

IDENTIFIKASI POTENSI PENIPUAN PADA TRANSAKSI BANK MENGUNAKAN *K-MEANS CLUSTERING*

Jessica Uly Sari Hutagalung[✉], Kristin Trivena Sihombing, Michael Owen Hutabarat,
Syanti Irviantina

Fakultas Informatika, Universitas Mikroskil, Medan, Indonesia

Email: jessicauly01@gmail.com

DOI: <https://doi.org/10.46880/methoda.Vol14No3.pp392-395>

ABSTRACT

Increasing cases of fraud in bank transactions are a serious concern for financial institutions, resulting in significant economic losses and undermining customer trust. This calls for identifying suspicious transaction patterns through machine-learning approaches to mitigate the risk of fraud. The methods used include problem identification, transaction data collection, and preprocessing to clean and prepare the data and after that, applying the K-Means Clustering algorithm to group transactions based on similar characteristics. The evaluation result obtained in this study using the Silhouette Score is 0.42, indicating a fairly good separation between normal and suspicious transactions. This research is expected to contribute to the development of a more accurate and efficient machine learning-based fraud detection system in banking institutions.

Keyword: *Fraud, Machine Learning, K-Means Clustering, Silhouette Score.*

ABSTRAK

Meningkatnya kasus penipuan dalam transaksi bank menjadi perhatian serius bagi institusi keuangan, mengakibatkan kerugian finansial yang signifikan dan merusak kepercayaan nasabah. Hal ini membutuhkan identifikasi pola transaksi yang mencurigakan melalui pendekatan machine learning untuk memitigasi resiko penipuan. Metode yang digunakan mencakup identifikasi masalah, pengumpulan data transaksi, dan preprocessing untuk membersihkan dan mempersiapkan data. Setelah itu, menerapkan algoritma K - Means Clustering untuk mengelompokkan transaksi berdasarkan kesamaan karakteristik. Hasil evaluasi yang diperoleh dalam penelitian ini dengan menggunakan Silhouette Score adalah 0.42, menunjukkan adanya pemisahan yang cukup baik antara transaksi normal dan yang mencurigakan. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem deteksi penipuan berbasis machine learning yang lebih akurat dan efisien di institusi perbankan.

Kata Kunci: *Penipuan, Machine Learning, K-Means Clustering, Silhouette Score.*

PENDAHULUAN

Sistem keuangan digital saat ini menghadapi masalah besar terkait kejahatan finansial. Proses perbankan *online* yang semakin kompleks memungkinkan pelaku penipuan untuk melakukan tindakan yang melanggar hukum. Karena banyaknya transaksi dan polanya yang beragam, pendekatan konvensional untuk mengidentifikasi *fraud* menjadi sulit. *Fraud* adalah tindakan yang

melanggar hukum dan dilakukan dengan sengaja untuk mendapatkan keuntungan secara tidak adil (Purnama & Arianto, 2024). Tindakan ini dapat dilakukan oleh individu, kelompok, ataupun organisasi. Oleh karena itu, dibutuhkan identifikasi potensi penipuan untuk mengurangi kerugian. Untuk menyelesaikan masalah tersebut, peneliti melakukan studi mengenai penerapan model *machine learning* dengan

pendekatan *unsupervised learning* menggunakan *K-Means Clustering*.

Metode pengelompokan yang paling umum adalah *K-Means*. Metode *clustering non-hierarchical* yang dikenal sebagai algoritma *K-Means* memiliki waktu komputasi yang cukup cepat. *Clustering* merupakan sebuah teknologi *machine learning* tanpa pengawasan yang mengelompokkan data berdasarkan tingkat kemiripan (Saputra, 2024). Tujuan dari penelitian ini adalah untuk mengevaluasi kemampuan metode *K - Means Clustering* dalam mengidentifikasi pola transaksi yang tidak biasa atau mencurigakan pada transaksi bank.

KAJIAN LITERATUR

Identifikasi Penipuan Pada Transaksi Bank

Fraud adalah tindakan penyimpangan atau pembiaran yang sengaja dilakukan untuk mengelabui, menipu, atau memanipulasi Bank, nasabah, atau pihak lain di lingkungan Bank dan/atau menggunakan sarana Bank sehingga mengakibatkan kerugian Bank, nasabah, atau pihak lain, atau pelaku *fraud* memperoleh keuntungan keuangan, baik secara langsung maupun tidak langsung (OJK, 2022). Semua jenis aktivitas perbankan, seperti penyetoran, penarikan, transfer, pengajuan pinjaman, dan investasi dapat terlibat dalam penipuan perbankan, yang merupakan jenis kejahatan keuangan yang melibatkan penggunaan praktik penipuan atau ilegal untuk mendapatkan keuntungan finansial yang tidak adil atau melanggar hukum. Penipuan perbankan dapat dilakukan oleh individu maupun organisasi, dan jenis penipuan ini berkisar dari yang kecil hingga yang besar (*Bank Fraud – Types and Examples*, 2024).

