



Algoritma Fuzzy C-Means untuk Klasterisasi Penggunaan Kartu Kredit

(*Fuzzy C-Means Algorithm for Clusterization of Credit Card Usage*)

Miftahul Jannah¹, Melani Nur Mudyawati², Acep Razif Andriyan³, Dinda Meysya Rochma⁴, Siti Lufia Dwi Agustini⁵

¹Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050061@student.uinsgd.ac.id

²Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050060@student.uinsgd.ac.id

³Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050002@student.uinsgd.ac.id

⁴Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050032@student.uinsgd.ac.id

⁵Teknik Informatika, UIN Sunan Gunung Djati Bandung, 1177050007@student.uinsgd.ac.id

Abstrak

Kartu kredit pada saat ini sudah tidak asing lagi ditelinga masyarakat. Kartu kredit merupakan sebuah alat pembayaran pengganti uang tunai dalam bentuk kartu yang diterbitkan oleh bank untuk memudahkan pada nasabah bertransaksi. Saat ini terdapat berbagai macam perusahaan jasa keuangan penerbitan kartu kredit di dunia termasuk Indonesia. Dengan berbagai manfaatnya sehingga kredit digandrungi oleh semua kalangan, sehingga semua orang berlomba lomba menggunakan kredit dengan memilih bank yang diinginkan. Metode data mining dapat memberi solusi untuk menggali pengetahuan dari data dengan mencari pola atau aturan tertentu dari sejumlah data dalam jumlah besar. Salah satu metode data mining adalah Clustering, dimana Clustering digunakan untuk mengelompokkan data dengan cara mengelompokkan data tersebut ke dalam beberapa cluster. Dengan menggunakan dataset kartu kredit dibagi menjadi 3 cluster menggunakan algoritma Fuzzy C-Means. Dari 3 cluster tersebut yang lebih diprioritaskan dengan nilai terbesar merupakan cluster yang banyak digunakan oleh banyak orang, yang mempunyai nilai pertengahan itu sedang, sedangkan cluster dengan nilai terkecil merupakan cluster yang sedikit peminatnya.

Kata kunci: Data Mining, Fuzzy C-Means, Kartu Kredit, Klasterisasi

Abstract

Credit cards are now familiar to the public. A credit card is a means of payment in lieu of cash in the form of a card issued by the bank to facilitate transactions for customers. Currently, there are various kinds of credit card issuing financial service companies in the world, including Indonesia. With various benefits so that credit is loved by all groups, so that everyone competes to use credit by choosing the desired bank. Data mining methods can provide solutions to extract knowledge from data by looking for certain patterns or rules from large amounts of data. One of the data mining methods is clustering, in which clustering is used to group data by grouping the data into several clusters. By using the credit card data set is divided into 3 clusters using the Fuzzy C-Means algorithm. Of the 3 clusters, the ones that are prioritized with the largest value are the clusters that are widely used by many people, which have a medium value, while the cluster with the smallest value is the cluster with the least interest.

Keywords: Data Mining, Clustering, Credit Card, Fuzzy C-Means

1 Pendahuluan

Kartu Kredit merupakan sebuah alat pembayaran yang berfungsi sebagai pengganti uang tunai, dimana alat tersebut dapat digunakan oleh konsumen untuk ditukarkan dengan berbagai barang dan jasa yang dibelinya di tempat-tempat yang bisa menerima pembayaran dengan menggunakan kartu kredit. Pada saat ini bagi sebagian besar pengguna kartu kredit seperti sebuah gaya hidup yang senantiasa dibutuhkan dan menjadi hal wajib di dalam kehidupan mereka. Pada bank-bank Indonesia, kartu kredit diciptakan untuk memudahkan dan cara kerjanya diatur oleh Bank Indonesia (BI). Dengan banyak bank yang menerbitkan kartu kredit sehingga pada metode clustering ini akan merekomendasikan kartu kredit.

