

**PENERAPAN AGGLOMERATIVE CLUSTERING PADA NOTA  
PELANGGAN PIUTANG RITEL FASHION DAN SEPATU**

Tugas Akhir



Disusun Oleh:

MICHAEL ABADI SANTOSO

71120067

PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS KRISTEN DUTA WACANA  
YOGYAKARTA

2016

**PENERAPAN AGGLOMERATIVE CLUSTERING PADA NOTA  
PELANGGAN PIUTANG PADA RITEL FASHION DAN  
SEPATU**

Tugas Akhir



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

Disusun Oleh:  
MICHAEL ABADI SANTOSO  
71120067

PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS KRISTEN DUTA WACANA  
YOGYAKARTA  
2016

## PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

### **PENERAPAN AGGLOMERATIVE CLUSTERING PADA NOTA PELANGGAN PIUTANG RITEL FASHION DAN SEPATU**

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan 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 kesarjanaan saya.

Yogyakarta, 18 Oktober 2016



MICHAEL ABADI SANTOSO

71120067

## HALAMAN PERSETUJUAN

Judul Skripsi : PENERAPAN AGGLOMERATIVE CLUSTERING  
PADA NOTA PELANGGAN PIUTANG RITEL  
FASHION DAN SEPATU

Nama Mahasiswa : MICHAEL ABADI SANTOSO

N I M : 71120067

Matakuliah : Skripsi (Tugas Akhir)

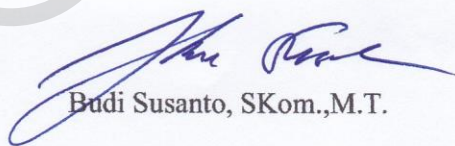
Kode : TIW276

Semester : Gasal

Tahun Akademik : 2016/2017

Telah diperiksa dan disetujui di  
Yogyakarta,  
Pada tanggal 18 Oktober 2016

Dosen Pembimbing I



Budi Susanto, SKom., M.T.

Dosen Pembimbing II



Gloria Virginia, S.Kom., MAI, Ph.D.

## HALAMAN PENGESAHAN

### PENERAPAN AGGLOMERATIVE CLUSTERING PADA NOTA PELANGGAN PIUTANG RITEL FASHION DAN SEPATU

Oleh: MICHAEL ABADI SANTOSO / 71120067

Dipertahankan di depan Dewan Penguji Skripsi  
Program Studi Teknik Informatika Fakultas Teknologi Informasi  
Universitas Kristen Duta Wacana - Yogyakarta  
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar  
Sarjana Komputer  
pada tanggal 17 Oktober 2016

Yogyakarta, 18 Oktober 2016  
Mengesahkan,

Dewan Penguji:

1. Budi Susanto, SKom., M.T.
2. Gloria Virginia, S.Kom., MAI, Ph.D.
3. Restyandito, S.Kom., MSIS, Ph.D
4. Laurentius Kuncoro Probo Saputra, S.T.,  
M.Eng.

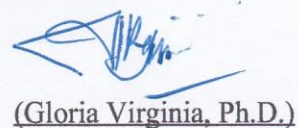


Dekan



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

Ketua Program Studi



(Gloria Virginia, Ph.D.)

## UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa karena atas berkat dan rahmat-Nya, skripsi yang berjudul “PENERAPAN *AGGLOMERATIVE CLUSTERING* PADA NOTA PELANGGAN PIUTANG RITEL FASHION DAN SEPATU “ ini dapat terselesaikan.

Penulis menyusun skripsi ini dalam rangka memenuhi salah satu persyaratan untuk mencapai gelar sarjana (S1) pada Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Duta Wacana Yogyakarta.

Penulis menyadari bahwa terselesaikannya Skripsi ini tak lepas dari campur tangan berbagai pihak. Untuk itulah penulis ingin menyampaikan terima kasih kepada:

1. Bapak Budi Susanto, S.Kom., M.T., selaku dosen pembimbing I, yang telah banyak memberikan bimbingan selama penyusunan dan penulisan Skripsi ini.
2. Ibu Gloria Virginia, Ph.D., selaku dosen pembimbing II yang juga telah banyak memberikan masukan dan arahan selama pembuatan skripsi.
3. Bapak Antonius Rachmat, S.Kom.,M.Kom., yang selalu memberikan kritik dan saran selama pembuatan skripsi.
4. Bapak Kristian Adi Nugraha, S.Kom.,M.Kom., yang juga selalu memberikan kritik dan saran selama pembuatan skripsi.
5. Kepada salah satu sumber perusahaan ritel fashion dan sepatu di Indonesia yang telah bersedia memberikan data yang digunakan untuk sumber penelitian.
6. Kepada keluarga terkasih, Bapak, Ibu, dan kakak saya yang memberikan dukungan, doa, nasehat, dan motivasi hingga sampai detik ini penulis menyelesaikan studi.

7. Kepada Laksmi Widya Astuti, S.Kom., yang selalu memberikan semangat dan cinta yang selalu abadi seperti nama saya.
8. Kepada teman-teman jurusan Teknik Informatika 2012 dan Twelvi, yang senantiasa ada untuk memberikan dukungan, dan sama-sama berjuang untuk menyelesaikan tugas akhir.
9. Kepada keluarga besar yang senantiasa memotivasi serta selalu mendoakan kelancaran studi hingga Skripsi ini terselesaikan, dan semua keluarga yang tak bisa disebutkan satu per satu.
10. Kepada teman masa kecil Yakobos Agung S.Kom, Eugenius Radityo Unggul, S.Par, dan Laksito Aji yang saling mendukung satu sama lain.
11. Kepada Big Data Skripsweet Squad yaitu Constantius Damar Wicaksono, S.Kom, Klaudius Jemly, Widnyana Santika, I Made Arya Beta dan Yoas Hernanda yang sejuangan bersama bahu membahu dalam suka dan duka menyelesaikan topik Big Data.
12. Terakhir, penulis hendak menyapa setiap nama yang tidak dapat penulis cantumkan satu per satu, terima kasih atas doa yang senantiasa mengalir tanpa sepengetahuan penulis.

