

**ANALISIS PROFILE PELANGGAN MENGGUNAKAN
METODE K-MEANS CLUSTERING BERDASARKAN DATA
ELECTRONIC DATA CAPTURER (EDC)**

Skripsi



oleh
JORDI RIADI
71140038

**PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2018**

**ANALISIS PROFILE PELANGGAN MENGGUNAKAN
METODE K-MEANS CLUSTERING BERDASARKAN DATA
ELECTRONIC DATA CAPTURER (EDC)**

Skripsi



Diajukan kepada Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh

JORDI RIADI
71140038

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2018

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

ANALISIS PROFILE PELANGGAN MENGGUNAKAN METODE K-MEANS CLUSTERING BERDASARKAN DATA ELECTRONIC DATA CAPTURER (EDC)

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi keserjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar keserjanaan saya.

Yogyakarta, 21 Juni 2018



JORDI RIADI
71140038

HALAMAN PERSETUJUAN

Judul Skripsi : ANALISIS POLA TRANSAKSI PELANGGAN
MENGUNAKAN METODE K-MEANS
CLUSTERING

Nama Mahasiswa : JORDI RIADI

N I M : 71140038

Matakuliah : Skripsi (Tugas Akhir)

Kode : TIW276

Semester : Genap

Tahun Akademik : 2017/2018


Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 9 Mei 2018

Dosen Pembimbing I



Budi Susanto, SKom., M.T.

Dosen Pembimbing II



R. Gunawan Santosa, Drs. M.Si.

HALAMAN PENGESAHAN

ANALISIS PROFILE PELANGGAN MENGGUNAKAN METODE K-MEANS CLUSTERING BERDASARKAN DATA ELECTRONIC DATA CAPTURER (EDC)

Oleh: JORDI RIADI / 71140038

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Data Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 30 Mei 2018

Yogyakarta, 21 Juni 2018
Mengesahkan,

Dewan Penguji:

1. Budi Susanto, S.Kom., M.T.
2. R. Gunawan Santosa, Drs. M.Si.
3. Gloria Virginia, S.Kom., MAI, Ph.D.
4. Antonius Rachmat C., S.Kom., M.Cs.



Dekan

(Budi Susanto, S.Kom., M.T.)

Ketua Program Studi

(Gloria Virginia, Ph.D.)

UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan kepada Tuhan yang Maha Esa atas berkat-Nya. Karena-Nya skripsi dengan judul “Analisis Profile Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Data EDC” dapat penulis selesaikan.

Penulis menyusun skripsi ini dalam rangka memenuhi salah satu persyaratan untuk mencapai gelar sarjana pada Program Studi Informatika, Fakultas Teknologi Informasi, Universitas Kristen Duta Wacana, Yogyakarta.

Penulis menyadari bahwa terselesaikannya skripsi ini tak lepas dari campur tangan berbagai pihak. Oleh karena itu penulis ingin menyampaikan terima kasih kepada:

1. Tuhan Yesus Kristus, atas pertolongan-Nya sehingga penulis dapat menyelesaikan skripsi
2. Bapak Budi Susanto, S.Kom., M.T., selaku dosen pembimbing I yang telah banyak memberikan bimbingan, saran, dan kritik yang membantu penulis selama pembuatan sistem dan penulisan skripsi ini.
3. Bapak Drs. R. Gunawan, M.Si., selaku dosen pembimbing II yang juga telah memberikan masukan dan arahan dalam melakukan evaluasi pada penelitian yang dibuat oleh penulis.
4. Bapak Willy Sudiarto Raharjo, S.Kom., M.cs dan Bapak Kristian Adi Nugraha, S.Kom., M.T., yang sudah memberikan kritik dan saran selama pembuatan skripsi.
5. Kepada Ibu saya yang selalu memberikan dukungan, doa, nasehat, dan motivasi kepada penulis dalam menyelesaikan studi.
6. Kepada salah satu bank swasta di Indonesia yang telah bersedia memperbolehkan menggunakan dan mempelajari karakteristik data transaksinya dan sistem basis datanya sebagai sumber penelitian.
7. Kepada Oey Jaya Angel yang selalu memberikan semangat, doa, dan dukungan untuk menyelesaikan skripsi.