K - Means

K-means merupakan salah satu algoritma dalam *machine learning* yang bersifat *unsupervised learning*. *K-Means* berfungsi untuk mengelompokkan data ke dalam data *cluster*. *K-Means Clustering* adalah metode *non-hierarchy* dan dapat menerima data tanpa label kategori (Orleans & Putra, 2022.). Cara kerja *K-Means Clustering* dimulai dengan memilih secara acak sejumlah titik awal yang dikenal

sebagai "*centroids*". Semua data akan diberikan ke kumpulan yang memiliki *centroid* terdekat berdasarkan jarak *Euclidean*. *Centroid* dari setiap kumpulan akan dihitung ulang sebagai rata-rata dari semua data dari kumpulan tersebut. Sampai penempatan data dalam kelompok tidak mengalami perubahan yang signifikan atau hingga jumlah iterasi yang ditetapkan, proses dilakukan berulang kali (Feby, 2022).

Clustering

Clustering merupakan metode untuk mengelompokkan data berdasarkan kesamaan maupun perbedaan di antara data tersebut (RevoU, 2024). *Clustering* membagi set data menjadi kelompok-kelompok, atau kluster, di mana masing-masing kelompok memiliki karakteristik yang sama dan berbeda. *Clustering* tidak mengelompokkan atau mengklasifikasikan objek data karena tidak memerlukan data pelatihan, atau *training* data. Sebaliknya, *clustering* mengidentifikasi pola intrinsik dalam data dan mengelompokkannya berdasarkan kesamaan tertentu.

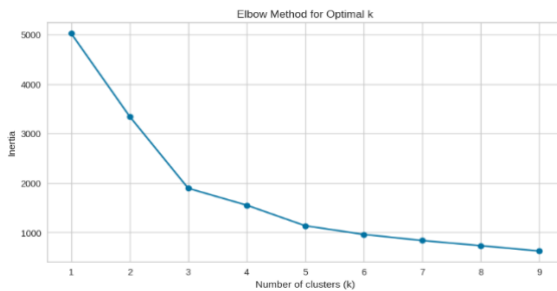
METODE PENELITIAN

Metode penelitian dimulai dengan identifikasi masalah, yang bertujuan untuk menentukan masalah apa yang akan diselesaikan. Selanjutnya, data yang relevan dikumpulkan untuk analisis lanjutan. Kemudian, proses *preprocessing* data melibatkan membersihkan dan mempersiapkan data agar siap digunakan, termasuk menangani masalah seperti *noise*, duplikasi, dan data yang hilang. Lalu, temukan jumlah *K* yang optimal kemudian algoritma *K-Means* diterapkan yang merupakan teknik *clustering* untuk mengelompokkan data berdasarkan kemiripan. Terakhir, evaluasi model dilakukan menggunakan *Silhouette Score* untuk mengevaluasi kualitas pengelompokan yang dihasilkan dan menentukan jumlah kluster yang paling cocok untuk analisis data.

HASIL DAN PEMBAHASAN

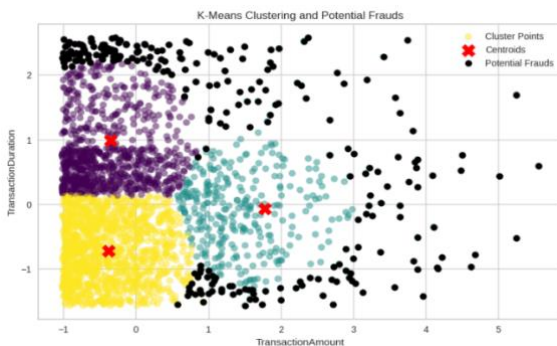
Proses analisis dimulai dengan menentukan jumlah kluster yang optimal. Grafik yang terlihat pada Gambar 1, menunjukkan hasil metode *Elbow* untuk menentukan jumlah kluster

optimal dalam algoritma *K-Means*. Sumbu *x* merepresentasikan jumlah kluster (*k*) yang diuji, mulai dari 1 hingga 9, sementara sumbu *y* menunjukkan nilai inersia, yang merupakan ukuran seberapa baik data dikelompokkan dalam kluster. Metode "Elbow" digunakan untuk menentukan jumlah kluster yang optimal dengan mencari titik di mana penurunan inerti mulai melambat, yang biasanya terlihat seperti "siku" (*elbow*) pada grafik.



