

TUGAS AKHIR

SEGMENTASI PELANGGAN DENGAN ALGORITMA DBSCAN CLUSTERING



Adelia Salsabina

NPM: 6181901029

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2024

FINAL PROJECT

**CUSTOMER SEGMENTATION WITH DBSCAN CLUSTERING
ALGORITHM**



Adelia Salsabina

NPM: 6181901029

**DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2024**

LEMBAR PENGESAHAN

SEGMENTASI PELANGGAN DENGAN ALGORITMA DBSCAN CLUSTERING

Adelia Salsabina

NPM: 6181901029

Bandung, 16 Januari 2024

Menyetujui,

Pembimbing

Digitally signed
by Luciana
Abednego

Luciana Abednego, M.T.

Ketua Tim Penguji

Digitally signed
by Rosa de Lima
E. Padmowati

Rosa De Lima, M.T.

Anggota Tim Penguji

Digitally
signed by
Husnul
Hakim

Husnul Hakim, M.T.

Mengetahui,

Ketua Program Studi

Digitally signed
by Lionov

Lionov, Ph.D.

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa tugas akhir dengan judul:

SEGMENTASI PELANGGAN DENGAN ALGORITMA DBSCAN CLUSTERING

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal 16 Januari 2024



Adelia Salsabina
NPM: 6181901029

ABSTRAK

Semakin banyaknya perusahaan membuat persaingan pasar semakin ketat. Oleh karena itu, perusahaan perlu memahami perkembangan pasar agar dapat bersaing secara efektif. Untuk mencapai hal ini, perusahaan harus mengembangkan strategi bisnis yang tepat dengan memahami kebiasaan dan kebutuhan pelanggan, karena hal ini mempengaruhi kesuksesan perusahaan.

Dalam pengembangan strategi bisnis yang fokus pada kebiasaan dan kebutuhan pelanggan, salah satu pendekatan yang dapat diambil adalah dengan melakukan segmentasi pelanggan. Pelanggan dengan karakteristik serupa dikelompokkan bersama, sementara pelanggan dengan karakteristik yang berbeda dikelompokkan secara terpisah. Metode yang dapat digunakan adalah *clustering*.

Pada tugas akhir ini, segmentasi pelanggan dilakukan menggunakan data dari PT. Jamkrindo cabang Jambi, sebuah perusahaan penjaminan kredit untuk UMKM. Teknik *clustering*, seperti algoritma DBSCAN dan *k*-Means, diterapkan untuk mendapatkan model segmentasi pelanggan. Selanjutnya, model RFM (*Recency, Frequency, Monetary*) diintegrasikan dengan hasil *clustering* untuk membentuk model segmentasi pelanggan yang menyeluruh. Evaluasi model dilakukan menggunakan metode *silhouette index* dan *davies bouldin index*.

Data pelanggan PT. Jamkrindo dibagi menjadi dua jenis, yaitu pelanggan PT. Jamkrindo Perorangan dan PT. Jamkrindo Badan Usaha. Dari pembuatan model, algoritma DBSCAN lebih unggul dibandingkan algoritma *k*-Means. Hal ini disebabkan oleh nilai evaluasi *clustering* yang diperoleh dari data pelanggan PT. Jamkrindo Perorangan. Algoritma DBSCAN menunjukkan nilai *silhouette index* yang lebih tinggi (mendekati 1) dibandingkan dengan algoritma *k*-Means. Nilai *silhouette index* untuk *clustering* data pelanggan PT. Jamkrindo Perorangan menggunakan DBSCAN adalah 0.598, sedangkan dengan *k*-Means adalah 0.550. Sementara itu, untuk data pelanggan PT. Jamkrindo Badan Usaha, pembuatan model menggunakan algoritma DBSCAN menghasilkan kelompok pelanggan yang lebih bermakna dan dapat dibedakan dengan jelas berdasarkan nilai RFM.

Kemudian dilakukan analisis karakteristik lebih mendalam untuk setiap kelompok pelanggan yang terbentuk, dan diberikan rekomendasi strategi bisnis. Implementasi dari hasil penelitian ini diterapkan dalam sebuah perangkat lunak berbasis *website* untuk menampilkan hasil segmentasi pelanggan beserta rekomendasi strategi bisnis.

Kata-kata kunci: Segmentasi Pelanggan, Pengelompokan, Algoritma DBSCAN, RFM.

ABSTRACT

The increasing number of companies intensifies market competition. Therefore, companies need to comprehend market developments to compete effectively. To achieve this, companies must develop the right business strategy by understanding customer habits and needs, as these factors significantly influence a company's success.

In the development of a business strategy focused on customer habits and needs, one approach is to conduct customer segmentation. Customers with similar characteristics are grouped together, while those with different characteristics are separated. The method that can be employed for this purpose is clustering.