8. Kepada Gunawan Tibi Santoso, Ericson Rumuy, dan Azhalia Amesa dalam membantu dan memberi saran dalam menyelesaikan skripsi.
9. Teman-teman yang berjuang bersama selama perkuliahan, Gunawan Tibi Santoso, Andre Erlando, Erwin Winata, Fidelia Vera, Olivia Citra, Budi Surya Halim, Dimas Pangestu, Anton Susilo, dan Ricky Fajar Adiputra.
10. Kepada teman-teman jurusan Informatika 2014, yang bersama-sama berjuang dalam menyelesaikan skripsi.
11. Terakhir, penulis hendak berterimakasih pada setiap nama yang tidak dapat penulis cantumkan satu per satu.

Tentunya penulis masih memiliki banyak kekurangan dalam melakukan penulisan skripsi ini. Oleh karena itu, penulis sangat menghargai jika ada berbagai masukan dari para pembaca baik berupa kritik maupun saran yang dapat membangun penulis dalam penyempurnaan penulisan skripsi di masa yang akan datang. Penulis juga meminta maaf bila terdapat kesalahan dalam penulisan skripsi ini.

Terima Kasih.

KATA PENGANTAR

Puji syukur Penulis Panjatkan ke Hadirat Tuhan Yesus Kristus karena atas kasih dan anugerah-Nya, sehingga penulis dapat menyelesaikan tugas akhir ini.

Dengan selesainya tugas akhir ini tidak lepas dari bantuan banyak pihak yang telah memberikan masukan-masukan kepada penulis. Untuk itu penulis mengucapkan banyak terimakasih.

Penulis menyadari bahwa laporan tugas akhir ini masih jauh dari kesempurnaan baik dari bentuk penyusunan maupun materinya. Oleh karena itu segala kritikan dan saran yang membangun akan penulis terima dengan baik. Akhir kata semoga laporan tugas akhir ini dapat memberikan manfaat kepada kita sekalian.

Yogyakarta, 9 Mei 2018

Penulis

INTISARI

ANALISIS PROFILE PELANGGAN MENGGUNAKAN METODE K-MEANS CLUSTERING BERDASARKAN DATA EDC

Penggunaan mesin EDC (Electronic Data Capturer) sebagai sarana transaksi sudah sangat sering digunakan oleh pelanggan. Data yang ada pasti sudah sangat banyak, tetapi terkadang data tersebut dibiarkan, tidak diolah untuk apapun, padahal data yang didapatkan dari setiap transaksi dapat diolah untuk dijadikan menjadi data yang berguna, pola transaksi misalnya. Dengan data yang sangat banyak untuk mendapatkan pola transaksi pelanggan secara manual kemungkinan memakan waktu yang lama dan tidak akurat. Klasterisasi salah satu cara untuk mendapatkan pola yang belum diketahui tersebut. Dengan cara melakukan klasterisasi, sistem dapat mengelompokkan pelanggan berdasarkan dengan kelompok yang memiliki kemiripannya masing-masing, dan setiap kelompok yang terbentuk pasti memiliki suatu kemiripan yang berujung dengan nominal transaksi yang mirip satu dengan yang lainnya. Penelitian ini akan membuat sebuah sistem yang dapat memenuhi kebutuhan diatas dengan mengimplementasikan algoritma *K-Means*. Dengan melakukan implementasi algoritma K-Means yang merupakan salah satu algoritma clustering berjenis unsupervised learning, sangat cocok sebagai algoritma klasterisasi yang belum diketahui kelompok seperti apakah yang akan terbentuk nantinya. Hasil pengelompokan klaster dari penelitian yang menggunakan algoritma *K-Means* ini sudah sesuai dengan karakteristik data yang tingkat dari jumlah nominal transaksi dipengaruhi oleh 2 atribut/fitur dengan nilai korelasi tertinggi dengan atribut nominal, yaitu umur dan jumlah penghasilan nasabah. Kesimpulan yang didapatkan bahwa semakin tinggi umur dan tingkat penghasilan, maka semakin tinggi kemungkinan bahwa jumlah nominal transaksi akan semakin tinggi juga.