Salah satu metode yang dapat digunakan untuk menyelesaikan permasalahan tersebut adalah dengan Clustering yang merupakan salah satu metode data mining [1]. Clustering berguna untuk mengelompokkan data yang didapat dari dataset didasari pada informasi yang ditemukan dalam data yang menggambarkan objek tersebut dan hubungan diantaranya. Dengan menggunakan algoritma Fuzzy C-Means yang merupakan salah satu algoritma dalam pengelompokkan data karena membuat suatu perkiraan yang efisien dan tidak memerlukan banyak parameter. Beberapa penelitian telah menghasilkan kesimpulan bahwa metode Fuzzy C-Means dapat digunakan untuk pengelompokan data berdasarkan atribut-atribut tertentu. Pada penelitian ini akan menganalisis penerapan metode Fuzzy C-Means untuk mengelompokkan data rekomendasi kartu kredit dengan dataset nasabah.

Terdapat beberapa penelitian terdahulu yang terkait, antara lain: (1) pengelompokkan citra motif batik dengan fitur tekstur dengan menggunakan algoritma Fuzzy C-Means [2]; (2) perbandingan algoritma Fuzzy C-Means dan algoritma DBSCAN untuk pengelompokkan pelanggan retail [3]; pengelompokkan dosen pembimbing tugas akhir berdasarkan profil dosen dan bidang keahliannya dengan menggunakan Fuzzy C-Means [4]; dan perbandingan algoritma Fuzzy C-Means dan K-Means untuk klasterisasi kedisiplinan kinerja pegawai [5].

2 Metodologi

Clustering atau klasterisasi adalah metode pengelompokan data. Menurut Tan, 2006 clustering adalah sebuah proses untuk mengelompokkan data ke dalam beberapa cluster atau kelompok sehingga data dalam satu cluster memiliki tingkat kemiripan yang maksimum dan data antar cluster memiliki kemiripan yang minimum. Metode Clustering yang digunakan dalam penelitian ini adalah Algoritma Fuzzy C-Means.

```

FCM Algorithm :
Input :  $\theta (k,l), N$ 
Output  $U^*_{FCM}, V^*_{FCM}$ 
1. Initialize Partition  $U^{(0)}$  randomly
2. for  $i = 1$  to  $n$ 
3.   for  $k = 1$  to  $c$ 
4.     Repeat for  $j = 1, 2, 3, \dots$ 
5.       Update centroid  $V^{(j)}$  with  $U^{(j-1)}$  Using (3)
6.       Compute Distance  $D^{(j)}$  with  $V^{(j)}$ 
7.       Update Partition Matrix  $U^{(j)}$  with  $D^{(j)}$  using (5)
8.     Until  $\|U^{(j)} - U^{(j-1)}\| < \epsilon$ 
9.   end
10. end
11. Return  $U^*_{FCM} \leftarrow U^{(j)}$  and  $V^*_{FCM} \leftarrow V^{(j)}$ 
    
```

Gambar 1 Pseudocode Algoritma C-Means

Algoritma Fuzzy C-Means adalah salah satu algoritma Clustering. Fuzzy C-Means (FCM) adalah suatu teknik pengclusteringan data yang mana keberadaan tiap-tiap data dalam suatu cluster ditentukan oleh derajat keanggotaan [6]–[8]. Teknik ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981 [9]. Output dari FCM bukan merupakan fuzzy interface system, melainkan merupakan deret pusat cluster dan beberapa derajat keanggotaan untuk setiap titik data. Algoritma FCM dalam bentuk Pseudocode terdapat pada Gambar 1.