Dalam penyusunan tugas akhir ini, tentunya penulis masih memiliki banyak kekurangan pada topik dalam Skripsi ini dan penulisannya yang masih banyak terdapat kekurangan.

Oleh karena itu, penulis sangat menghargai dan menerima jika ada berbagai masukan dari para pembaca baik berupa kritik maupun saran yang sifatnya membangun demi penyempurnaan penulisan-penulisan Skripsi di masa yang akan datang. Penulis meminta maaf bila ada kesalahan dalam penulisan Skripsi ini.

Terima Kasih

## **KATA PENGANTAR**

Puji syukur Penulis Panjatkan ke Hadirat Tuhan Yang Maha Esa karena atas Rahmat dan Karunia-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, Oktober 2016

## INTISARI

### **PENERAPAN AGGLOMERATIVE CLUSTERING PADA NOTA PELANGGAN PIUTANG RITEL FASHION DAN SEPATU**

*Agglomerative Clustering* merupakan salah satu metode yang membentuk tingkatan tertentu seperti struktur pada pohon dikarenakan proses dilakukan secara bertahap untuk mendapatkan *cluster*. Ada beberapa algoritma yang dapat digunakan dalam melakukan *Agglomerative Clustering*, diantaranya adalah algoritma *Single Linkage* dan *Complete Linkage*. Dalam penelitian ini, penulis menggunakan algoritma *Single Linkage* dan *Complete Linkage* beserta metode normalisasi *Min-Max* dan *Z-Score* serta metode *similarity Euclidean Distance* dan *Cosine Similarity* untuk mengelompokkan perilaku pelanggan menjadi *cluster* pada transaksi penjualan bon dan tunai pada perusahaan ritel fashion dan sepatu dari tahun 2010-2012. Evaluasi dilakukan dengan menghitung nilai *purity cluster* menggunakan periode waktu kuartal. Pengujian menggunakan metode yang sama berdasar nilai *purity* yang terbaik untuk setiap kuartalnya. Berdasarkan hasil analisis yang dilakukan, ditemukan bahwa *Complete Linkage* dengan menggunakan *Min-Max* serta *Euclidean* menghasilkan nilai *purity* paling baik serta sistem tidak dapat memprediksi perilaku pelanggan dalam pembelian barang pada setiap tahunnya baik pada penjualan bon atau tunai dikarenakan pola klaster dominan yang berbeda dan tidak dapat memprediksi golongan umur yang dominan dikarenakan golongan yang dominan dalam pembelian berdasar pada nilai minimum dan maksimum data pada setiap kuartal. Selain itu ditemukan bahwa data penelitian tidak cocok untuk dilakukan klasterisasi maupun analisis perilaku pelanggan piutang.

**Kata Kunci:** Agglomerative Clustering, Single Linkage, Complete Linkage, Data Mining, Implementasi Algoritma, Big Data, Data Warehouse, Euclidean Distance, Cosine Similarity, Z-Score, Min-Max

## DAFTAR ISI

|                                                  |      |
|--------------------------------------------------|------|
| HALAMAN JUDUL .....                              | ii   |
| PERNYATAAN KEASLIAN TUGAS AKHIR .....            | iii  |
| HALAMAN PERSETUJUAN.....                         | iv   |
| HALAMAN PENGESAHAN .....                         | v    |
| UCAPAN TERIMA KASIH.....                         | vi   |
| KATA PENGANTAR .....                             | viii |
| INTISARI .....                                   | ix   |
| DAFTAR ISI.....                                  | x    |
| DAFTAR GAMBAR .....                              | xiv  |
| DAFTAR TABEL.....                                | xvii |
| BAB I PENDAHULUAN.....                           | 1    |
| 1.1 Latar Belakang.....                          | 1    |
| 1.2 Rumusan Masalah .....                        | 3    |
| 1.3 Batasan Sistem.....                          | 3    |
| 1.4 Tujuan Penelitian.....                       | 4    |
| 1.5 Metode Penelitian.....                       | 4    |
| 1.6 Sistematika Penulisan.....                   | 5    |
| BAB II TINJAUAN PUSTAKA .....                    | 7    |
| 2.1 Tinjauan Pustaka .....                       | 7    |
| 2.2 Landasan Teori .....                         | 9    |
| 2.2.1 Analisis Clustering .....                  | 9    |
| 2.2.2 Hierarchical Agglomerative Clustering..... | 9    |
| 2.2.3 Single Linkage.....                        | 11   |

