

Teknik Clustering Menggunakan Pendekatan Algoritma K-Means Pada Data Disbursement Fintech

Muhammad Shabri Arrahim Mardi

Alamat : Jl. Perintis No. 46A, Baubau, Sulawesi Tenggara
e-mail: shabri@student.telkomuniversity.ac.id

ABSTRAK

Tulisan ini dimaksudkan untuk melakukan pengelompokan data *Disbursement Fintech* dengan menggunakan teknik *clustering* menggunakan pendekatan algoritma *K-Means*. Sehingga akan diketahui *region – region* yang berpotensi untuk dibukakan cabang *fintech* baru yang dibagi menjadi tiga kelompok yaitu 1 = *low* (kurang potensial), 2 = *medium* (potensial), dan 3 = *high* (sangat potensial).

Kata kunci: Algoritma K-Means, Disbursement Fintech, Data Mining, Fintech.

1. PENDAHULUAN

1.1 Latar Belakang Masalah

Data *disbursement* yang ada pada *fintech* semakin lama semakin banyak jumlahnya, hal tersebut sangat disayangkan jika tidak dimanfaatkan dengan baik karena hanya akan menjadi kuburan data, alangkah lebih baik jika data tersebut dapat dimanfaatkan untuk mencari informasi yang bermanfaat untuk strategi pembuatan cabang baru. Pemahaman yang baik terhadap data pertumbuhan pelanggan pada tiap *branch* (cabang) dapat digunakan perusahaan *fintech* untuk berinvestasi dengan membuat cabang baru pada *region – region* yang dianggap memiliki potensi. Masalah yang mungkin dihadapi adalah kesulitan dalam menganalisa *region – region* mana saja yang memiliki nilai pertumbuhan pelanggan yang baik dan memiliki potensi untuk dibuatkan cabang yang baru.

Segmentasi *region* adalah metode yang dapat digunakan untuk memilih *region* mana saja yang tepat untuk dibuatkan cabang yang baru. Dengan segmentasi *region* berdasarkan pertumbuhan pelanggannya, kita dapat menargetkan *region* mana yang dapat dibuatkan cabang *fintech* yang baru. Namun untuk menganalisa data dalam jumlah besar memerlukan tenaga dan waktu yang banyak. Clustering data untuk menentukan *region – region* mana saja yang memiliki potensi untuk dibuatkan cabang baru dirasa penting karena untuk meningkatkan pelayanan kepada customer.

1.2 Tujuan

Tujuan dari penelitian ini adalah untuk melakukan pengelompokan data *disbursement fintech* dengan menggunakan teknik clustering. Sehingga akan diketahui *region – region* yang berpotensi untuk dibukakan cabang *fintech* baru yang dibagi menjadi tiga kelompok yaitu 1 = *low* (kurang potensial), 2 = *medium* (potensial), dan 3 = *high* (sangat potensial).

1.3 Deskripsi Data

Data yang digunakan pada penelitian ini adalah data *disbursement fintech*. Jumlah *record* dari data tersebut adalah 3703 dengan 11 *variable*. Dataset *disbursement fintech* terdiri dari beberapa atribut antara lain *branch* bertipe *String*, *cutoff_date* bertipe *Date*, *area* bertipe *String*, *region* bertipe *String*, *first_date_disbursement* bertipe *Date*, *active_borrowers* bertipe *Integer*, *active_agent* bertipe *Integer*, *delinquency_rate* bertipe *Numeric*, *outstanding* bertipe *Integer*, *weekly_disbursement* bertipe *Integer*, dan *weekly_new_borrower_per_bp* bertipe *Integer*. Hasil evaluasi terhadap kualitas data adalah masih terdapat data yang memiliki noise dan ditemukan banyak nilai kosong atau null yang disebut sebagai missing value.