In this final project, customer segmentation is carried out using data from PT. Jamkrindo branch in Jambi, a credit guarantee company for Micro, Small, and Medium Enterprises (MSMEs). Clustering techniques, such as DBSCAN and k -Means algorithms, are applied to obtain a customer segmentation model. Subsequently, the RFM (Recency, Frequency, Monetary) model is integrated with clustering results to form a comprehensive customer segmentation model. Model evaluation is conducted using silhouette index and Davies Bouldin index methods.

PT. Jamkrindo customer data is divided into two types: individual customers (PT. Jamkrindo Perorangan) and corporate customers (PT. Jamkrindo Badan Usaha). In the model creation process, the DBSCAN algorithm outperforms the k -Means algorithm. This is due to the evaluation scores obtained from the individual customer data of PT. Jamkrindo. The DBSCAN algorithm shows a higher silhouette index value (approaching 1) compared to the k -Means algorithm. The silhouette index value for clustering the individual customer data using DBSCAN is 0.598, while with k -Means, it is 0.550. Meanwhile, for corporate customer data (PT. Jamkrindo Badan Usaha), the model created using the DBSCAN algorithm results in more meaningful and distinguishable customer clusters based on RFM values.

Furthermore, a more in-depth analysis of the characteristics of each customer group is conducted, and business strategy recommendations are provided. The implementation of this research result is applied in a website-based software to display customer segmentation results along with business strategy recommendations.

Keywords: Customer Segmentation, Clustering, DBSCAN Algorithm, RFM.

Tugas akhir ini dipersembahkan kepada diri sendiri, orang tua, dan keluarga. . .

KATA PENGANTAR

Puji dan syukur penulis ucapkan kepada Tuhan Yang Maha Esa atas berkat-Nya penulis dapat menyelesaikan tugas akhir yang berjudul “Segmentasi Pelanggan dengan Algoritma DBSCAN *Clustering*”. Penulis juga ingin mengucapkan terima kasih kepada berbagai pihak yang telah membantu dan mendukung penulis selama masa perkuliahan sampai dengan pengerjaan tugas akhir. Oleh karena itu, penulis ingin menyampaikan terima kasih kepada :

1. Kepada kedua orang tua, kakak, dan adik penulis yang telah memberikan dukungan, semangat, dan doa selama perkuliahan dan pengerjaan tugas akhir.
2. Kepada pembimbing, Ibu Luciana Abednego, M.T. yang telah memberikan waktu untuk membimbing pengerjaan tugas akhir.
3. Kepada penguji, Ibu Rosa De Lima, M.T. dan Bapak Husnul Hakim, M.T yang telah memberikan saran dan masukan dari hasil tugas akhir.
4. Kepada anggota keluarga lain yang telah memberikan dukungan dan semangat selama perkuliahan dan pengerjaan tugas akhir.
5. Kepada teman-teman seperjuangan penulis selama kuliah terutama Ira, Alma, Kinan, dan Levi yang telah menemani, membantu, dan memberi dukungan kepada penulis selama masa perkuliahan dan pengerjaan tugas akhir.
6. Kepada teman penulis Sindi dan Fairuz yang telah memberi dukungan dan semangat kepada penulis.
7. Kepada seluruh pihak lain yang tidak dapat disebutkan satu per satu.

Dalam pengerjaan dan penulisan tugas akhir tentunya masih banyak kekurangan dan kesalahan. Oleh karena itu, penulis ingin meminta maaf atas kekurangan dan kesalahan yang ada pada tugas akhir ini. Penulis juga ingin berterima kasih kepada para pembaca tugas akhir ini dan semoga bermanfaat bagi para pembaca.

Bandung, Januari 2024

Penulis

DAFTAR ISI

KATA PENGANTAR	xv
DAFTAR ISI	xvii
DAFTAR GAMBAR	xix
DAFTAR TABEL	xxiii
DAFTAR KODE PROGRAM	xxv
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
1.4 Batasan Masalah	3
1.5 Metodologi	3
1.6 Sistematika Pembahasan	3
2 LANDASAN TEORI	5
2.1 PT. Jamkrindo	5
2.2 Segmentasi Pelanggan[1]	7
2.3 <i>Data Science</i> [2]	9
2.3.1 <i>Prior Knowledge</i>	9
2.3.2 <i>Preparation</i>	10
2.3.3 <i>Modeling</i>	12
2.3.4 <i>Application</i>	13
2.3.5 <i>Knowledge</i>	13
2.4 <i>Recency, Frequency, Monetary</i> (RFM)[3]	14
2.5 Teknik <i>Clustering</i> [4]	14
2.5.1 <i>Partitioning Methods</i>	16
2.5.2 <i>Density-Based Methods</i>	18
2.6 Algoritma DBSCAN <i>Clustering</i> [4]	19
2.7 Teknik Evaluasi <i>Clustering</i>	23
2.8 <i>Principal Component Analysis</i> (PCA)	25
2.9 Library sklearn pada Python[5]	26
2.10 Visualisasi Data	27
2.10.1 Tipe Visualisasi	27
2.10.2 Tipe Data	32
3 ANALISIS PENYELESAIAN MASALAH	33
3.1 Deskripsi Masalah	33
3.2 Eksperimen menggunakan Studi Kasus Kecil	33
3.2.1 Data Pengguna <i>Credit Card</i>	34