|                                                      |                                            |           |
|------------------------------------------------------|--------------------------------------------|-----------|
| 2.2.4                                                | Studi Kasus Single Linkage.....            | 11        |
| 2.2.5                                                | Complete Linkage .....                     | 18        |
| 2.2.6                                                | Studi Kasus Agglomerative Clustering ..... | 21        |
| 2.2.7                                                | Purity .....                               | 24        |
| 2.2.8                                                | Data Warehouse.....                        | 27        |
| 2.2.9                                                | Bussiness Inteligence.....                 | 28        |
| 2.2.10                                               | Big Data.....                              | 28        |
| <b>BAB III ANALISIS DAN PERANCANGAN SISTEM .....</b> |                                            | <b>29</b> |
| 3.1                                                  | Kebutuhan Sistem.....                      | 29        |
| 3.1.1                                                | Kebutuhan Fungsional.....                  | 29        |
| 3.1.2                                                | Kebutuhan Non Fungsional.....              | 30        |
| 3.2                                                  | Use Case .....                             | 30        |
| 3.2.1                                                | Diagram Use Case .....                     | 30        |
| 3.2.2                                                | Model Use Case.....                        | 31        |
| 3.3                                                  | Environment Percobaan.....                 | 35        |
| 3.3.1                                                | Spesifikasi Kebutuhan Perangkat Keras..... | 35        |
| 3.3.2                                                | Spesifikasi Kebutuhan Perangkat Lunak..... | 36        |
| 3.4                                                  | Arsitektur Sistem.....                     | 37        |
| 3.5                                                  | Rancangan Proses.....                      | 38        |
| 3.5.1                                                | Proses ETL .....                           | 39        |
| 3.5.2                                                | Proses Normalisasi dan Pivoting.....       | 40        |
| 3.5.3                                                | Proses Similarity.....                     | 43        |
| 3.5.4                                                | Proses Analisis Klasterisasi.....          | 46        |
| 3.6                                                  | Rancangan Database.....                    | 47        |
| 3.6.1                                                | Database Transaksional .....               | 48        |
| 3.6.2                                                | Database Dimensional .....                 | 53        |
| 3.7                                                  | Rancangan User Interface.....              | 55        |
| 3.7.1                                                | Halaman Login .....                        | 55        |
| 3.7.2                                                | Halaman Input Data Uji.....                | 55        |
| 3.7.3                                                | Halaman Proses Tabel Fakta .....           | 56        |
| 3.7.4                                                | Halaman Normalisasi .....                  | 56        |

|                                        |                                                           |    |
|----------------------------------------|-----------------------------------------------------------|----|
| 3.7.5                                  | Halaman Pivoting dan Scatter Plot.....                    | 57 |
| 3.7.6                                  | Halaman Similarity.....                                   | 57 |
| 3.7.7                                  | Halaman Dendrogram .....                                  | 58 |
| 3.7.8                                  | Halaman Lihat Data.....                                   | 58 |
| 3.7.9                                  | Halaman Manajemen User .....                              | 59 |
| 3.8                                    | Rancangan Pengujian dan Analisis .....                    | 60 |
| 3.8.1                                  | Validasi Data .....                                       | 61 |
| 3.8.1.1                                | Normalisasi.....                                          | 62 |
| 3.8.1.2                                | Pengujian Pivoting.....                                   | 64 |
| 3.8.1.3                                | Pengujian Similarity .....                                | 64 |
| 3.8.1.4                                | Pengujian Pembentukan Kluster.....                        | 66 |
| 3.8.2                                  | Analisis Sistem .....                                     | 67 |
| BAB IV IMPLEMENTASI DAN ANALISIS ..... |                                                           | 68 |
| 4.1                                    | Implementasi Sistem .....                                 | 69 |
| 4.1.1                                  | Konfigurasi Server.....                                   | 69 |
| 4.1.1.1                                | Konfigurasi hostname.....                                 | 70 |
| 4.1.1.2                                | Konfigurasi <i>IP Address</i> .....                       | 71 |
| 4.1.1.3                                | Konfigurasi <i>Hadoop</i> .....                           | 71 |
| 4.1.1.4                                | Konfigurasi <i>Hive</i> .....                             | 76 |
| 4.1.1.5                                | Konfigurasi <i>Zookeeper</i> .....                        | 77 |
| 4.1.1.6                                | Konfigurasi <i>Tez</i> .....                              | 78 |
| 4.1.2                                  | Proses Extract Transformation Load (ETL).....             | 80 |
| 4.1.2.1                                | ETL Tabel Dimensi .....                                   | 87 |
| 4.1.2.2                                | ETL Tabel Fakta.....                                      | 90 |
| 4.1.3                                  | Implementasi Proses <i>Cleaning</i> .....                 | 91 |
| 4.1.4                                  | Implementasi Proses Data <i>Filtering</i> .....           | 91 |
| 4.1.5                                  | Implementasi Algoritma <i>Min-Max Normalization</i> ..... | 92 |
| 4.1.6                                  | Implementasi Algoritma Altman Z-Score Normalization.....  | 93 |
| 4.1.7                                  | Implementasi Pivoting.....                                | 96 |
| 4.1.8                                  | Implementasi Euclidean Distance Similarity .....          | 97 |
| 4.1.9                                  | Implementasi Cosine Similarity .....                      | 99 |

|                                           |     |
|-------------------------------------------|-----|
| 4.1.10 Implementasi Single Linkage .....  | 101 |
| 4.1.11 Implementasi Complete Linkage..... | 102 |
| 4.1.12 Implementasi Antarmuka .....       | 103 |
| 4.1.12.1 Halaman Login .....              | 103 |
| 4.1.12.2 Halaman Tabel Fakta.....         | 104 |
| 4.1.12.3 Halaman Normalisasi .....        | 105 |
| 4.1.12.4 Halaman Pivoting .....           | 105 |
| 4.1.12.5 Halaman Scatter.....             | 106 |
| 4.1.12.6 Halaman Similarity.....          | 107 |
| 4.1.12.7 Halaman Klasterisasi .....       | 108 |
| 4.1.12.8 Halaman Lihat Data.....          | 109 |
| 4.1.12.9 Halaman Manajemen User .....     | 109 |
| 4.2 Analisis Sistem .....                 | 110 |
| 4.2.1 Analisis Validasi Sistem.....       | 111 |
| 4.2.2 Analisis Pemotongan Cluster .....   | 113 |
| 4.2.3 Analisis Akurasi Metode .....       | 120 |
| 4.2.4 Analisis Perilaku.....              | 126 |
| BAB V KESIMPULAN DAN SARAN .....          | 134 |
| 5.1 Kesimpulan.....                       | 134 |
| 5.2 Saran .....                           | 135 |
| DAFTAR PUSTAKA .....                      | 137 |
| LAMPIRAN.....                             | 139 |