Kata Kunci: *Unsupervised Learning*, Klasterisasi, *K-Means*, *Scater Plot*, *Euclidean Distance*

DAFTAR ISI

HALAMAN JUDUL.....	i
PERNYATAAN KEASLIAN TUGAS AKHIR.....	ii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN.....	v
UCAPAN TERIMA KASIH.....	vi
KATA PENGANTAR	viii
INTISARI.....	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL.....	xiii
BAB I.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah	2
1.3. Baatasan Masalah	2
1.4. Tujuan Penelitian.....	3
1.5. Manfaat Penelitian.....	3
1.6. Metode Penelitian.....	4
1.7. Sistematika Penulisan.....	5
BAB II.....	7
2.1. Tinjauan Pustaka	7
2.2. Landasan Teori	8
2.2.1. Dataset.....	8
2.2.2. Clustering	12
2.2.3. K-Means	12
2.2.4. Min Max.....	16
2.2.5. Purity	17
BAB III	20
3.1. Kebutuhan Sistem.....	20
3.1.1. Kebutuhan Fungsional	20
3.1.2. Kebutuhan Non-Fungsional	20
3.2. Use Case	22
3.2.1. Diagram Use Case.....	22

3.2.2.	Model Use Case	23
3.3.	Rancangan Environment	25
3.4.	Arsitektur Sistem	26
3.5.	Rancangan Proses	27
3.5.1.	Proses Normalisasi	28
3.5.2.	Proses Klasterisasi	28
3.6.	Rancangan Database	29
3.7.	Rancangan Antarmuka	35
3.7.1.	Halaman Login	35
3.7.2.	Halaman Tabel Data dan Normalisasi	36
3.7.3.	Halaman Klasterisasi	37
3.8.	Rancangan Pengujian	37
3.8.1.	Validasi Data	37
BAB IV	39
4.1.	Implementasi Sistem	39
4.1.1.	Implementasi Proses <i>Extract Transform Load</i> (ETL)	39
4.1.2.	Implementasi Min Max Normalization	41
4.1.3.	Implementasi K-Means Clustering	41
4.1.4.	Implementasi Antarmuka	42
4.2.	Analisis Sistem	46
4.2.1.	Analisis <i>Purity</i> Klaster	46
BAB V	57
5.1.	Kesimpulan	57
5.2.	Saran	58
DAFTAR PUSTAKA	59
LAMPIRAN	61

DAFTAR GAMBAR

<i>Gambar 2.1 Scatter Plot Graph</i>	18
<i>Gambar 3.1 Penyebaran Data</i>	22
<i>Gambar 3.2 Use Case Diagram</i>	23
<i>Gambar 3.3 Arsitektur Sistem</i>	26
<i>Gambar 3.4 Alur Rancangan Proses Sistem</i>	27
<i>Gambar 3.5 Flowchart Normalisasi Min Max</i>	28
<i>Gambar 3.6 Flowchart Algoritma K-Means</i>	29
<i>Gambar 3.7 Rancangan skema pemodelan basis data</i>	34
<i>Gambar 3.8 Rancangan skema data warehouse</i>	35
<i>Gambar 3.9 Halaman login</i>	35
<i>Gambar 3.10 Halaman Tabel Data</i>	36
<i>Gambar 3.11 Halaman Tabel Normalisasi</i>	36
<i>Gambar 3.12 Halaman Klasterisasi</i>	37
<i>Gambar 4.1 Query tabel dimensi waktu</i>	39
<i>Gambar 4.2 Query tabel dimensi merchant</i>	40
<i>Gambar 4.3 Query tabel dimensi waktu</i>	40
<i>Gambar 4.4 Query tabel fakta</i>	40
<i>Gambar 4.5 Halaman login sistem</i>	42
<i>Gambar 4.6 Antarmuka input range tanggal</i>	43
<i>Gambar 4.7 Antarmuka tampilan tabel fakta</i>	44
<i>Gambar 4.8 Antarmuka tampilan normalisasi</i>	44
<i>Gambar 4.9 Antarmuka tampilan hasil klasterisasi</i>	45
<i>Gambar 4.10 Antarmuka tampilan hasil klasterisasi berupa scatter plot</i>	45
<i>Gambar 4.11 Antarmuka tampilan hasil klasterisasi</i>	47
<i>Gambar 4.12 Antarmuka tampilan hasil klasterisasi</i>	50

