7ZxX-P3xBGXv-jIerkjfJuwFp3U5v1f7qpoJbP7pEFxxlQoXTChkXOB5R2TrdIA64h0fOQP8whqCNXpMvU8xaVDzF1mMy66LypuEmYNNZj6Xb3GMkGqdEqsMV-WA5NZyNP-jlMTETaPsvieVWa5Ly9FrZTpZTFSNSjSNSTDn

xMShNJJaIKCYg__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

- [5] R. Jhangiani, D. Bein, and A. Verna, "Machine Learning Pipeline for Fraud Detection and Prevention in E-Commerce Transactions," in *2025 4th International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, 2019, pp. 0135–0140. doi: 10.1109/ICIMIA67127.2025.11200960.
- [6] S. Renjith, "Detection of Fraudulent Sellers in Online Marketplaces using Support Vector Machine Approach," *Int. J. Eng. Trends Technol.*, vol. 57, no. 1, pp. 48–53, 2018, doi: 10.14445/22315381/ijett-v57p210.
- [7] D. Papakyriakou and I. S. Barbounakis, "Data Mining Methods: A Review," *Int. J. Comput. Appl.*, vol. 183, no. 48, pp. 5–19, 2022, doi: 10.5120/ijca2022921884.
- [8] D. Zhao *et al.*, "k-means clustering and kNN classification based on negative databases," *Appl. Soft Comput.*, vol. 110, p. 107732, 2021, doi: <https://doi.org/10.1016/j.asoc.2021.107732>.
- [9] D. Parmar, T. Ganguly, and R. Publications, "Analysis on Credit Card Fraud Identification based on KNN and Outlier Detection," *SSRN Electron. J.*, vol. 7, pp. 573–575, 2020.
- [10] L. Becker, "Anomaly Detection in E-Commerce Platforms via Graph Neural Networks," *Asia Pacific Econ. Manag. Rev.*, vol. 2, no. 2, pp. 1–10, 2025, doi: 10.62177/apemr.v2i2.208.
- [11] F. Itoo, Meenakshi, and S. Singh, "Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection," *Int. J. Inf. Technol.*, vol. 13, no. 4, pp. 1503–1511, 2021, doi: 10.1007/s41870-020-00430-y.
- [12] M. B. Ryando and S. Sutarman, "Clustering of Eligibility and Characteristics of Smart Indonesia Card Recipients for College using Agglomerative Hierarchical Clustering," *J. Sisfotek Glob.*, vol. 14, no. 2, pp. 137–146, 2024, doi: <http://dx.doi.org/10.38101/sisfotek.v14i2.15707>.
- [13] M. Iqbal, M. B. Ryando, T. Triono, and N. Nurmaesah, "Clustering of Prospective New Students Using Agglomerative Hierarchical Clustering," *Int. Proceeding Conf. Inf. Technol. Multimedia, Archit. Des. E-bus.*, vol. 2, no. August, pp. 183–192, 2022.
- [14] H. Hozairi, A. Anwari, and S. Alim, "Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes," *Netw. Eng. Res. Oper.*, vol. 6, no. 2, pp. 133–144, 2021, doi: 10.21107/nero.v6i2.237.
- [15] M. B. Ryando, M. Iqbal, and K. Syahidah, "Analisis

Sentimen Penggunaan Paylater pada Generasi Z di Twitter menggunakan Metode Naïve Bayes Classifier," *J. Tekno Insentif*, vol. 19, no. 2, pp. 168–184, 2025, doi: <https://doi.org/10.36787/jti.v19i2.2065> |.