Pada penelitian kali ini disini kita menggunakan salah satu metode dari data mining, yaitu menggunakan algoritma FCM (Fuzzy C-Means), algoritma FCM ini fungsi utamanya untuk mengcluster. Pada artikel ini disajikan untuk mengetahui clustering pada penggunaan kartu kredit. Pada tahapan dalam coding, ada preprocessing, data training dan visualisasi data. Kami memulainya dengan preprocessing yaitu awalnya data itu berupa data mentah selanjutnya kita bersihkan dan nanti hasilnya akan menjadi data yang siap untuk diolah dengan data mining. Selanjutnya data training pada tahap ini digunakan untuk melatih algoritma mencari model yang cocok yang menghasilkan suatu model machine learning. Dan yang terakhir tahap visualisasi data pada proses ini menemukan kandungan yang tersembunyi sehingga mendapatkan hasil clustering penjualan handphone sesuai yang diharapkan menyajikan hasil visual dari data.

3 Hasil dan Pembahasan

3.1 Pengumpulan Data

Data yang kami pakai merupakan data kuantitatif yaitu berisi kumpulan angka customer id, balance, balance frequency, payment dan lainnya. Data ini merupakan data mentah yang harus melalui tahap preprocessing data terlebih dahulu. Data mentah perlu dibersihkan dan dikodekan menjadi nilai numerik sebelum diberikan ke model pembelajaran mesin, proses pembersihan dan pengkodean ini disebut sebagai teks preprocessing. Oleh karena itu, diperlukan proses pengubahan bentuk menjadi data yang terstruktur untuk kebutuhan lebih lanjut (sentiment analysis, topic modelling, dan lain-lain).

3.2 Pre-processing Data

Gambar 2 menunjukkan proses pre-processing data yang memastikan tidak ada missing value.

```
df["MINIMUM_PAYMENTS"].fillna(0, inplace = True)
df
```

CUST_ID	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	CASH_ADVANCE	PURCHASES_FREQUENCY	ONEOFF_PURC
0	C10001	40.900749	0.818182	95.40	0.00	95.40	0.000000	0.169867
1	C10002	3202.467416	0.909091	0.00	0.00	6442.945403	0.000000	0.000000
2	C10003	2495.148892	1.000000	773.17	773.17	0.00	0.000000	1.000000
3	C10004	1696.670542	0.636364	1499.00	1499.00	0.00	205.788017	0.083333
4	C10005	817.714335	1.000000	16.00	16.00	0.00	0.000000	0.083333
...
8948	C19196	28.462517	1.000000	291.12	0.00	291.12	0.000000	1.000000
8948	C19187	19.163215	1.000000	300.00	0.00	300.00	0.000000	1.000000
8947	C19188	23.388873	0.833333	144.40	0.00	144.40	0.000000	0.833333
8948	C19189	13.457564	0.833333	0.00	0.00	36.556778	0.000000	0.000000
8948	C19190	372.708075	0.696969	1093.25	1093.25	0.00	127.040008	0.696967

```
df.sum()
```

```
CUST_ID          0
BALANCE          0
BALANCE_FREQUENCY 0
PURCHASES        0
ONEOFF_PURCHASES 0
INSTALLMENTS_PURCHASES 0
PURCHASES_FREQUENCY 0
CASH_ADVANCE     0
PURCHASES_FREQUENCY 0
ONEOFF_PURCHASES_FREQUENCY 0
PURCHASES_INSTALLMENTS_FREQUENCY 0
CASH_ADVANCE_FREQUEN... 0
CASH_ADVANCE_TXR 0
PURCHASES_TXR    0
CREDIT_LIMIT     0
PAYMENTS         0
MINIMUM_PAYMENTS 0
PRC_FULL_PAYMENT 0
TENURE           0
dtype: int64
```

```
df.ising = df[['BALANCE', 'PURCHASES', 'ONEOFF_PURCHASES', 'CREDIT_LIMIT', 'PAYMENTS', 'MINIMUM_PAYMENTS']]
X = (op.asarray(df.ising))
```