DAFTAR TABEL

Tabel 2.1 Tingkatan <i>income</i>	10
Tabel 2.2 Tabel Data	13
Tabel 2.3 Tabel Centroid Iterasi 1.....	13
Tabel 2.4 Tabel Data Setelah Normalisasi	14
Tabel 2.5 Tabel Centroid Setelah Normalisasi	14
Tabel 2.6 Iterasi 1.....	14
Tabel 2.7 Centroid Iterasi 2.....	15
Tabel 2.8 Iterasi 2.....	15
Tabel 2.9 Centroid Iterasi 3.....	16
Tabel 2.10 Iterasi 3.....	16
Tabel 2.11 Normalisasi data tabel iterasi 3 dengan metode Min Max.....	17
Tabel 2.12 Normalisasi centroid iterasi 3	17
Tabel 3.1 Nilai korelasi	21
Tabel 3.2 Daftar kategori	21
Tabel 4.1 Analisis hasil kluster pada kategori 1 atribut pada Gambar 4.12.....	48
Tabel 4.2 Analisis <i>Purity</i> kategori 1 atribut pada kluster Gambar 4.12.....	48
Tabel 4.3 Analisis hasil kluster pada Gambar 4.12.....	49
Tabel 4.4 Analisis <i>Purity</i> pada kluster Gambar 4.12	49
Tabel 4.5 Analisis hasil kluster pada kategori 1 atribut pada Gambar 4.13.....	51
Tabel 4.6 Analisis <i>Purity</i> kategori 1 atribut pada kluster Gambar 4.13.....	51
Tabel 4.7 Analisis hasil kluster pada Gambar 4.13.....	52
Tabel 4.8 Analisis <i>Purity</i> pada kluster Gambar 4.13	52
Tabel 4.9 Rangkuman Analisis	53
Tabel 4.10 Analisis nilai <i>purity</i> terhadap jumlah k pada kluster	53
Tabel 4.11 Analisis perilaku nasabah pada kluster <i>rule</i> KEMENKEU	54
Tabel 4.12 Analisis perilaku nasabah pada kluster <i>rule</i> KEMENKEU	54
Tabel 4.13 Analisis perilaku nasabah pada kluster <i>rule binning</i>	55
Tabel 4.14 Analisis perilaku nasabah pada kluster <i>rule binning</i>	55

INTISARI

ANALISIS PROFILE PELANGGAN MENGGUNAKAN METODE K-MEANS CLUSTERING BERDASARKAN DATA EDC

Penggunaan mesin EDC (Electronic Data Capturer) sebagai sarana transaksi sudah sangat sering digunakan oleh pelanggan. Data yang ada pasti sudah sangat banyak, tetapi terkadang data tersebut dibiarkan, tidak diolah untuk apapun, padahal data yang didapatkan dari setiap transaksi dapat diolah untuk dijadikan menjadi data yang berguna, pola transaksi misalnya. Dengan data yang sangat banyak untuk mendapatkan pola transaksi pelanggan secara manual kemungkinan memakan waktu yang lama dan tidak akurat. Klasterisasi salah satu cara untuk mendapatkan pola yang belum diketahui tersebut. Dengan cara melakukan klasterisasi, sistem dapat mengelompokkan pelanggan berdasarkan dengan kelompok yang memiliki kemiripannya masing-masing, dan setiap kelompok yang terbentuk pasti memiliki suatu kemiripan yang berujung dengan nominal transaksi yang mirip satu dengan yang lainnya. Penelitian ini akan membuat sebuah sistem yang dapat memenuhi kebutuhan diatas dengan mengimplementasikan algoritma *K-Means*. Dengan melakukan implementasi algoritma K-Means yang merupakan salah satu algoritma clustering berjenis unsupervised learning, sangat cocok sebagai algoritma klasterisasi yang belum diketahui kelompok seperti apakah yang akan terbentuk nantinya. Hasil pengelompokan klaster dari penelitian yang menggunakan algoritma *K-Means* ini sudah sesuai dengan karakteristik data yang tingkat dari jumlah nominal transaksi dipengaruhi oleh 2 atribut/fitur dengan nilai korelasi tertinggi dengan atribut nominal, yaitu umur dan jumlah penghasilan nasabah. Kesimpulan yang didapatkan bahwa semakin tinggi umur dan tingkat penghasilan, maka semakin tinggi kemungkinan bahwa jumlah nominal transaksi akan semakin tinggi juga.