- Contoh data bernilai range :

delinquency_rate

1%-2.5%

- Contoh data dengan nilai *Null/Missing Value* dan mempunyai *noise*:

4	2,100	-	NaN
---	-------	---	-----

2. METODE

2.1 Praproses

Pada penelitian ini terdapat 2 pra-proses yang dilakukan yaitu :

- Pembersihan Data (*Data Cleaning*)
Digunakan untuk menghapus noise dan memperbaiki data yang tidak konsisten. Karena pada dataset masih banyak terdapat *noise* dan

missing value maka dilakukan Pembersihan Data (*Data Cleaning*) untuk menangani hal tersebut.

- **Reduksi Data (*Data Reduction*)**

Digunakan untuk me-*resize* data dengan agregasi, eliminasi fitur yang sama, atau pengelompokan data. Karena pada dataset terdapat atribut yang tidak terlalu berpengaruh dalam pembentukan kluster, sehingga atribut tersebut harus dieliminasi dan dilakukan pengurangan jumlah *record* data sebab dari total 3703 record data hanya diambil 250 record untuk tiap – tiap *region* yang ada, dengan menggunakan metode sampling kuota (*Nonprobability sampling*). Karena terdapat 5 region pada dataset, jadi jumlah *record* yang digunakan adalah $5 \times 250 = 1250$ record data.

Berikut adalah potongan data yang telah melalui pra-proses :

	region	active_borrowers	active_agent	weekly_new_borrower_per_bp
0	REGION_JR1	1108	5	3
1	REGION_JH	1515	7	1
2	REGION_JH	1092	5	7
3	REGION_JR2	683	3	17
4	REGION_JR1	2294	8	7
5	REGION_JR1	54	2	4

Figure 1 Dataset hasil pre-proses

2.2 Analisis Pemilihan Algoritma

Penelitian ini menggunakan algoritma *K-means* untuk menyelesaikan masalah yang sudah dijabarkan sebelumnya, *K-means* merupakan suatu algoritma pengklasteran yang cukup sederhana yang mempartisi dataset kedalam beberapa kluster *k*. Algoritmanya cukup mudah untuk diimplementasi dan dijalankan, relatif cepat, mudah disesuaikan dan banyak digunakan[1]. Prinsip utama dari teknik ini adalah menyusun *k* buah partisi/pusat massa (*centroid*)/rata-rata (*mean*) dari sekumpulan data. Algoritma *K-means* dimulai dengan pembentukan partisi kluster di awal kemudian secara iteratif partisi kluster ini diperbaiki hingga tidak terjadi perubahan yang signifikan pada partisi kluster[2].

2.3 Analisis Penentuan Parameter

Pemilihan sebagian kecil kelompok data yang mungkin mengandung pola dari permasalahan. Variabel yang digunakan adalah *active_borrowers*, *active_agent*, *weekly_new_borrower_per_bp*. Beberapa variable tersebut dianggap memiliki andil yang besar untuk menentukan apakah dalam suatu region memiliki potensi untuk dibuatkan cabang *fintech* yang baru. Proses penentuan nilai *k* pada algoritma *K-means* menggunakan metode *Elbow*. Berikut ini merupakan grafik pencarian nilai *k* yang dianggap cocok :

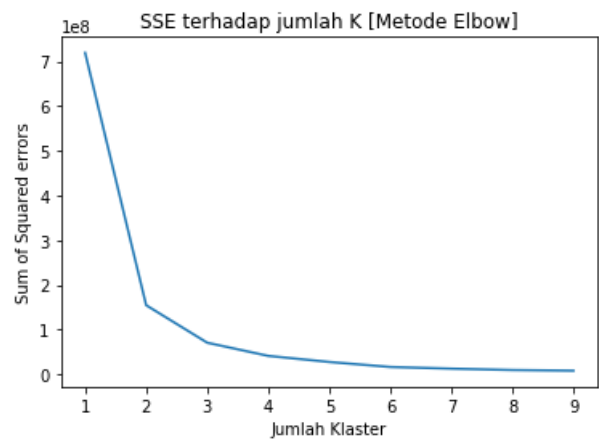


Figure 2 Grafik penentuan nilai *k*

Berdasarkan grafik diatas, dapat dilihat bahwa bentuk *elbow* (siku) terlihat saat jumlah kluster adalah 3. Oleh karena itu, nilai *k* yang baik adalah 3.