3.2.2	Data Pelanggan Bank	49
3.3	Eksperimen <i>Tools</i> dan <i>Library</i>	64
4	PENAMBANGAN DATA	69
4.1	Pengumpulan Data	69
4.2	Penyiapan Data	70
4.2.1	Data PT. Jamkrindo Perorangan	70
4.2.2	Data PT. Jamkrindo Badan Usaha	75
4.3	Eksplorasi Data	78
4.3.1	Data PT. Jamkrindo Perorangan	78
4.3.2	Data PT. Jamkrindo Badan Usaha	84
4.4	Ekstraksi Fitur	90
4.4.1	Data PT. Jamkrindo Perorangan	90
4.4.2	Data PT. Jamkrindo Badan Usaha	93
4.5	Pembuatan Model dan Evaluasi	97
4.5.1	Data PT. Jamkrindo Perorangan	97
4.5.2	Data PT. Jamkrindo Badan Usaha	99
4.6	Analisis Hasil Model	101
4.6.1	Data PT. Jamkrindo Perorangan	101
4.6.2	Data PT. Jamkrindo Badan Usaha	110
4.7	Rekomendasi Strategi Bisnis	116
4.7.1	Data PT. Jamkrindo Perorangan	116
4.7.2	Data PT. Jamkrindo Badan Usaha	118
5	IMPLEMENTASI MODEL DAN PENGUJIAN	123
5.1	Pembangunan Perangkat Lunak	123
5.1.1	Deskripsi Perangkat Lunak	123
5.1.2	Tampilan Antarmuka Perangkat Lunak	124
5.1.3	Implementasi Perangkat Lunak	127
5.2	Pengujian Fungsionalitas Perangkat Lunak	129
6	KESIMPULAN DAN SARAN	131
6.1	Kesimpulan	131
6.2	Saran	131
	DAFTAR REFERENSI	133
	A KODE PROGRAM	135

DAFTAR GAMBAR

1.1	Segmentasi Pelanggan berdasarkan Karakteristik Tertentu	1
2.1	Segmentasi Pelanggan berdasarkan berbagai Karakteristik	8
2.2	Deskripsi RFM Cell	9
2.3	Proses <i>data science</i>	10
2.4	<i>Scatterplot</i> untuk <i>credit score vs loan interest rate</i>	11
2.5	Tahapan pemodelan prediktif	12
2.6	<i>Scatterplot</i> untuk data <i>training</i> dan <i>testing</i>	13
2.7	Inisialisasi <i>clustering</i>	17
2.8	Proses iterasi	18
2.9	Hasil <i>clustering</i>	18
2.10	<i>Cluster</i> dengan bentuk <i>arbitrary</i>	19
2.11	<i>p</i> <i>directly density-reachable</i> dari <i>q</i>	20
2.12	<i>p</i> <i>density-reachable</i> dari <i>q</i>	20
2.13	<i>p</i> dan <i>q</i> <i>density-connected</i> oleh <i>o</i>	21
2.14	Contoh grafik <i>elbow method</i>	25
2.15	Contoh plot PCA	26
2.16	<i>Bar chart</i> berbentuk horizontal	28
2.17	<i>Bar chart</i> berbentuk vertikal	28
2.18	<i>Pie chart</i>	28
2.19	Istilah pada <i>box plot</i>	29
2.20	Contoh <i>box plot</i>	29
2.21	Ilustrasi <i>box plot</i> dan <i>violin plot</i>	30
2.22	Contoh <i>violin plot</i>	30
2.23	Contoh <i>heatmap</i>	30
2.24	Contoh <i>histogram</i>	31
2.25	<i>Line chart</i>	31
2.26	Contoh <i>parallel coordinate plot</i>	32
3.1	Plot korelasi data pelanggan kartu kredit	36
3.2	Persebaran ‘BALANCE’	38
3.3	Persebaran ‘BALANCE_FREQUENCY’	39
3.4	Persebaran ‘PURCHASES’	39
3.5	Persebaran ‘PURCHASES_FREQUENCY’	40
3.6	Persebaran ‘PURCHASES_TRX’	40
3.7	Persebaran ‘CASH_ADVANCE’	41
3.8	Persebaran ‘CASH_ADVANCE_FREQUENCY’	41
3.9	Persebaran ‘CASH_ADVANCE_TRX’	42
3.10	Persebaran ‘ONEOFF_PURCHASES’	42
3.11	Persebaran ‘ONEOFF_PURCHASES_FREQUENCY’	43
3.12	Persebaran ‘INSTALLMENTS_PURCHASES’	43
3.13	Persebaran ‘PURCHASES_INSTALLMENTS_FREQUENCY’	44
3.14	Persebaran ‘CREDIT_LIMIT’	44