## DAFTAR GAMBAR

|                                                                           |    |
|---------------------------------------------------------------------------|----|
| Gambar 2. 1 Contoh Dendrogram dalam Average Linkage (Laeli, 2014) .....   | 10 |
| Gambar 2. 2 Langkah Single Linkage (Handoyo, Rumani & Nasution, 2014).... | 12 |
| Gambar 2. 3 Dendrogram Studi Kasus Single Linkage .....                   | 18 |
| Gambar 2. 4 Dendrogram Studi Kasus Complete Linkage.....                  | 21 |
| Gambar 2. 5 Dendrogram Single Linkage .....                               | 22 |
| Gambar 2. 6 Dendrogram Ward Linkage .....                                 | 23 |
| Gambar 2. 7 Dendrogram Complete Linkage.....                              | 23 |
| Gambar 2. 8 Dendrogram Average Linkage.....                               | 24 |
| Gambar 3. 1 Use Case Diagram Sistem.....                                  | 31 |
| Gambar 3. 2 Arsitektur Sistem.....                                        | 37 |
| Gambar 3. 3 Alur proses sistem.....                                       | 39 |
| Gambar 3. 4 Flowchart Proses ETL.....                                     | 40 |
| Gambar 3. 5 Flowchart Min-Max Normalization .....                         | 41 |
| Gambar 3. 6 Flowchart Z-Score Normalization .....                         | 42 |
| Gambar 3. 7 . Flowchart Pivoting.....                                     | 43 |
| Gambar 3. 8 Flowchart Euclidean Distance .....                            | 44 |
| Gambar 3. 9 Flowchart Cosine Similarity .....                             | 45 |
| Gambar 3. 10 Flowchart Single Linkage .....                               | 46 |
| Gambar 3. 11 Flowchart Complete Linkage.....                              | 47 |
| Gambar 3. 12 Model Dimensional Database .....                             | 54 |
| Gambar 3. 13 Halaman Login.....                                           | 55 |
| Gambar 3. 14 Halaman Input Data Uji .....                                 | 55 |
| Gambar 3. 15 Halaman Proses Tabel Fakta.....                              | 56 |
| Gambar 3. 16 Halaman Normalisasi .....                                    | 56 |
| Gambar 3. 17 Halaman Pivoting.....                                        | 57 |

|                                                                                 |    |
|---------------------------------------------------------------------------------|----|
| Gambar 3. 18 Halaman Similarity .....                                           | 57 |
| Gambar 3. 19 Halaman Dendrogram .....                                           | 58 |
| Gambar 3. 20 Halaman Lihat Data .....                                           | 59 |
| Gambar 3. 21 Halaman Manajemen User .....                                       | 59 |
| Gambar 3. 22 Alur pengujian validasi sistem .....                               | 60 |
| Gambar 3. 23 Alur pengujian analisis sistem .....                               | 61 |
| Gambar 4. 1 Topologi arsitektur client-server .....                             | 69 |
| Gambar 4. 2 Struktur Hadoop Ecosystem.....                                      | 70 |
| Gambar 4. 3 Daftar aplikasi yang berjalan pada cluster .....                    | 75 |
| Gambar 4. 4 Tampilan Hadoop di Browser .....                                    | 75 |
| Gambar 4. 5 Hiveserver2 .....                                                   | 77 |
| Gambar 4. 6 Hive dalam HDFS .....                                               | 77 |
| Gambar 4. 7 Query transformasi jumlah per strip menjadi per transaksi bon ..... | 85 |
| Gambar 4. 8 Hasil transformasi jumlah per strip menjadi per transaksi bon.....  | 86 |
| Gambar 4. 9 Proses ETL untuk membentuk model dimensional .....                  | 88 |
| Gambar 4. 10 Transformasi get_table_name .....                                  | 88 |
| Gambar 4. 11 Transformasi mysql_to_hive.....                                    | 88 |
| Gambar 4. 12 Query tabel dimensi_pelanggan.....                                 | 89 |
| Gambar 4. 13 Query tabel dimensi_strip .....                                    | 90 |
| Gambar 4. 14 Query tabel fakta .....                                            | 91 |
| Gambar 4. 15 Query proses data cleaning .....                                   | 91 |
| Gambar 4. 16 Query proses data filtering.....                                   | 92 |
| Gambar 4. 17 Source code Min-Max Normalization.....                             | 92 |
| Gambar 4. 18 Source code Min-Max Normalization (lanjutan) .....                 | 93 |
| Gambar 4. 19 Source code Z-Score Normalization .....                            | 94 |
| Gambar 4. 20 Source code Z-Score Normalization (lanjutan).....                  | 95 |
| Gambar 4. 21 Source code pivoting.....                                          | 96 |
| Gambar 4. 22 Source code pivoting (lanjutan) .....                              | 97 |
| Gambar 4. 23 Source code Euclidean Distance Similarity .....                    | 98 |
| Gambar 4. 24 Source code Euclidean Distance Similarity (lanjutan).....          | 99 |
| Gambar 4. 25 Source code Cosine Similarity .....                                | 99 |

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| Gambar 4. 26 Source code Cosine Similarity (lanjutan) .....       | 100 |
| Gambar 4. 27 Source code Agglomerative Clustering .....           | 101 |
| Gambar 4. 28 Source code Agglomerative Clustering (lanjutan)..... | 102 |
| Gambar 4. 29 Source code Single Linkage.....                      | 102 |
| Gambar 4. 30 Source code Complete Linkage .....                   | 103 |
| Gambar 4. 31 Interface halaman login.....                         | 104 |
| Gambar 4. 32 Interface halaman tabel fakta .....                  | 104 |
| Gambar 4. 33 Interface halaman tambah data uji .....              | 105 |
| Gambar 4. 34 Interface halaman normalisasi.....                   | 105 |
| Gambar 4. 35 Interface halaman pivoting.....                      | 106 |
| Gambar 4. 36 Interface halaman scatter.....                       | 107 |
| Gambar 4. 37 Interface halaman similarity.....                    | 107 |
| Gambar 4. 38 Interface halaman klasterisasi .....                 | 108 |
| Gambar 4. 39 Interface halaman dendrogram.....                    | 108 |
| Gambar 4. 40 Interface halaman lihat data .....                   | 109 |
| Gambar 4. 41 Interface halaman manajemen user .....               | 110 |
| Gambar 4. 42 Interface halaman edit user .....                    | 110 |
| Gambar 4. 43 Scatter klaster Maret-Mei 2010 Metode I .....        | 124 |
| Gambar 4. 44 Scatter klaster Maret-Mei 2010 Metode II.....        | 124 |
| Gambar 4. 45 Scatter klaster Maret-Mei 2010 Metode III .....      | 125 |
| Gambar 4. 46 Scatter klaster Maret-Mei 2010 Metode IV .....       | 125 |