Kata Kunci: *Unsupervised Learning*, Klasterisasi, *K-Means*, *Scater Plot*, *Euclidean Distance*

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Perkembangan teknologi di era modern membuat hampir segala hal sudah berdasarkan sistem komputerisasi, salah satunya adalah transaksi menggunakan kartu debit atau kredit dengan mesin *EDC* (*Electronic Data Capturer*). Tingkat konsumsi masyarakat pun bertambah karena semakin lama jumlah kebutuhan yang dibutuhkan dan diinginkan semakin banyak, sehingga jumlah *merchant* atau toko juga semakin banyak karena tuntutan kebutuhan atas konsumsi tersebut. Dengan jumlah transaksi yang banyak tersebut, tidak sedikit yang memilih untuk melakukan transaksi menggunakan mesin *EDC* selain karena praktis, terkadang tergiur juga dengan promo-promo yang ditawarkan dan mengharuskan calon pembeli menggunakan kartu debit atau kredit untuk mendapatkan promo tersebut. Dengan jumlah transaksi yang banyak menggunakan mesin *EDC* tersebut terdapat banyak data yang dapat diolah.

Berdasarkan pada hal tersebut, penulis akan melakukan penelitian terhadap salah satu bank swasta terbesar di Indonesia. Bank swasta ini sudah memiliki banyak mesin *EDC* yang tersebar hampir di seluruh Indonesia. Dengan itu dapat disimpulkan bahwa jumlah data yang ada dapat dikatakan sangat banyak. Data mentah yang ada tersebut dapat diolah menjadi sebuah nilai yang nantinya akan menghasilkan pola dari transaksi yang menggunakan mesin *EDC*, dari data tersebut pihak bank swasta akan mendapatkan informasi bagaimana kinerja mesin *EDC* yang digunakan oleh setiap *merchant*, apakah mesin *EDC* tersebut digunakan secara rutin atau tidak. Pola transaksi yang akan dianalisis berupa fitur-fitur seperti umur, jenis kelamin, pekerjaan, pendidikan, status pernikahan, gaji pelanggan, dan nominal dari transaksi yang dilakukan nasabah. Dari pola transaksi tersebut pihak bank akan dipermudah dalam mendapatkan informasi berupa kelompok range umur nasabah yang menggunakan kartu milik bank tersebut berdasarkan tingkatan

nominal transaksi yang terdekat. Pihak bank swasta tersebut juga dapat mengetahui dimana wilayah yang tingkat transaksi menggunakan mesin *EDC*-nya masih dibawah rata-rata sebagai target pemasaran misalnya diberikan promo khusus di wilayah tersebut agar jumlah transaksi yang menggunakan kartu debit atau kredit bank tersebut meningkat dan dapat memperluas wilayah cakupan mesin *EDC* bank tersebut untuk kedepannya. Sebagai sarana yang berguna untuk memudahkan pihak bank dalam mendapatkan pola transaksi tersebut, penelitian ini akan menggunakan metode *clustering K-Means* dalam mengelompokkan data yang ada berdasarkan kemiripan dari sekumpulan data yang sudah ada. *Clustering* berguna untuk mempermudah pihak bank dalam membaca data yang telah diolah untuk mengetahui pola dari data transaksi yang ada. Dari hasil *clustering* data yang ada, akan terlihat data yang mana saja yang memiliki kemiripan sifat dengan data lainnya.

Penelitian ini akan menerapkan metode *K-Means clustering* untuk melakukan analisa pola transaksi pelanggan dengan mengelompokkan data berdasarkan kemiripan sifat. Sistem akan menampilkan hasil dari klasterisasi dalam bentuk tabel dan grafik pada sebuah *website*. Hasil dari *clustering* akan dievaluasi menggunakan metode *purity* untuk mengetahui tingkat optimalitas hasil *clustering* yang dilakukan oleh sistem.

1.2. Rumusan Masalah

- Bagaimana implementasi *K-Means clustering* pada pola transaksi pelanggan yang menggunakan mesin *EDC* pada salah satu bank.
- Bagaimana kualitas hasil *clustering* berdasarkan nilai *purity*.

1.3. Batasan Masalah

- Pengguna atau *user* sistem ini adalah *Customer Service Officer (CSO)* dari pihak bank.
- Data yang akan dianalisa merupakan data *dummy* yang di-*generate* sendiri oleh penulis yang memiliki pola berdasarkan dari data transaksi dari pelanggan *merchant* yang menggunakan mesin *EDC* milik salah satu bank swasta di Indonesia.