3. PEMBAHASAN DAN HASIL

3.1 Hasil Percobaan

Setelah dilakukan percobaan terhadap dataset *disbursement Fintech* dengan menggunakan algoritma *K-means* dengan parameter yang telah disebutkan pada point sebelumnya didapatkan hasil *cluster* sebagai berikut :

	active_borrowers	active_agent	weekly_new_borrower_per_bp
cluster			
1.0	984.1	4.4	11.0
2.0	1012.9	4.3	11.7
3.0	1020.0	4.7	8.8

Figure 3 Hasil cluster dan masing - masing nilai atributnya

Figure 3 memperlihatkan nilai masing – masing atribut berdasarkan klasternya.

Berikut ini merupakan hasil cluster yang didapatkan untuk tiap – tiap *region*. Nilai kluster akhir didapatkan dengan menghitung *Modus* pada atribut '*cluster*' untuk tiap *region*nya.

Class untuk REGION_JH : 3

	region	active_borrowers	active_agent	weekly_new_borrower_per_bp	cluster
0	REGION_JH	1515.0	7.0		1 3.0
1	REGION_JH	1092.0	5.0		7 3.0
2	REGION_JH	1973.0	8.0		4 2.0
3	REGION_JH	1258.0	9.0		4 3.0
4	REGION_JH	314.0	3.0		14 1.0
5	REGION_JH	257.0	2.0		17 1.0
6	REGION_JH	1586.0	7.0		5 3.0
7	REGION_JH	277.0	2.0		18 1.0
8	REGION_JH	146.0	2.0		1 1.0
9	REGION_JH	98.0	1.0		24 1.0

Figure 4 Hasil cluster untuk REGION_JH

Pada Figure 4, dapat dilihat bahwa REGION_JH masuk kedalam klaster 3 (High = Sangat Potensial).

Class untuk REGION_JR1 : 1

	region	active_borrowers	active_agent	weekly_new_borrower_per_bp	cluster
250	REGION_JR1	1108.0	5.0	3	2.0
251	REGION_JR1	2294.0	8.0	7	1.0
252	REGION_JR1	54.0	2.0	4	2.0
253	REGION_JR1	452.0	3.0	16	1.0
254	REGION_JR1	108.0	2.0	15	3.0
255	REGION_JR1	2057.0	8.0	5	1.0
257	REGION_JR1	1914.0	7.0	7	1.0
258	REGION_JR1	1854.0	8.0	4	1.0
259	REGION_JR1	310.0	3.0	30	2.0

Figure 5 Hasil cluster untuk REGION_JR1

Pada Figure 5, dapat dilihat bahwa REGION_JR1 masuk kedalam klaster 1 (Low = Kurang Potensial).

Class untuk REGION_JR2 : 1

	region	active_borrowers	active_agent	weekly_new_borrower_per_bp	cluster
500	REGION_JR2	683.0	3.0	17	1.0
501	REGION_JR2	243.0	2.0	15	2.0
502	REGION_JR2	516.0	3.0	22	2.0
503	REGION_JR2	434.0	3.0	17	1.0
504	REGION_JR2	82.0	2.0	19	1.0
505	REGION_JR2	716.0	3.0	24	2.0
506	REGION_JR2	2089.0	6.0	7	2.0
507	REGION_JR2	604.0	3.0	10	1.0
508	REGION_JR2	685.0	3.0	29	1.0
509	REGION_JR2	2482.0	7.0	3	1.0

Figure 6 Hasil cluster untuk REGION_JR2

Pada Figure 6, dapat dilihat bahwa REGION_JR1 masuk kedalam klaster 1 (Low = Kurang Potensial).