3.15	Persebaran ‘PAYMENTS’	45
3.16	Persebaran ‘MINIMUM_PAYMENTS’	45
3.17	Persebaran ‘PRC_FULL_PAYMENT’	46
3.18	Persebaran ‘TENURE’	46
3.19	Visualisasi Hasil <i>Cluster</i>	47
3.20	Plot korelasi data pelanggan bank	51
3.21	Persebaran <i>age</i>	52
3.22	Persebaran <i>balance</i>	52
3.23	Persebaran <i>day</i>	53
3.24	Persebaran <i>duration</i>	53
3.25	Persebaran <i>campaign</i>	54
3.26	Persebaran <i>pdays</i>	54
3.27	Persebaran <i>previous</i>	55
3.28	Persebaran <i>job</i>	55
3.29	Persebaran <i>marital</i>	56
3.30	Persebaran <i>education</i>	56
3.31	Persebaran <i>default</i>	57
3.32	Persebaran <i>housing</i>	57
3.33	Persebaran <i>loan</i>	58
3.34	Persebaran <i>contact</i>	58
3.35	Persebaran <i>month</i>	59
3.36	Persebaran <i>poutcome</i>	59
3.37	Persebaran <i>y</i>	60
3.38	Visualisasi Hasil <i>Cluster</i>	61
3.39	Persebaran data pelanggan mall	65
3.40	Jumlah <i>cluster</i> dengan kenaikan ϵ	65
3.41	Jumlah <i>cluster</i> dengan kenaikan ϵ	66
3.42	Jumlah <i>cluster</i> dengan kenaikan <i>min_sample</i>	66
3.43	Jumlah <i>cluster</i> dengan kenaikan <i>min_sample</i>	67
3.44	Jumlah <i>cluster</i> dengan kenaikan <i>min_sample</i> dan ϵ bersamaan	67
3.45	Nilai <i>silhouette index</i> untuk berbagai <i>metric</i>	68
4.1	Perbandingan distribusi nilai <i>recency</i> data asli dan data sampel	73
4.2	Perbandingan distribusi nilai <i>frequency</i> data asli dan data sampel	73
4.3	Perbandingan distribusi nilai <i>monetary</i> data asli dan data sampel	74
4.4	Plot korelasi data pelanggan PT. Jamkrindo Perorangan	78
4.5	Persebaran ‘Usia’	79
4.6	Persebaran ‘Penjaminan/Kafalah’	79
4.7	Persebaran ‘IJP/IJK’	80
4.8	Persebaran ‘Feebase’	80
4.9	Persebaran ‘Feeagen’	81
4.10	Persebaran ‘Santunan Kebakaran’	81
4.11	Persebaran ‘Produk’	82
4.12	Persebaran ‘Bank/Lembaga Keuangan’	82
4.13	Persebaran ‘Peruntukan’	82
4.14	Persebaran ‘Nama Agen’	83
4.15	Persebaran ‘Jenis KUR’	83
4.16	Persebaran ‘Tahun’	83
4.17	Persebaran ‘Waktu’	84
4.18	Plot korelasi data pelanggan PT. Jamkrindo Badan Usaha	84
4.19	Persebaran ‘Penjaminan/Kafalah’	85
4.20	Persebaran ‘IJP/IJK’	86