- Hasil klasterisasi data akan divisualisasikan ke dalam grafik *Scatter Plot*.
- Hasil akhir dari serangkaian analisa tersebut adalah terbentuknya *cluster* pola transaksi pelanggan sejenis.
- Hasil *clustering* akan dievaluasi menggunakan metode *purity*.

1.4. Tujuan Penelitian

Untuk membuat sistem yang dapat menghasilkan *cluster* hasil dari analisa sekumpulan data transaksi dengan menerapkan metode K-Means terhadap data transaksi *dummy* yang telah dikonsultasikan dengan salah satu karyawan dari bank swasta yang dan transaksi yang menggunakan mesin *EDC* milik salah satu bank swasta di Indonesia yang bertujuan untuk mendapatkan pola perilaku pelanggan dalam transaksi yang menggunakan mesin *EDC* milik bank tersebut. Pola transaksi yang didapatkan berupa berapakah range umur pelanggan yang melakukan transaksi dengan tingkatan nominal tertentu, apakah ada fitur lain yang dapat mempengaruhi pola nasabah dalam melakukan transaksi. Dari pola transaksi yang didapatkan, pihak bank dapat mengetahui fitur-fitur apa saja yang dapat mempengaruhi nasabah dalam melakukan transaksi, dan dengan diketahuinya fitur yang mempengaruhi pola transaksi nasabah, target pemasaran dapat tertuju pada nasabah-nasabah yang memiliki potensi.

1.5. Manfaat Penelitian

Manfaat bagi bank swasta dalam menggunakan analisa pola transaksi ini adalah untuk mengetahui pola dari kebiasaan nasabah/masyarakat dalam melakukan transaksi menggunakan kartu debit yang dapat diolah menjadi informasi yang berguna untuk memudahkan pemilihan target pemasaran, meningkatkan penyebaran mesin *EDC* ke wilayah yang memiliki tingkat penggunaan mesin *EDC* diatas rata-rata, sehingga mesin *EDC* milik bank swasta tersebut digunakan secara optimal, demikian sebaliknya pihak bank swasta tersebut juga dapat mengetahui dimana mesin *EDC* yang jarang digunakan. Hal ini merupakan salahsatu masalah yang dimiliki oleh pihak bank swasta tersebut, karena mesin *EDC* yang tersebar banyak yang jumlah transaksinya sangat rendah dari rata-rata transaksi yang ada, dan lebih baik mesin tersebut ditarik kembali dan diberikan

ke *merchant* yang baru dan memiliki potensi untuk mendapatkan transaksi yang lebih banyak. Diharapkan sistem ini juga membantu pihak bank swasta dalam menggunakan hasil dari analisa pola ini untuk mendapatkan pola berupa kelompok umur tertentu, maupun nasabah dengan fitur lainnya yang memiliki korelasi dalam melakukan transaksi dengan nominal transaksi yang semirip mungkin yang melakukan transaksi menggunakan kartu debit dari setiap tempat perbelanjaan ataupun *merchant* yang memiliki mesin *EDC* milik bank swasta tersebut. Jika pihak bank swasta tersebut ingin mengadakan promo khusus pengguna kartu debit atau kredit milik bank swasta tersebut, pihak bank sudah mengetahui pelanggan dengan umur berapa yang dapat menjadi sasaran promo tersebut yang sesuai dengan produk yang akan dijadikan promo tersebut.

1.6. Metode Penelitian

Metode-metode yang digunakan penulis dalam menyelesaikan penelitian ini adalah sebagai berikut:

1. Metode Pengumpulan Data

a. Studi Pustaka

Penulis melakukan studi pustaka dengan cara mengumpulkan dan mempelajari teori-teori dan literature yang berhubungan dan membantu dalam penyelesaian penelitian. Teori-teori yang digunakan adalah teori-teori yang berkaitan dengan *Clustering*, algoritma *K-Means*, purity, serta perangkat lunak yang akan digunakan dalam membangun sistem.

b. Pengambilan Data

Penulis akan membuat data *dummy* dengan karakteristik data yang semirip mungkin dan sudah dikonsultasikan dengan salahsatu karyawan bank swasta tersebut, dan sistem basis data yang hampir menyerupai sistem basis data dari salah satu bank swasta di Indonesia sebagai bahan data untuk melakukan penelitian.