## DAFTAR TABEL

|                                                                             |    |
|-----------------------------------------------------------------------------|----|
| Tabel 2. 1 Contoh Dataset.....                                              | 12 |
| Tabel 2. 2 Hasil Perhitungan Similarity Euclidean Distance.....             | 14 |
| Tabel 2. 3 Hasil Perhitungan Similarity Manhattan Distance.....             | 16 |
| Tabel 2. 4 Hasil Perhitungan Similarity Manhattan Distance (lanjutan) ..... | 16 |
| Tabel 2. 5 Hasil perhitungan awal Single Linkage .....                      | 17 |
| Tabel 2. 6 Hasil akhir perhitungan Single Linkage .....                     | 17 |
| Tabel 2. 7 Contoh Dataset Complete Linkage .....                            | 19 |
| Tabel 2. 8 Contoh Dataset Complete Linkage (lanjutan).....                  | 19 |
| Tabel 2. 9 Hasil perhitungan Similarity dengan Manhattan Distance .....     | 19 |
| Tabel 2. 10 Hasil perhitungan awal Complete Linkage.....                    | 20 |
| Tabel 2. 11 Hasil perhitungan akhir Complete Linkage .....                  | 21 |
| Tabel 2. 12 Label Cluster Single Linkage.....                               | 22 |
| Tabel 2. 13 Label Cluster Ward Linkage.....                                 | 23 |
| Tabel 2. 14 Label Cluster Complete Linkage .....                            | 24 |
| Tabel 2. 15 Label Cluster Average Linkage .....                             | 24 |
| Tabel 2. 16 Data Perhitungan Complete Linkage .....                         | 25 |
| Tabel 2. 17 Data Cluster Complete Linkage.....                              | 26 |
| Tabel 2. 18 Data Sampling Metode Binning .....                              | 26 |
| Tabel 3. 1 Tabel d_bon .....                                                | 48 |
| Tabel 3. 2 Tabel d_nota_tunai .....                                         | 49 |
| Tabel 3. 3 Tabel m_bon .....                                                | 50 |
| Tabel 3. 4 Tabel m_nota_tunai .....                                         | 50 |
| Tabel 3. 5 Tabel m_nota_tunai (lanjutan).....                               | 51 |

|                                                                                               |     |
|-----------------------------------------------------------------------------------------------|-----|
| Tabel 3. 6 Tabel strip .....                                                                  | 51  |
| Tabel 3. 7 Tabel pelanggan .....                                                              | 52  |
| Tabel 3. 8 Tabel pelanggan (lanjutan) .....                                                   | 53  |
| Tabel 3. 9 Tabel 10 Data fakta acak penjualan bon 1 Januari – 3 Januari 2011...               | 62  |
| Tabel 3. 10 Tabel Min-Max Normalization bon 1 Januari – 3 Januari 2011 .....                  | 62  |
| Tabel 3. 11 Tabel Z-Score Normalization bon 1 Januari – 3 Januari 2011 .....                  | 63  |
| Tabel 3. 12 Tabel Z-Score bon 1 Januari – 3 Januari 2011 (lanjutan).....                      | 63  |
| Tabel 3. 13 Tabel data pivoting bon 1 Januari – 3 Januari 2011 .....                          | 64  |
| Tabel 3. 14 Tabel Similarity Euclidean Distance 1 Januari – 3 Januari 2011 .....              | 65  |
| Tabel 3. 15 Tabel Similarity Cosine Similarity 1 Januari – 3 Januari 2011 .....               | 65  |
| Tabel 3. 16 Tabel Similarity Cosine Similarity 1 Januari – 3 Januari 2011<br>(lanjutan) ..... | 65  |
| Tabel 3. 17 Tabel Single Linkage 1 Januari – 3 Januari 2011 .....                             | 66  |
| Tabel 3. 18 Tabel Complete Linkage 1 Januari – 3 Januari 2011.....                            | 66  |
| Tabel 4. 1 Statistik frekuensi jumlah transaksi pembelian di setiap nota .....                | 80  |
| Tabel 4. 2 Statistik frekuensi jumlah transaksi pembelian di setiap nota (lanjutan)<br>.....  | 81  |
| Tabel 4. 3 Statistik frekuensi jumlah transaksi pembelian di setiap nota (lanjutan)<br>.....  | 82  |
| Tabel 4. 4 Statistik frekuensi jumlah transaksi pembelian di setiap nota (lanjutan)<br>.....  | 83  |
| Tabel 4. 5 Statistik data nota bon.....                                                       | 83  |
| Tabel 4. 6 Struktur tabel transaksidatabon.....                                               | 84  |
| Tabel 4. 7 Hasil normalisasi data validasi .....                                              | 111 |
| Tabel 4. 8 Hasil pivoting.....                                                                | 111 |
| Tabel 4. 9 Hasil pivoting (lanjutan) .....                                                    | 112 |
| Tabel 4. 10 Hasil Cosine Similarity .....                                                     | 112 |
| Tabel 4. 11 Perbandingan nilai sistem dan manual.....                                         | 113 |
| Tabel 4. 12 Metode penelitian .....                                                           | 114 |
| Tabel 4. 13 Data uji penelitian .....                                                         | 114 |
| Tabel 4. 14 Sample klaster manual dari metode binning.....                                    | 115 |