4.21	Persebaran ‘Feeagen’	86
4.22	Persebaran ‘Produk’	87
4.23	Persebaran ‘Bank/Lembaga Keuangan’	87
4.24	Persebaran ‘Peruntukan’	87
4.25	Persebaran ‘Nama Agen’	88
4.26	Persebaran ‘Jenis KUR’	88
4.27	Persebaran ‘Tahun’	88
4.28	Persebaran ‘Waktu’	89
4.29	Distribusi <i>recency</i>	91
4.30	Distribusi <i>monetary</i>	92
4.31	Distribusi <i>frequency</i>	94
4.32	Distribusi <i>recency</i>	94
4.33	Distribusi <i>monetary</i>	95
4.34	Distribusi <i>frequency</i>	97
4.35	Visualisasi Hasil <i>Cluster</i>	98
4.36	Visualisasi Hasil <i>Cluster</i>	99
4.37	Visualisasi Hasil <i>Cluster</i>	100
4.38	Visualisasi Hasil <i>Cluster</i>	101
4.39	Distribusi <i>recency</i> tiap <i>cluster</i>	108
4.40	Distribusi <i>frequency</i> tiap <i>cluster</i>	108
4.41	Distribusi <i>monetary</i> tiap <i>cluster</i>	109
4.42	Visualisasi Hasil <i>Cluster</i>	110
4.43	Distribusi <i>recency</i> tiap <i>cluster</i>	114
4.44	Distribusi <i>frequency</i> tiap <i>cluster</i>	114
4.45	Distribusi <i>monetary</i> tiap <i>cluster</i>	114
4.46	Visualisasi Hasil <i>Cluster</i>	116
4.47	Persebaran <i>monetary</i> pelanggan <i>cluster 3</i>	117
4.48	Persebaran <i>monetary</i> pelanggan <i>cluster 4</i>	118
4.49	Persebaran <i>monetary</i> pelanggan <i>cluster 5</i>	118
4.50	Persebaran <i>monetary</i> pelanggan <i>cluster 1</i>	119
4.51	Persebaran <i>monetary</i> pelanggan <i>cluster 2</i>	120
4.52	Persebaran <i>monetary</i> pelanggan <i>cluster 3</i>	120
4.53	Persebaran <i>monetary</i> pelanggan <i>cluster 4</i>	121
4.54	Persebaran <i>monetary</i> pelanggan <i>cluster 5</i>	122
5.1	<i>Flow chart</i> perangkat lunak	124
5.2	Antarmuka halaman utama	125
5.3	Antarmuka halaman hasil segmentasi	125
5.4	Antarmuka halaman hasil segmentasi	126
5.5	Antarmuka halaman hasil segmentasi	126
5.6	Halaman utama	127
5.7	Halaman hasil segmentasi	128
5.8	Halaman deskripsi <i>cluster</i>	128
5.9	Halaman <i>insight</i>	129
5.10	Halaman hasil untuk data PT. Jamkrindo Perorangan	130
5.11	Halaman hasil untuk data PT. Jamkrindo Badan Usaha	130

DAFTAR TABEL

2.1	Karakteristik metode <i>clustering</i>	19
3.1	Deskripsi atribut dari data 'Customer_Data.csv'	34
3.2	Jumlah anggota tiap <i>cluster</i>	47
3.3	Nilai <i>Mean</i>	48
3.4	Nilai <i>Min</i>	48
3.5	Nilai <i>Max</i>	48
3.6	Deskripsi atribut dari data 'bankfull.csv'	50
3.7	Jumlah anggota tiap <i>cluster</i>	61
3.8	Nilai <i>Mean</i>	61
3.9	Nilai <i>Min</i>	62
3.10	Nilai <i>Max</i>	62
3.11	Jumlah setiap kategori fitur per <i>cluster</i>	62
3.12	Nilai <i>Mean</i>	63
3.13	Nilai <i>Min</i>	64
3.14	Nilai <i>Max</i>	64
4.1	Deskripsi atribut data pelanggan PT. Jamkrindo	69
4.2	Perubahan 'Waktu'	71
4.3	Perbandingan nilai data keseluruhan dan data sampel	74
4.4	Deskripsi atribut data pelanggan PT. Jamkrindo Perorangan	75
4.5	Perubahan 'Waktu'	75
4.6	Deskripsi atribut data pelanggan PT. Jamkrindo Badan Usaha	77
4.7	Perbedaan Nilai PT. Jamkrindo Perorangan dan PT. Jamkrindo Badan Usaha	89
4.8	Deskripsi nilai RFM	90
4.9	Jumlah anggota tiap <i>cluster</i>	98
4.10	Jumlah anggota tiap <i>cluster</i>	98
4.11	Jumlah anggota tiap <i>cluster</i>	100
4.12	Jumlah anggota tiap <i>cluster</i>	100
4.13	Nilai <i>Mean</i>	101
4.14	Nilai <i>Min</i>	102
4.15	Nilai <i>Max</i>	102
4.16	Jumlah setiap kategori 'Produk' per <i>cluster</i>	103
4.17	Keterangan untuk kategori 'Produk'	103
4.18	Jumlah setiap kategori 'Bank/Lembaga Keuangan per <i>cluster</i>	103
4.19	Keterangan untuk kategori 'Bank/Lembaga Keuangan'	104
4.20	Jumlah setiap kategori 'Peruntukan' per <i>Cluster</i>	104
4.21	Keterangan untuk kategori 'Peruntukan'	104
4.22	Jumlah setiap kategori 'Nama Agen' per <i>cluster</i>	105
4.23	Keterangan untuk kategori 'Nama Agen'	105
4.24	Jumlah setiap kategori 'Jenis KUR' per <i>cluster</i>	105
4.25	Keterangan untuk kategori 'Jenis KUR'	105
4.26	Jumlah setiap kategori 'Tahun' per <i>cluster</i>	106