Gambar 1. Visualisasi Nilai Inersia untuk Berbagai Jumlah Kluster

Kemudian, penerapan algoritma *K-Means* untuk pengelompokan data. Pertama, *model K-Means* diinisialisasi dengan tiga kluster dan parameter *random_state* untuk memastikan reproduktivitas. Setelah model dilatih menggunakan data yang telah diskalakan, label kluster untuk setiap data *point* diperoleh dan ditampilkan. Lalu, mendeteksi potensi *fraud* dalam *dataset* menggunakan hasil dari algoritma *K-Means*. Jumlah data yang memenuhi kriteria ini dihitung dan ditampilkan, menunjukkan berapa banyak potensi *fraud* yang terdeteksi yaitu sebanyak 252. Selanjutnya, memvisualisasikan hasil kluster *K - Means* dapat dilihat pada Gambar 2.



Gambar 2. Hasil Kluster K-Means dan Potensi Penipuan Dalam Dataset

Gambar tersebut menunjukkan hasil dari algoritma *K-Means clustering* pada *dataset* yang memiliki dua fitur, yaitu *TransactionAmount* dan *TransactionDuration*. Titik-titik berwarna yang tersebar di area *plot* mewakili data transaksi yang telah dikelompokkan ke dalam kluster dapat dilihat pada Tabel 1.

Tabel 1. Cluster Points

<i>Cluster</i>	<i>Transaction Amount</i>	<i>Transaction Duration</i>
0	Rendah	Tinggi
1	Rendah	Rendah
2	Tinggi	Rendah

Titik-titik berwarna hitam pada Gambar 2, menunjukkan transaksi yang dianggap berpotensi penipuan. Titik-titik ini dapat tersebar di seluruh area, tetapi beberapa mungkin berada di dekat *centroid* atau di area yang memiliki kluster tertentu, menunjukkan bahwa transaksi tersebut memiliki karakteristik yang mirip dengan transaksi yang telah dikelompokkan.

KESIMPULAN

Hasil analisis *clustering* menggunakan *Silhouette Score* memiliki nilai sebesar 0.42, menunjukkan bahwa model cukup baik. Nilai *Silhouette Score* ini berada di antara -1 hingga 1, di mana nilai yang lebih tinggi mengindikasikan pemisahan kluster yang lebih baik. Dalam hal ini, angka 0.42 menunjukkan bahwa meskipun kluster memiliki pemisahan yang cukup baik namun masih perlu perbaikan. Nilai ini menunjukkan bahwa beberapa titik data mungkin berada di tepi antara kluster, yang dapat mengindikasikan bahwa pemisahan kluster tidak sepenuhnya optimal. Dari hasil evaluasi model, dapat disimpulkan bahwa walaupun *clustering* yang dilakukan sudah memberikan pemisahan yang dapat diterima, hasilnya masih dapat ditingkatkan.

Beberapa langkah yang dapat diambil untuk meningkatkan kualitas *clustering* adalah menguji berbagai jumlah kluster yang bisa membantu menemukan konfigurasi yang lebih baik dengan pemisahan yang lebih jelas, atau meninjau fitur yang digunakan dalam proses

clustering. Mungkin ada fitur lain yang lebih relevan yang bisa ditambahkan atau metode *preprocessing* yang lebih baik yang dapat diterapkan untuk meningkatkan hasil. Bisa dipertimbangkan juga untuk menggunakan algoritma *clustering* yang lain.

DAFTAR PUSTAKA

- Bank Fraud – Types and examples*. (2024). Fraud.Com. <https://www.fraud.com/post/bank-fraud>
- Feby, D. (2022). *Konsep Model Machine Learning K-Means Clustering*. DQLab. Retrieved Oktober 18, 2024, from <https://dqlab.id/konsep-model-machine-learning-k-means-clustering>
- OJK. (2022). Panduan Strategi Anti-Fraud Bagi Bank Perkreditan Rakyat (BPR). *Otoritas Jasa Keuangan*, 1–118.
- Orleans, B., & Putra, E. P. (2022). *Clustering Algoritma (K-Means)*. BINUS UNIVERSITY School of Information Systems. Retrieved Oktober 18, 2024, from <https://sis.binus.ac.id/2022/01/31/clustering-algoritma-k-means/>
- Purnama, V., & Arianto, D. B. (2024). *Penerapan K-Means Clustering pada Data Pembayaran Tagihan Kartu Kredit Untuk Menganalisis Potensi Fraud*. 4(2), 52–57. <https://doi.org/https://doi.org/10.34001/jjister.v4i2.1199>
- RevoU. *Clustering*. (2024). Retrieved September 18, 2024, from <https://www.revou.co/kosakata/clustering>
- Saputra, C. H. (2024). Integrasi Audit dan Teknik *Clustering* untuk Segmentasi dan Kategorisasi Aktivitas Log. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 11(1), 209–214. <https://doi.org/10.25126/jtiik.20241118071>