



FINAL REPORT
TUGAS PROYEK PENAMBANGAN DATA

Judul
Implementasi Algoritma K-Nearest Neighbor (KNN) untuk
klasifikasi Data Penipuan Kartu Kredit

Dosen: Ardi Susanto, S. Kom., M.Cs

Anggota Kelompok:
Avriansyah Bahtiar (21090024), Fajri Abdul Ghani (21090067), Zidan Afkar Maula
(21090109)

ABSTRAK

Kemudahan dalam mengakses informasi dapat menyebabkan penyalahgunaan informasi terutama pada tindak kejahatan mulai dari penipuan, penyerangan dan berbagai kasus lainnya. Salah satu bentuk manipulasi data pada bidang e-commerce adalah penipuan pada kartu kredit. Transaksi kartu kredit adalah metode pembayaran yang paling umum dalam beberapa tahun terakhir. Meskipun kartu kredit bukanlah alat pembayaran yang paling diminati di Indonesia. Tingginya angka penipuan kartu kredit menyebabkan banyaknya kasus yang terjadi dan sulit untuk dilakukan pencegahan. Sehingga perlu dilakukan deteksi penipuan kartu kredit sedini mungkin. Deteksi penipuan adalah proses mendeteksi apakah suatu transaksi sah atau tidak. Sistem deteksi penipuan otomatis diperlukan terutama mengingat lalu lintas data transaksi yang besar, dan tidak mungkin bagi manusia untuk memeriksa secara manual setiap transaksi satu per satu apakah itu curang atau tidak. Sistem deteksi penipuan otomatis dengan menggunakan teknik pembelajaran mesin dan metode lain sudah banyak berkembang. Dalam upaya mengatasi masalah ini, implementasi algoritma K-Nearest Neighbors (KNN) dapat digunakan untuk mengklasifikasikan data penipuan kartu kredit.

Kata kunci : kartu kredit, *KNN*

1. Permasalahan

Kemudahan dalam mengakses informasi dapat menyebabkan penyalahgunaan informasi terutama pada tindak kejahatan mulai dari penipuan, penyerangan dan berbagai kasus lainnya. Salah satu bentuk manipulasi data pada bidang e-commerce adalah penipuan pada kartu kredit. Transaksi kartu kredit adalah metode pembayaran yang paling umum dalam beberapa tahun terakhir. Meskipun kartu kredit bukanlah alat pembayaran yang paling diminati di Indonesia. Tingginya angka penipuan kartu kredit menyebabkan banyaknya kasus yang terjadi dan sulit untuk dilakukan pencegahan. Sehingga perlu dilakukan deteksi penipuan kartu kredit sedini mungkin. Deteksi penipuan adalah proses mendeteksi apakah suatu transaksi sah atau tidak. Sistem deteksi penipuan otomatis diperlukan terutama mengingat lalu lintas data transaksi yang besar, dan tidak mungkin bagi manusia untuk memeriksa secara manual setiap transaksi satu per satu. Sistem deteksi penipuan otomatis dengan menggunakan teknik pembelajaran mesin dan metode lain sudah banyak berkembang. Dalam upaya mengatasi masalah ini, implementasi algoritma K-Nearest Neighbors (KNN) dapat digunakan untuk mengklasifikasikan data penipuan kartu kredit. Algoritma KNN dalam konteks ini bekerja dengan mencari K transaksi penipuan terdekat berdasarkan jarak atau kesamaan atribut dengan transaksi baru. Metrik jarak seperti jarak Euclidean atau jarak Manhattan digunakan untuk mengukur kesamaan antara transaksi. Setelah mendapatkan tetangga terdekat, mayoritas label penipuan atau bukan penipuan dari tetangga tersebut akan digunakan untuk mengklasifikasikan transaksi baru.

2. Dataset

Dataset yang digunakan diambil dari <https://www.kaggle.com/>, dengan URL <https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud/code?select=creditcard.csv> tentang Credit Card Fraud Detection. Terdapat 284806 kolom data dengan 31 atribut diantaranya 'Time', 'Amount', 'V1 -V28', 'Class', contoh dataset dapat dilihat pada gambar di bawah.

	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	...	V21	V22	V23	V24	V25	V26	V27	V28	Amount	Class
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.363787	...	-0.018307	0.277838	-0.110474	0.066928	0.128539	-0.189115	0.133558	-0.021053	149.62	0
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803	0.085102	-0.255425	...	-0.225775	-0.638672	0.101288	-0.339846	0.167170	0.125895	-0.008983	0.014724	2.69	0
2	1.0	-1.358354	-1.340163	1.773209	0.378780	-0.503198	1.800499	0.791461	0.247676	-1.514654	...	0.247998	0.771679	0.909412	-0.689281	-0.327642	-0.139097	-0.055353	-0.059752	378.66	0
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609	0.377436	-1.387024	...	-0.108300	0.005274	-0.190321	-1.175575	0.647376	-0.221929	0.062723	0.061458	123.50	0
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941	-0.270533	0.817739	...	-0.009431	0.798278	-0.137458	0.141267	-0.206010	0.502292	0.219422	0.215153	69.99	0

5 rows x 31 columns

3. Metode

A. Preprocessing

1. Memuat Dataset: Mengimport dataset yang akan digunakan untuk melakukan KNN.

2. Mengimpor Libraries: Import library yang diperlukan, seperti NumPy, Pandas, dan scikit-learn.

B. Feature Extraction

3. Memisahkan Fitur dan Label: Pisahkan kolom fitur (variabel independen) dan kolom label (variabel dependen) dari dataset. Kolom fitur berisi atribut yang digunakan untuk melakukan prediksi, sedangkan kolom label berisi nilai yang ingin diprediksi.