2. Metode Perancangan Sistem

Tahap ini penulis akan melakukan perancangan sistem yang akan dibuat, mulai dari pembuatan data mentah yang sesuai dengan data asli yang sudah

dipelajari oleh penulis, proses normalisasi data sebelum ditampilkan kedalam grafik, perancangan desain antarmuka pada sistem, serta melakukan evaluasi pada pola yang telah dibuat oleh sistem dengan membandingkan dengan nilai dari *purity*.

3. Metode Pengembangan Sistem

Metode pengembangan sistem yang akan digunakan oleh penulis adalah algoritma *K-Means Clustering* pada data dummy transaksi mesin EDC salah satu bank swasta di Indonesia.

4. Metode Evaluasi

Tahap evaluasi hasil dari *clustering* akan dilakukan menggunakan metode *Purity* untuk menganalisa apakah hasil *clustering* yang dijalankan sudah optimal atau belum dengan melihat berapa persen dari total jumlah data yang sudah di *cluster* dengan benar berdasarkan nilai *purity*.

1.7. Sistematika Penulisan

Sistematika penulisan penelitian ini akan dalam sebuah laporan yang terdiri dari lima bab, sebagai berikut:

Bab I. Pendahuluan

Bab ini berisikan latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, metodologi penelitian, dan sistematika penulisan laporan skripsi.

Bab II. Tinjauan Pustaka

Bab ini berisikan tinjauan pustaka dan landasan teori. Bab ini menguraikan teori maupun hasil dari penelitian serupa yang telah dilakukan sebelumnya untuk dipergunakan sebagai referensi dalam perancangan sistem. Landasan teori berisikan konsep yang berguna sebagai pemecahan masalah dalam penelitian ini yang diperoleh dari berbagai sumber.

Bab III. Perancangan Sistem Penelitian

Bab ini berisikan perancangan sistem yang meliputi pengolahan data ke *database*, perancangan implementasi algoritma *K-Means Clustering*

sebagai algoritma yang digunakan dalam kalsterisasi di penelitian ini, serta perancangan sistem antar muka dari sistem yang akan dibuat.

Bab IV. Implementasi dan Analisis Sistem

Bab ini berisikan hasil implementasi dan analisis dari penelitian mengenai setiap proses yang ada berdasarkan perancangan sistem yang telah dibuat di bab sebelumnya.

Bab V. Kesimpulan dan Saran

Bab ini berisikan kesimpulan dari penelitian setelah menganalisis hasil implementasi sistem yang telah dibuat. Bab ini juga berisikan saran yang dapat digunakan dalam melakukan penelitian yang serupa atau mengembangkan sistem ini di masa mendatang.

©UKDW

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Dalam penelitian ini, penggunaan *K-Means Clustering* untuk melakukan analisa pola transaksi nasabah, penulis menggunakan *rule* dari KEMENKEU dan penggunaan metode *binning* sebagai aturan dalam menentukan kategori pada tingkatan *income* dari nasabah. Transaksi berhasil diterapkan dan dikembangkan menjadi sebuah program. Berdasarkan hasil pengujian dan analisis yang telah dilakukan oleh penulis, dapat disimpulkan:

1. Berdasarkan hasil pengujian dengan jumlah $k=6$ ditentukan bahwa penggunaan *rule* untuk kategori *income* dari KEMENKEU dengan kategori 1 atribut menghasilkan hasil *purity* paling baik, dan pada urutan berikutnya KEMENKEU dengan 2 atribut, *binning* dengan 1 atribut, serta *binning* dengan 2 atribut dengan nilai *purity* paling rendah
2. Berdasarkan hasil analisis *purity* untuk setiap jumlah k , penggunaan *rule* untuk kategori *income* yang paling optimal adalah penggunaan metode *rule* kategori *income* dari KEMENKEU dengan nilai rata-rata *purity* tertinggi, dan pada urutan berikutnya *binning* dengan 1 atribut, KEMENKEU dengan 2 atribut, serta *binning* dengan 2 atribut.
3. Berdasarkan hasil analisis perilaku nasabah pada setiap *rule* untuk kategori *income* baik yang menggunakan *rule* dari KEMENKEU maupun metode *binning*, ditemukan bahwa semakin tinggi rata-rata umur pada kategori maka semakin tinggi nilai rata-rata nominal transaksi yang dilakukan. Tetapi rata-rata umur yang terendah tidak memiliki nilai rata-rata nominal transaksi yang paling rendah.