Gambar 2 Pre-processing Data

3.3 Implementasi Algoritma

Setelah selesai processing data selanjutnya kita masukkan jumlah cluster yang kita inginkan, disini kelompok kita memasukkan 3 cluster, $m=2$, error 0,005 dan iterasi 10 (terdapat pada Gambar 3).

```

[ ] fcm = FCM(n_clusters=3, m=2, error=0.005, max_iter=10)
    fcm.fit(X)
    fcm_centers = fcm.centers
    fcm_labels = fcm.u.argmax(axis=1)

print(fcm_centers)
print(fcm_labels)

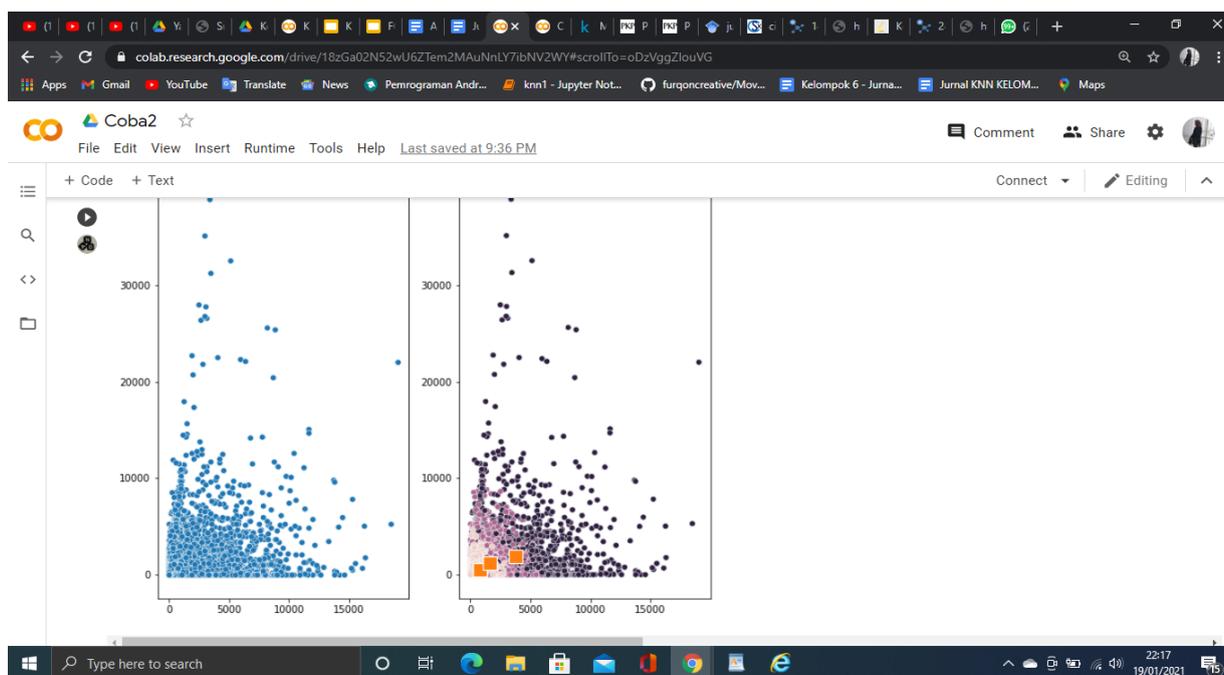
[[ 754.5901  462.2795  222.11754 2098.5378  815.9255  442.1327 ]
 [1644.3856 1216.6401  721.37274 5647.649  1974.5881  831.8733 ]
 [3762.9329 1852.9008 1188.1357 9735.11  3331.443  1515.4458 ]
 [0 1 1 ... 0 0 0]]

[ ] f, axes, = plt.subplots(1,2, figsize=(10,10))
    scatter(X[:,0], X[:,1], ax = axes[0])
    scatter(X[:,0], X[:,1], ax = axes[1], hue=fcm_labels)
    scatter(fcm_centers[:,0], fcm_centers[:,1], ax=axes[1], marker="s", s=200)
    plt.show()

/usr/local/lib/python3.6/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following variables as keyword args: x, y. From version
FutureWarning
/usr/local/lib/python3.6/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following variables as keyword args: x, y. From version
FutureWarning
    
```