4.27	Keterangan untuk kategori ‘Tahun’	106
4.28	Jumlah setiap kategori ‘Waktu’ per <i>cluster</i>	106
4.29	Keterangan untuk kategori ‘Waktu’	107
4.30	Pembagian kategori nilai <i>Recency</i>	109
4.31	Pembagian kategori nilai <i>Frequency</i>	109
4.32	Pembagian kategori nilai <i>Monetary</i>	109
4.33	Interpretasi Hasil <i>Cluster</i>	110
4.34	Nilai <i>Mean</i>	110
4.35	Nilai <i>Min</i>	111
4.36	Nilai <i>Max</i>	111
4.37	Jumlah setiap kategori fitur per <i>Cluster</i>	111
4.38	Keterangan untuk Tabel 4.37	112
4.39	Pembagian kategori nilai <i>Recency</i>	115
4.40	Pembagian kategori nilai <i>Frequency</i>	115
4.41	Pembagian kategori nilai <i>Monetary</i>	115
4.42	Interpretasi Hasil <i>Cluster</i>	115

DAFTAR KODE PROGRAM

2.1	Potongan kode untuk <i>clustering</i>	26
2.2	Potongan kode untuk prepemrosesan data	26
2.3	Potongan kode untuk melakukan evaluasi model	27
2.4	Potongan kode untuk menyimpan model ke pickle	27
3.1	Potongan kode untuk mengubah kategorik menjadi numerik	50
4.1	Kode untuk menghitung <i>recency</i>	90
4.2	Kode untuk menghitung <i>monetary</i>	91
4.3	Kode untuk menghitung <i>frequency</i> dan penjumlahan <i>monetary</i>	92
4.4	Kode untuk menghitung <i>recency</i>	94
4.5	Kode untuk menghitung <i>monetary</i>	95
4.6	Kode untuk menghitung <i>frequency</i> dan penjumlahan <i>monetary</i>	96
A.1	Kode untuk eksplorasi parameter DBSCAN	135
A.2	Kode untuk menyimpan fungsi yang digunakan pada pemodelan	137
A.3	Kode pemodelan data pengguna credit card	138
A.4	Kode pemodelan data pelanggan bank	141
A.5	Kode pemodelan data pelanggan PT. Jamkrindo Perorangan	145
A.6	Kode pemodelan data pelanggan PT. Jamkrindo Badan Usaha	154
A.7	Kode CSS untuk halaman hasil	162
A.8	Kode CSS untuk halaman deskripsi <i>cluster</i>	163
A.9	Kode CSS untuk halaman <i>insight</i>	164
A.10	<i>Script</i> untuk pemodelan data pelanggan PT. Jamkrindo Perorangan	164
A.11	Kode tampilan halaman utama perangkat lunak	168
A.12	Kode tampilan halaman hasil PT. Jamkrindo Perorangan	169
A.13	Kode tampilan halaman hasil <i>cluster</i> 1 PT. Jamkrindo Perorangan namun juga diaplikasikan pada <i>cluster</i> lainnya	170
A.14	Kode tampilan halaman <i>insight</i> PT. Jamkrindo Perorangan	172
A.15	<i>Script</i> untuk pemodelan data pelanggan PT. Jamkrindo Badan Usaha	173
A.16	Kode tampilan halaman hasil PT. Jamkrindo Badan Usaha	176
A.17	Kode tampilan halaman hasil <i>cluster</i> 1 PT. Jamkrindo Badan Usaha namun juga diaplikasikan pada <i>cluster</i> lainnya	177
A.18	Kode tampilan halaman <i>insight</i> PT. Jamkrindo Badan Usaha	180

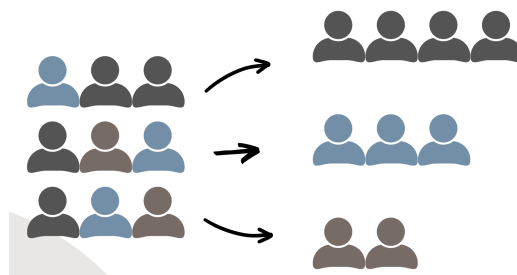
BAB 1

PENDAHULUAN

1.1 Latar Belakang

Perusahaan yang semakin banyak membuat persaingan pasar semakin ketat. Hal tersebut mengharuskan perusahaan untuk selalu memahami pasar yang terus berkembang. Setiap perusahaan memiliki pelanggan yang dapat membantu dalam kemajuan perusahaan dan berdampak pada pertambahan jumlah data transaksi maupun data pelanggan dari perusahaan. Dengan adanya data-data dari pelanggan tersebut dapat diolah untuk menciptakan strategi bisnis yang dapat menjaga hubungan dengan pelanggan dan juga ikut mengembangkan perusahaan. Untuk menciptakan strategi bisnis yang tepat, penting untuk perusahaan memahami pelanggan beserta kebiasaan dan kebutuhannya karena kebiasaan pelanggan merupakan faktor penentu untuk kesuksesan dan kegagalan dari sebuah perusahaan[6]. Setelah mengenali pelanggan, dapat dibentuk strategi bisnis yang sesuai dengan karakteristiknya.