Class untuk REGION_JT : 1

	region	active_borrowers	active_agent	weekly_new_borrower_per_bp	cluster
750	REGIONID_JT	430.0	2.0	19	3.0
751	REGIONID_JT	385.0	2.0	23	3.0
752	REGIONID_JT	533.0	3.0	24	1.0
753	REGIONID_JT	1568.0	7.0	9	3.0
754	REGIONID_JT	1322.0	7.0	10	1.0
755	REGIONID_JT	1370.0	8.0	8	3.0
756	REGIONID_JT	639.0	3.0	12	1.0
757	REGIONID_JT	511.0	3.0	22	1.0
758	REGIONID_JT	1722.0	9.0	7	1.0
759	REGIONID_JT	351.0	3.0	11	3.0

Figure 8 Hasil cluster untuk REGION_JT

Pada Figure 7, dapat dilihat bahwa REGION_JT masuk kedalam klaster 1 (Kurang = Tidak Potensial).

Class untuk REGION_N : 1

	region	active_borrowers	active_agent	weekly_new_borrower_per_bp	cluster
1000	REGIONID_N	235.0	2.0	13	1.0
1001	REGIONID_N	424.0	2.0	6	3.0
1002	REGIONID_N	1740.0	7.0	4	1.0
1003	REGIONID_N	1152.0	8.0	5	1.0
1004	REGIONID_N	411.0	3.0	11	1.0
1005	REGIONID_N	569.0	2.0	28	3.0
1006	REGIONID_N	319.0	2.0	33	3.0
1007	REGIONID_N	295.0	2.0	2	1.0
1008	REGIONID_N	360.0	2.0	12	3.0
1009	REGIONID_N	938.0	5.0	3	1.0

Figure 9 Hasil cluster untuk REGION_N

Pada Figure 8, dapat dilihat bahwa REGION_N masuk kedalam klaster 1 (Low = Kurang Potensial).

Dari hasil klaster masing – masing *region* diatatas, diketahui bahwa *region* yang Sangat berpotensi untuk dibuatkan *branch* (cabang) untuk *fintech* yang baru adalah REGION_JH karena memiliki nilai klaster = 3 (High = Sangat Potensial).

3.2 Ringkasan Model

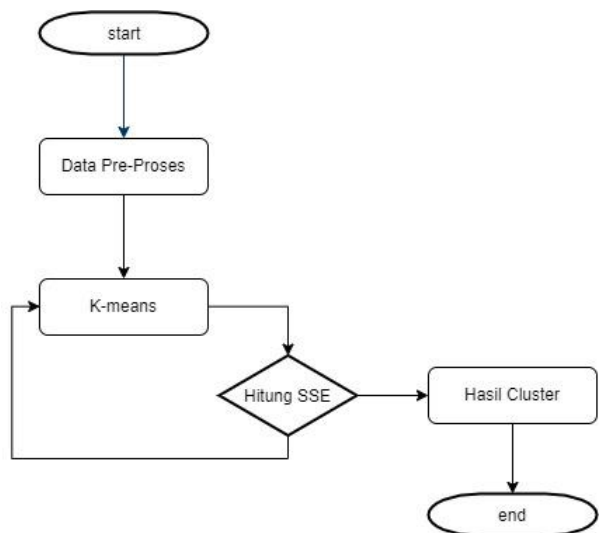


Figure 7 Model Teknik Clustering

3.3 Interpretasi Model

Teknik *clustering* dengan algoritma K-Means seperti model pada Figure 9 dapat membantu pengelompokkan dataset dalam pengambilan keputusan untuk pembuatan branch (cabang) *fintech* pada *region-region* yang dibagi menjadi 3 kelompok yaitu, 1 = low (Kurang Potensial), 2 = medium (Potensial), dan 3 = high (Sangat Potensial). Tujuan dari penelitian ini tercapai dengan menggunakan model pada Figure 9 karena model tersebut berhasil menentukan klaster dari masing – masing *region* yang ada pada dataset.

DAFTAR PUSTAKA

- [1] Wu, Xindong & Kumar, Vipin. (2009). The Top Ten Algorithms in Data Mining. London: CRC Press.
- [2] Witten, I. H., Frank, E., & Hall, M. A. (2011). Data Mining: Practical Machine Learning and Tool. Burlington: Morgan Kaufmann Publisher.