Gambar 3 Implementasi Algoritma Fuzzy C-Means

Sehingga akan keluar hasil clusteringnya, dari gambar dibawah terdapat gambar kiri dan kanan. Pada gambar kiri data sebelum di cluster dan di sebelah kanan data sesudah di cluster, disitu terdapat tiga warna pada tiap pixel yang menunjukkan perbedaan warna itu dari tiga cluster. Hasil klusterisasi dari implementasi algoritma Fuzzy C-Means divisualisasikan pada Gambar 4.



Gambar 4 Visualisasi Hasil Kluster

4 Simpulan

Berdasarkan penelitian kami menggunakan pengklasteran penggunaan kartu kredit dengan algoritma Fuzzy C-Means. Penelitian ini menghasilkan kluster mengenai rekomendasi kartu kredit mana yang sering dipakai oleh banyak orang dengan membagi menjadi 3 kluster. Untuk penelitian berikutnya perlu melakukan analisis atau interpretasi lebih dalam terhadap hasil kluster yang dihasilkan.

Referensi

- [1] H. Jiawei, M. Kamber, J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2006.
- [2] A. A. Pratama, N. Suciati, and D. Purwitasari, "Implementasi Fuzzy C-Means untuk Pengelompokan Citra Batik Berdasarkan Motif dengan Fitur Tekstur," *J. Tek. POMITS Vol. 1, No. 1, 2012*, 2012.
- [3] M. I. Chanafi, D. P. Hapsari, R. K. Hapsari, and T. Indriyani, "Implementasi Algoritma Clustering Untuk Pengelompokan Pelanggan Retail Berdasarkan Skor Recency, Frequency, Dan Monetary," *Pros. Semin. Nas. Sains dan Teknol. Terap.*, vol. 1, no. 1, pp. 797–810, 2019, [Online]. Available: <https://ejurnal.itats.ac.id/sntekpan/article/view/783>.
- [4] Y. S. Siregar and P. Harliana, "Analisis perancangan algoritma fuzzy c-means dalam menentukan dosen pembimbing tugas akhir," *J. Penelit. Tek. Inform.*, vol. 3, no. 1, pp. 181–185, 2019.
- [5] N. Agustina and Prihandoko, "Perbandingan Algoritma K-Means Dengan Algoritma Fuzzy C-Means Untuk Clustering Tingkat Kedisiplinan Kinerja Karyawan," 2018. doi: <https://doi.org/10.29207/resti.v2i3.492>.
- [6] J. Tamaela, E. Sedyono, and A. Setiawan, "Cluster Analysis Menggunakan Algoritma Fuzzy C-means dan K-means Untuk Klasterisasi dan Pemetaan Lahan Pertanian di Minahasa Tenggara," *J. Buana Inform.*, vol. 8, no. 3, 2017, doi: 10.24002/jbi.v8i3.1317.
- [7] R. Shugara, E. Ernawati, and D. Andreswari, "IMPLEMENTASI ALGORITMA FUZZY C - MEANS CLUSTERING DAN SIMPLE ADDITIVE WEIGHTING DALAM PEMBERIAN BANTUAN PROGRAM PENINGKATAN KUALITAS KAWASAN PERMUKIMAN," *Pseudocode*, vol. 3, no. 2, pp. 91–97, 2017, doi: 10.33369/pseudocode.3.2.91-97.
- [8] N. Muhandi, "PENENTUAN PENERIMA BEASISWA DENGAN ALGORITMA FUZZY C-MEANS DI UNIVERSITAS MEGOW PAK TULANG BAWANG," *J. Teknol. Inf. Magister Darmajaya*, vol. 1, no. 02, pp. 158–174, 2015.
- [9] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Comput. Geosci.*, vol. 10, no. 2–3, pp. 191–203, 1984, doi: 10.1016/0098-3004(84)90020-7.