Untuk dapat mengenali kebiasaan pelanggan dan karakteristiknya dapat dilakukan segmentasi pelanggan. Pada segmentasi pelanggan membagi pelanggan yang heterogen menjadi beberapa kelompok berdasarkan berbagai karakteristik seperti kebutuhan, perilaku konsumsi, geografis (kota, lokasi, negara), dan demografis (usia, jenis kelamin, pendidikan, pekerjaan). Pelanggan dengan kemiripan karakteristik akan bergabung pada kelompok yang sama dan pelanggan dengan karakteristik yang berbeda akan bergabung pada kelompok yang berbeda seperti pada ilustrasi Gambar 1.1. Berdasarkan hasil dari *clustering* dapat dibentuk strategi bisnis yang sesuai berdasarkan karakteristik dari setiap kelompok yang terbentuk.



Gambar 1.1: Segmentasi Pelanggan berdasarkan Karakteristik Tertentu

Teknik yang dapat digunakan untuk melakukan segmentasi pelanggan yaitu teknik *clustering*. Teknik *clustering* merupakan pembelajaran mesin *unsupervised* untuk menyelesaikan masalah. *Unsupervised* berarti data tidak memiliki label atau tidak diklasifikasikan. *Clustering* dapat mengidentifikasi kelompok dalam suatu kumpulan data sehingga titik data dalam kelompok yang sama akan sangat mirip karakteristiknya dan titik data dalam kelompok yang berbeda akan memiliki karakteristik yang sangat berbeda dengan setiap kelompok diberi label yang sesuai. Terdapat berbagai metode untuk algoritma *clustering* yaitu *clustering* yang bersifat iteratif dengan mempartisi kumpulan data menjadi sub kelompok berbeda disebut dengan *partition based clustering* dan metode *hierarchical clustering* yaitu metode untuk menganalisis kelompok dengan membentuk

tingkatan seperti pohon yang disebut dendrogram. Selain itu terdapat metode *density based clustering* di mana data dikelompokkan berdasarkan area dengan titik data yang memiliki kepadatan yang sama dianggap sebagai kelompok.

Hasil dari algoritma *clustering* dapat dievaluasi dengan berbagai teknik untuk mengetahui apakah hasil dari *clustering* tersebut baik. Kelompok yang baik memiliki karakteristik jumlah kelompok yang optimal dan dengan jarak seminimal mungkin antar anggota di dalam kelompok yang sama dan jarak semaksimal mungkin dengan anggota di kelompok lain. Untuk mengetahui apakah sudah memenuhi karakteristik tersebut dapat dilakukan dengan berbagai teknik evaluasi antara lain *silhouette index*, *elbow method*, *davies bouldin index*, *mutual information*, *rand index*, *dunn index*, *calinski harabasz index*, dan lainnya.

Algoritma *clustering* yang digunakan pada tugas akhir ini adalah teknik *clustering* dengan metode berdasarkan kepadatan yaitu algoritma DBSCAN *clustering*. Algoritma DBSCAN *clustering* melakukan pengelompokan terhadap titik data dengan tetangganya yang berada dalam jarak ϵ tertentu dan memenuhi jumlah titik minimum (*min_sample*). ϵ merupakan ukuran radius untuk *neighborhood* dan *min_sample* merupakan jumlah titik tetangga minimum dalam suatu *neighborhood*. Selain itu juga digunakan algoritma *k*-Means *clustering* sebagai pembandingan. Hasil dari algoritma *clustering* akan dilakukan evaluasi menggunakan beberapa teknik evaluasi yang ada. Dari hasil kelompok yang terbentuk dilakukan analisis mengenai karakteristik yang dimiliki dan diberikan rekomendasi strategi bisnis per kelompok.

Pada tugas akhir ini, dilakukan penelitian mengenai segmentasi pelanggan dengan menggunakan data pelanggan dari perusahaan PT. Jamkrindo. Perusahaan tersebut merupakan perusahaan penjaminan kredit untuk UMKM yang menawarkan dalam bentuk skim konvensional dan syariah antara lain kredit umum, mikro, konstruksi, penjaminan bank garansi, penjaminan distribusi, penjaminan program dari surety bond, dan penjaminan dalam rangka sinergi BUMN. Selain itu juga dalam pembuatan model segmentasi pelanggan untuk data pelanggan PT. Jamkrindo akan mengkombinasikan model *clustering* dengan penerapan model RFM atau *recency, frequency, dan monetary*. Hasil segmentasi pelanggan ditampilkan pada sebuah perangkat lunak berbasis *website* yang dapat menampilkan informasi kelompok pelanggan, karakteristik setiap kelompok, dan rekomendasi strategi bisnis.

1.2 Rumusan Masalah

Rumusan masalah dari tugas akhir ini adalah sebagai berikut :

1. Bagaimana mengumpulkan data, proses awal, dan mengolah data untuk segmentasi pelanggan?
2. Bagaimana cara kerja algoritma DBSCAN dan algoritma *k*-Means untuk melakukan segmentasi pelanggan?
3. Bagaimana melakukan evaluasi hasil segmentasi pelanggan dengan *silhouette index* dan *davies bouldin index*?
4. Bagaimana menampilkan hasil segmentasi pelanggan pada perangkat lunak?