|                                                                                           |     |
|-------------------------------------------------------------------------------------------|-----|
| Tabel 4. 15 Sample klaster manual dari metode binning (lanjutan) .....                    | 115 |
| Tabel 4. 16 Hasil klaster Juni 2012 dengan Complete Linkage.....                          | 117 |
| Tabel 4. 17 Hasil klaster Juni 2012 dengan Complete Linkage (lanjutan) .....              | 118 |
| Tabel 4. 18 Jumlah dominan pada klaster Juni 2012 dengan Complete Linkage                 | 119 |
| Tabel 4. 19 Hasil purity dari data uji I hingga IV .....                                  | 120 |
| Tabel 4. 20 Hasil purity dari data uji I hingga VII.....                                  | 121 |
| Tabel 4. 21 Hasil pengurutan metode yang terbaik .....                                    | 122 |
| Tabel 4. 22 Jarak similarity untuk pembentukan 12 klaster .....                           | 122 |
| Tabel 4. 23 Jarak similarity untuk pembentukan 12 klaster (lanjutan).....                 | 123 |
| Tabel 4. 24 Data uji analisis perilaku.....                                               | 126 |
| Tabel 4. 25 Kuartal pengujian.....                                                        | 126 |
| Tabel 4. 26 Hasil pembelian dari klaster tahun 2010 dengan purity 75% .....               | 127 |
| Tabel 4. 27 Hasil pembelian dari klaster tahun 2011 dengan purity 75% .....               | 127 |
| Tabel 4. 28 Hasil pembelian dari klaster tahun 2011 dengan purity 75% (lanjutan)<br>..... | 128 |
| Tabel 4. 29 Hasil pembelian dari klaster tahun 2012 dengan purity 75% .....               | 128 |
| Tabel 4. 30 Hasil pembelian dari klaster tahun 2010 dengan purity 100% .....              | 128 |
| Tabel 4. 31 Hasil pembelian dari klaster tahun 2011 dengan purity 100% .....              | 129 |
| Tabel 4. 32 Hasil pembelian dari klaster tahun 2012 dengan purity 100% .....              | 129 |
| Tabel 4. 33 Hasil nilai pembentukan klaster.....                                          | 130 |
| Tabel 4. 34 Hasil analisis dominan pembelian nota bon 2010-2012 .....                     | 131 |
| Tabel 4. 35 Hasil analisis dominan pembelian nota tunai 2010-2012 .....                   | 131 |

## INTISARI

### **PENERAPAN AGGLOMERATIVE CLUSTERING PADA NOTA PELANGGAN PIUTANG RITEL FASHION DAN SEPATU**

*Agglomerative Clustering* merupakan salah satu metode yang membentuk tingkatan tertentu seperti struktur pada pohon dikarenakan proses dilakukan secara bertahap untuk mendapatkan *cluster*. Ada beberapa algoritma yang dapat digunakan dalam melakukan *Agglomerative Clustering*, diantaranya adalah algoritma *Single Linkage* dan *Complete Linkage*. Dalam penelitian ini, penulis menggunakan algoritma *Single Linkage* dan *Complete Linkage* beserta metode normalisasi *Min-Max* dan *Z-Score* serta metode *similarity Euclidean Distance* dan *Cosine Similarity* untuk mengelompokkan perilaku pelanggan menjadi *cluster* pada transaksi penjualan bon dan tunai pada perusahaan ritel fashion dan sepatu dari tahun 2010-2012. Evaluasi dilakukan dengan menghitung nilai *purity cluster* menggunakan periode waktu kuartal. Pengujian menggunakan metode yang sama berdasar nilai *purity* yang terbaik untuk setiap kuartalnya. Berdasarkan hasil analisis yang dilakukan, ditemukan bahwa *Complete Linkage* dengan menggunakan *Min-Max* serta *Euclidean* menghasilkan nilai *purity* paling baik serta sistem tidak dapat memprediksi perilaku pelanggan dalam pembelian barang pada setiap tahunnya baik pada penjualan bon atau tunai dikarenakan pola klaster dominan yang berbeda dan tidak dapat memprediksi golongan umur yang dominan dikarenakan golongan yang dominan dalam pembelian berdasar pada nilai minimum dan maksimum data pada setiap kuartal. Selain itu ditemukan bahwa data penelitian tidak cocok untuk dilakukan klasterisasi maupun analisis perilaku pelanggan piutang.

**Kata Kunci:** Agglomerative Clustering, Single Linkage, Complete Linkage, Data Mining, Implementasi Algoritma, Big Data, Data Warehouse, Euclidean Distance, Cosine Similarity, Z-Score, Min-Max

## DAFTAR ISI

|                                                  |      |
|--------------------------------------------------|------|
| HALAMAN JUDUL .....                              | ii   |
| PERNYATAAN KEASLIAN TUGAS AKHIR .....            | iii  |
| HALAMAN PERSETUJUAN.....                         | iv   |
| HALAMAN PENGESAHAN .....                         | v    |
| UCAPAN TERIMA KASIH.....                         | vi   |
| KATA PENGANTAR .....                             | viii |
| INTISARI .....                                   | ix   |
| DAFTAR ISI.....                                  | x    |
| DAFTAR GAMBAR .....                              | xiv  |
| DAFTAR TABEL.....                                | xvii |
| BAB I PENDAHULUAN.....                           | 1    |
| 1.1 Latar Belakang.....                          | 1    |
| 1.2 Rumusan Masalah .....                        | 3    |
| 1.3 Batasan Sistem.....                          | 3    |
| 1.4 Tujuan Penelitian.....                       | 4    |
| 1.5 Metode Penelitian.....                       | 4    |
| 1.6 Sistematika Penulisan.....                   | 5    |
| BAB II TINJAUAN PUSTAKA .....                    | 7    |
| 2.1 Tinjauan Pustaka .....                       | 7    |
| 2.2 Landasan Teori .....                         | 9    |
| 2.2.1 Analisis Clustering .....                  | 9    |
| 2.2.2 Hierarchical Agglomerative Clustering..... | 9    |
| 2.2.3 Single Linkage.....                        | 11   |