5.2. **Saran**

Saran-saran yang dapat digunakan dalam pengembangan aplikasi selanjutnya antara lain:

1. Data transaksi yang diteliti merupakan data transaksi yang real, sehingga dapat ditemukan pola transaksi nasabah yang lain
2. Data transaksi yang diteliti memiliki jumlah yang lebih banyak, serta rentang waktu yang lebih lama sehingga dapat mendapatkan hasil pengujian metode, dan analisis pola yang lebih akurat.

©UKDW

DAFTAR PUSTAKA

- Agusta, Y. (2007). *Jurnal Sistem dan Informatika. K-Means – Penerapan, Permasalahan*, 47-60.
- Albert. (2017, May 8). *Age Categories, Life Cycle Groupings*. Retrieved from statcan: <https://statcan.gc.ca/eng/concepts/definitions/age2>
- Amogh, K., & Mahesh, K. (2014). *International Journal of Emerging Technology and Advanced Engineering. Survey on Hadoop and Introduction to YARN*, 82 - 87.
- Dharayani, R., Laksitowening, K. A., & P, Y. A. (2015). e-Proceeding of Engineering : Vol.2, No.1. *Implementasi ETL (Extract, Transform, Load) Pangkalan Data Perguruan Tinggi dengan Menggunakan State-Space Problem*, 1159-1165.
- Han, J., Kamber, M., & Pei, J. (2000). *Data Mining Concepts and Techniques. 3th Ed.*
- Hartanto, J., Herman, B., Indra, M., & Yuliana, M. (2011). *Prosiding Konferensi Nasional “Inovasi dalam Desain dan Teknologi”*. *Data Transformation Pada Data Mining*, 93-99.
- Husni, N. Y. (2015). *Jurnal Simantec. Clusterisasi Dokumen Web (Berita) Bahasa Indonesia Menggunakan algoritma K-Means*, 159-166.
- Manning, C. D., Raghavan, P., & Schütze, H. (2009). *An Introduction to Information Retrieval*. England: Cambridge University Press.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big Data: The Next Frontier For Innovation, Competition, and Productivity*, 1-2.
- Marzuki, I. (2015). *Smatika Jurnal. Temu Kembali Informasi Big Data Menggunakan K-Means Clustering*, 1-7.
- Patro, G. K., & Sahu, K. K. (2015). *Normalization: A Preprocessing Stage*.
- Rahadian, B. A., Kurnianingtyas, D., Mahardika, D. P., & Maghfira, T. N. (2017). *Jurnal Teknologi Informasi dan Ilmu Komputer . Analisis Judul Majalah Kawanku Menggunakan Clustering K-Means Dengan Konsep Simulasi Big Data Pada Hadoop Multi Node Cluster*, 75 - 80.
- Russom, P. (2011). *Big Data Analytics*. Renton: TDWI Research.
- Santosa, B. (2007). *Data Mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Sripada, S. C., & Rao, M. S. (2011). *Indian Journal of Computer Science and Engineering . Comparison of Purity and Entropy of K-Means Clustering and Fuzzy C Means Clustering*, 2-3.

- Sulianta, F. (2014). Jurnal Ilmiah Teknologi Informasi Terapan . *Customer Profiling Pada Supermarket Menggunakan Algoritma K-Means Dalam Memilih Produk Berdasarkan Selera Konsumen Dengan Daya Beli Maksimum*, 41-45.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Introduction to Data Mining*. Boston: Pearson Education.
- Trisminingsih, R., & Kiswari, I. Y. (2016). Jurnal Siliwangi. *Modul Extract, Transform*, 108-103.
- Wagstaf, K., Cardie, C., Rogers, S., & Schroedl, S. (2001). *Constrained K-means Clustering with Background Knowledge*.
- Widiatmanti. (2015, April 29). *Penghasilan Kelas Menengah*. Retrieved from BPPK KEMENKEU:
<http://www.bppk.kemenkeu.go.id/publikasi/artikel/167-artikel-pajak/21014-penghasilan-kelas-menengah-naik-potensi-pajak>

©UKDW