1.3 Tujuan

Tujuan dari tugas akhir ini adalah sebagai berikut :

1. Mengumpulkan data pelanggan dengan menghubungi narasumber dari PT. Jamkrindo cabang Jambi dan melakukan beberapa tahapan penyiapan data sampai dengan data siap untuk digunakan sebagai fitur *clustering*.
2. Mempelajari cara kerja algoritma DBSCAN dan algoritma *k*-Means sebagai pembandingan untuk segmentasi pelanggan dan diterapkan pada data yang digunakan.
3. Melakukan evaluasi hasil segmentasi pelanggan menggunakan teknik evaluasi *silhouette index* dan *davies bouldin index*.
4. Menampilkan hasil segmentasi pelanggan dan karakteristiknya.

1.4 Batasan Masalah

Batasan masalah untuk tugas akhir ini adalah sebagai berikut :

1. Data kasus nyata yang digunakan pada tugas akhir ini adalah data pelanggan dari PT. Jamkrindo cabang Jambi dalam rentang waktu Januari 2020 sampai dengan Desember 2021.
2. Hasil rekomendasi strategi yang dihasilkan dari segmentasi pelanggan hanya berlaku untuk PT. Jamkrindo di wilayah Kota Jambi.
3. Algoritma *clustering* yang digunakan sebagai pembanding adalah algoritma *k-Means clustering*.

1.5 Metodologi

Langkah-langkah pengerjaan tugas akhir adalah sebagai berikut :

1. Melakukan studi literatur mengenai segmentasi pelanggan dan perusahaan yang akan digunakan datanya.
2. Melakukan studi literatur mengenai algoritma *clustering* yang digunakan, teknik evaluasi hasil segmentasi, serta *tools* dan *library* yang digunakan.
3. Melakukan analisis penyelesaian masalah untuk algoritma *clustering* dengan studi kasus kecil.
4. Melakukan eksperimen dengan *tools* dan *library* yang digunakan.
5. Pengumpulan data pelanggan yang merupakan data kasus nyata.
6. Membersihkan, mengolah, dan menganalisis dataset sampai dengan siap untuk digunakan lebih lanjut.
7. Melakukan pembuatan model *clustering* untuk segmentasi pelanggan.
8. Mengevaluasi performa algoritma *clustering* yang digunakan dengan teknik evaluasi dalam *machine learning*.
9. Melakukan eksperimen dan analisis hasil eksperimen.
10. Membangun perangkat lunak berbasis *website* yang digunakan untuk menampilkan hasil *clustering* data pelanggan PT. Jamkrindo.
11. Melakukan eksperimen terhadap perangkat lunak dan pengujian fungsionalitas.
12. Menarik kesimpulan dari hasil segmentasi pelanggan dan memberikan rekomendasi strategi bisnis yang cocok berdasarkan hasil segmentasi.
13. Membuat dokumentasi tugas akhir yang menjelaskan setiap tahapan dan hasil yang didapat.

1.6 Sistematika Pembahasan

Sistematika pembahasan untuk tugas akhir adalah sebagai berikut :

1. Bab 1. Pendahuluan
Bab 1 berisi latar belakang pembuatan tugas akhir, rumusan masalah yang muncul dari adanya latar belakang, tujuan yang ingin dicapai, batasan-batasan masalah pada tugas akhir, metodologi penelitian yang digunakan, serta sistematika pembahasan untuk tugas akhir.
2. Bab 2. Landasan Teori
Bab 2 berisi teori-teori yang digunakan dalam pengerjaan tugas akhir yaitu konsep dari segmentasi pelanggan, konsep dasar *clustering*, teknik *clustering* DBSCAN, teknik evaluasi untuk algoritma *clustering*, serta teori lainnya yang digunakan pada pengerjaan tugas akhir.
3. Bab 3. Analisis Penyelesaian Masalah
Bab 3 berisi analisis mengenai penyelesaian masalah yang dilakukan mengenai pembuatan model *clustering* yang menghasilkan segmentasi pelanggan menggunakan studi kasus kecil serta eksplorasi.
4. Bab 4. Penambangan Data
Bab 4 berisi mengenai seluruh tahapan yang dilakukan pada data kasus nyata yaitu pengumpulan data, penyiapan data, eksplorasi data, ekstraksi fitur, pembangunan model dan evaluasi, analisis hasil model, dan rekomendasi strategi bisnis.

5. Bab 5. Implementasi Model dan Pengujian

Bab 5 berisi pembuatan perangkat lunak yang digunakan untuk menampilkan hasil dari segmentasi pelanggan serta pengujian fungsionalitas perangkat lunak.

6. Bab 6. Kesimpulan dan Saran

Bab 6 berisi kesimpulan dari keseluruhan pengerjaan tugas akhir dan saran berdasarkan pekerjaan yang telah dilakukan.