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Dalam dunia bisnis pada jaman sekarang, para pelaku bisnis senantiasa selalu berusaha mengembangkan cara-cara untuk dapat mengembangkan usaha mereka dan memperhatikan segala aspek yang ingin diperhatikan agar tidak salah langkah dalam pengambilan keputusan. Salah satu aspek yang berpengaruh terhadap stabilitas perusahaan ialah aspek keuangan. Perusahaan perlu mengatur neraca keuangan yang terdapat dalam perusahaan tersebut agar tidak mengalami kerugian yang sangat berdampak bagi kelangsungan perusahaan. Untuk dapat mengatur aspek keuangan tersebut ada banyak cara yang dapat ditempuh salah satunya ialah dengan melakukan analisis data perusahaan. Salah satu perusahaan yang bergerak di bidang ritel fashion dan sepatu akan diteliti oleh penulis. Di dalam bisnis ritel tersebut perilaku konsumen sangat berpengaruh terhadap keseimbangan keuangan perusahaan. Perusahaan perlu memperhatikan dan menganalisis perilaku pelanggan yang secara tidak langsung berpengaruh terhadap stabilitas keuangan perusahaan contohnya ialah perilaku pembelian barang piutang para pelanggan. Dengan diketahuinya perilaku pelanggan piutang, maka perusahaan dapat menyeleksi kelompok jenis barang yang berpotensi untuk dijual pada kuartal tertentu terhadap para pelanggan.

Salah satu cara agar dapat menganalisis perilaku pelanggan piutang adalah dengan mengoleksi seluruh data yang berkaitan dengan pelanggan tersebut termasuk informasi-informasi yang dibutuhkan agar dapat menggali data yang tepat dalam pengambilan keputusan untuk solusi bisnis. Cara tersebut dinamakan dengan *data mining*. Data yang akan digunakan dalam *data mining* yang akan diteliti akan menggunakan *big data* dari perusahaan ritel fashion dan sepatu yang akan nantinya diolah dengan metode *data warehouse*. Konsep dari *data*

*warehouse* sendiri ialah mampu memberikan kemampuan *query* terhadap *big data* dengan cepat. Sumber yang didapat merupakan data dari perusahaan tersebut. Data yang sudah diolah tersebut kemudian akan diolah lagi agar mempermudah analisis dengan menggunakan konsep *Business Intelligence*. Dalam pengolahan data, *Business Intelligence* menggunakan *data warehouse* sebagai sumber datanya. Contoh kasus dalam dunia nyata ialah apabila ada ribuan lembar nota penjualan yang dibayarkan secara piutang dan digambarkan suatu model relasional transaksi penjualan maka yang terjadi ialah banyaknya tabel yang berelasi. Kondisi tersebut tidaklah bagus dalam *Business Intelligence* dan oleh sebab itu perlu untuk dikonversi dengan pendekatan ETL (*Extract, Transform, dan Load*) menjadi model *data warehouse* (model dimensi). Oleh karena yang akan dilakukan adalah analisis *clustering* untuk analisis perilaku penjualan piutang, maka perlu dilakukan proses *pivoting* terhadap data-data penjualan yang sudah tersimpan dalam model *data warehouse* yang per record akan menyatakan seperti sebuah nota. Dengan dikelompokkan secara otomatis, dapat dilihat kelompok-kelompok transaksi penjualan jika dilakukan *clustering* berdasarkan karakter barang yang dijual.

Salah satu metode dalam *data mining* yang digunakan dalam penelitian ini adalah *clustering* dimana metode tersebut mengidentifikasi objek yang memiliki kesamaan karakteristik tertentu, kemudian menggunakan karakteristik tersebut sebagai “vektor karakteristik” atau “centroid”. Ada beberapa metode yang dapat digunakan untuk mengolah karakteristik tersebut yaitu *Single Linkage, Complete Linkage, Average Linkage, Ward Method* dan *Centroid Method*. Proses *clustering* yang akan dilakukan menggunakan *Agglomerative Hierarchical Clustering Algorithm (HAC Algorithm)* *Single Linkage* dan *Complete Linkage*. Sistem yang akan dibuat akan menerapkan konsep *data warehouse* dan *Business Intelligence*, mulai dari penggambaran arsitektur sistem yang ada, membuat ETL (*Extract, Transformation dan Load*) dan dimension model serta *pivoting* data, lalu hasil yang didapat dikeluarkan dalam bentuk OLAP (*Online Analytical Processing*) dan kemudian akan dimulai proses *clustering* data.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang ada, rumusan masalah yang akan dibahas dalam penelitian ini adalah :

- Bagaimana sistem dapat menerapkan data mining clustering pada nota pelanggan piutang di salah satu perusahaan ritel fashion dan sepatu ?
- Bagaimana hasil perbandingan kualitas *clustering* berdasarkan nilai *purity* antara metode *Single Linkage* dan *Complete Linkage* untuk studi kasus nota pelanggan piutang di salah satu perusahaan ritel fashion dan sepatu ?

## 1.3 Batasan Sistem

Adapun batasan sistem yang akan dibuat oleh peneliti adalah sebagai berikut :

1. Penelitian akan membandingkan antara metode *Complete Linkage Hierarchical* dan *Single Linkage Hierarchical* dengan metode normalisasi *Z-Score*, *Min-Max*, dan metode *similarity Euclidean Distance* atau *Cosine Similarity*.
2. *Big data* akan diolah di *Hadoop* dan *Hive*, kemudian *back-end* akan diolah di *java server pages* menggunakan *Tomcat Apache Server*.
3. Persebaran data akan digambarkan dengan grafik *scatter plot*.
4. Output *cluster* akan berbentuk grafik *dendrogram*, *scatter* grafik *cluster* dan tabel serta sistem berbasis web.
5. Hasil dari analisa tersebut adalah terbentuknya *cluster* perilaku piutang pelanggan di sebuah perusahaan ritel fashion dan sepatu.