4. Normalisasi Fitur (Opsional): Jika skala fitur-fitur dalam dataset berbeda-beda, normalisasi fitur dapat dilakukan untuk menghindari dominasi fitur dengan skala yang lebih besar.

C. Classification

5. Membagi Dataset: Bagi dataset menjadi subset data latih (training set) dan subset data uji (test set). Data latih digunakan untuk melatih model KNN, sedangkan data uji digunakan untuk menguji performa model.

6. Menentukan Nilai K: Tentukan jumlah tetangga terdekat (nilai K) yang akan digunakan dalam proses KNN. Pemilihan nilai K dapat dilakukan dengan menggunakan teknik validasi silang atau metode lainnya.

7. Menentukan Kelas Prediksi: Menentukan kelas prediksi berdasarkan mayoritas kelas dari tetangga terdekat. Jika $K = 1$, kelas dari tetangga terdekat langsung dijadikan prediksi.

8. Evaluasi Model: Melakukan evaluasi performa model KNN menggunakan matrik evaluasi yang sesuai, seperti akurasi, presisi, recall, atau metrik lainnya.

4. Hasil Eksperimen

	precision	recall	f1-score	support
0	1.00	1.00	1.00	85273
1	0.92	0.78	0.84	170
accuracy			1.00	85443
macro avg	0.96	0.89	0.92	85443
weighted avg	1.00	1.00	1.00	85443

Hasil eksperimen klasifikasi pada dataset credit card menggunakan *K-Nearest Neighbor (KNN)* menunjukkan matrik evaluasi sebagai berikut:

1. Precision: $0 = 1.00$, $1 = 0.92$

Precision mengukur seberapa baik model dalam mengidentifikasi kelas positif dengan benar. Semakin tinggi nilai precision, semakin sedikit kesalahan yang dilakukan dalam mengklasifikasikan sampel sebagai kelas positif.

2. Recall (Sensitivitas atau True Positive Rate): 0 = 1.00, 1 = 0.78

Recall mengukur seberapa baik model dapat menemukan kembali (mendeteksi) sampel-sampel yang benar dari kelas positif. Semakin tinggi nilai recall, semakin sedikit sampel dari kelas positif yang terlewatkan oleh model.

3. F1-Score: 0 = 1.00, 1 = 0.84

F1-score merupakan harmonic mean dari precision dan recall. F1-score memberikan gambaran tentang keseimbangan antara precision dan recall. Nilai F1-score yang tinggi menunjukkan performa yang baik dalam mengklasifikasikan kelas positif.

4. Support: 0 = 85273, 1 = 170

Support mengindikasikan jumlah kemunculan setiap kelas dalam dataset. Support memberikan informasi tentang seberapa banyak sampel yang ada untuk setiap kelas.

Dalam keseluruhan, model memiliki akurasi (accuracy) sebesar 0.9993504441557529, yang menunjukkan bahwa model dapat mengklasifikasikan dengan sangat baik. Meskipun demikian, perlu diperhatikan bahwa kelas 1 memiliki precision, recall, dan f1-score yang lebih rendah dibandingkan dengan kelas 0. Hal ini menunjukkan adanya ketidakseimbangan kinerja dalam memprediksi kelas positif, yang dapat menjadi fokus untuk peningkatan model di masa mendatang.

5. Kontribusi

- a. Avriansyah Bahtiar dalam proyek ini melakukan pencarian dataset dan metode yang akan digunakan, melakukan pengujian code dan hasil eksperimen.
- b. Fajri Abdul Ghani dalam proyek ini melakukan pencarian dataset dan metode yang akan digunakan, mencari referensi jurnal, membuat abstrak.
- c. Ziidan Afkar Maula dalam proyek ini melakukan pencarian dataset dan metode yang akan digunakan, mencari referensi jurnal, membuat permasalahan.

6. Referensi

Habibie, M. I. (2019). Deteksi Fraud Menggunakan Metode K-Means dan Euclidean Distance dalam Sensor IoT. Tesis tidak diterbitkan, Teknik Elektro, Teknik, Universitas Indonesia, Depok 16951, Indonesia. (Link jurnal <https://conferences.ittikom-pwt.ac.id/index.php/centive/article/view/80/79>)

Febriady, M., Samsuryadi*, & Rini, D. P. (2022). Klasifikasi Transaksi Penipuan Pada Kartu Kredit Menggunakan Metode Resampling Dan Pembelajaran Mesin. Jurnal Media Informatika Budidarma, 6(2), 1010-1016. (Link Jurnal : <http://ejurnal.stmik-budidarma.ac.id/index.php/mib/article/view/3515>)