#### 1.4 Tujuan Penelitian

Berdasarkan permasalahan yang diteliti, maka tujuan diadakan penelitian ini adalah membuat sebuah aplikasi yang dapat menerapkan proses klasterisasi pada nota pelanggan piutang terhadap salah satu perusahaan ritel fashion dan sepatu dengan membandingkan hasil *clustering* antara metode *Single Linkage* dan *Complete Linkage*.

#### 1.5 Metode Penelitian

Berikut ini adalah metodologi – metodologi yang akan digunakan dalam penelitian ini :

- **Metode Pengumpulan Data**

- **Studi Pustaka**

Penulis melakukan studi pustaka dengan cara mempelajari teori-teori dan literature yang mendukung penyelesaian penelitian yang berhubungan dengan *data mining*, *data warehouse*, *big data*, *Agglomerative Clustering* (*Single Linkage* dan *Complete Linkage*) serta perangkat lunak yang akan dipakai untuk membangun sistem.

- **Pengambilan Data**

Penulis mendapatkan *big data* dari salah satu perusahaan ritel fashion dan sepatu di Indonesia sebagai bahan dalam pengolahan *data warehouse* yang digunakan dalam penelitian. Data yang akan dianalisa dan digunakan merupakan data penjualan barang dari salah satu perusahaan ritel fashion dan sepatu dalam jangka waktu 2 tahun terhitung dari 2010 – 2012.

- **Metode Perancangan Sistem**

Tahap ini merupakan tahap peneliti melakukan perancangan dimulai dari sistem yang akan dibangun dan perancangan antarmuka. Output hasil yang akan dikeluarkan nantinya akan berbentuk sebuah grafik *dendrogram* dan tabel yang menjelaskan hasil klaster pada nota piutang pelanggan ritel fashion dan sepatu sesuai dengan penentuan *feature*.

*Feature* yang akan digunakan peneliti adalah nama umur dan empat kelompok jenis barang yang terdapat pada data berupa kelompok barang berjenis Anak, Wanita, Pria dan Sepatu.

- **Metode Pendekatan Pivoting**

Pivoting merupakan suatu cara yang membuat peneliti dapat melakukan pengaturan ulang dan perubahan suatu kolom dan baris dari suatu tabel *database* untuk mendapatkan suatu tabel baru yang terlihat dari sudut pandang yang lain tanpa mengubah tabel atau *database* yang lama sebelum dimasukkan ke dalam *data warehouse*.

- **Metode Pengembangan Sistem**

Metode pengembangan sistem yang dipakai oleh peneliti adalah *Single Linkage* dan *Complete Linkage* yang merupakan pengembangan dari *Hierarchical Clustering*.

- **Konsultasi**

Melakukan konsultasi dengan dosen pembimbing agar dapat melakukan penelitian secara terarah.

- **Metode Evaluasi**

Metode evaluasi yang akan digunakan adalah melakukan perbandingan antara hasil *clustering* baik dari *Single Linkage* maupun *Complete Linkage* yang berasal dari keluaran dari sistem dengan hasil *clustering* secara manual oleh penulis serta antar nota piutang dan juga nota tunai.

## 1.6 Sistematika Penulisan

Sistematika penulisan laporan Tugas Akhir ini dikelompokkan menjadi 5 bab, yaitu:

Bab I Pendahuluan. Bab ini berisi gambaran tentang penelitian yang dilakukan. Bab ini terbagi menjadi beberapa bagian, yaitu latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian dan sistematika penulisan.

Bab II Tinjauan Pustaka. Bab ini berisi tentang teori yang menjadi landasan dalam pembuatan penelitian. Teori-teori tersebut diambil dari jurnal penelitian, maupun sumber-sumber lain yang dapat mendukung penelitian. Dasar teori dan tinjauan pustaka ini menjadi acuan dalam pembuatan analisis data dan pengambilan kesimpulan penelitian.

Bab III Analisis dan Perancangan Sistem. Bab ini menjelaskan mengenai tahap perancangan sistem yang meliputi kebutuhan *hardware* dan *software*, spesifikasi sistem, *flowchart* program, *use case* dari sistem, skema basis data, desain antar muka serta rancangan pengujian sistem.

Bab IV Implementasi dan Analisis Sistem. Bab ini berisi tentang hasil implementasi algoritma serta analisa berkaitan dengan masalah yang ada. Penjelasan disajikan dalam bentuk deskripsi, tabel, grafik, dan gambar pendukung.

Bab V Kesimpulan dan Saran. Bab ini berisi kesimpulan dari penelitian yang dilakukan beserta saran pengembangan sistem untuk penelitian yang memiliki topik yang sama dengan topik penelitian ini.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Dalam penelitian ini, pembentukan klaster perilaku pelanggan piutang perusahaan ritel fashion dan sepatu menggunakan *Agglomerative Clustering* telah berhasil dikembangkan. Berdasarkan hasil pengujian dan analisis yang telah dilakukan oleh penulis, maka dapat disimpulkan:

1. Untuk dapat melakukan analisis klaster menggunakan algoritma *Agglomerative Clustering* diperlukan tahapan *data preprocessing*. *Data preprocessing* yang harus dilakukan terhadap data transaksi penjualan adalah menghapus transaksi yang memiliki *field* bernilai *NULL*. Transaksi yang digunakan hanyalah transaksi pembelian per pelanggan yang memiliki jumlah transaksi lebih dari 0. Untuk menentukan batasan jumlah barang yang terbeli dilakukan penghitungan *IQR*. Penentuan *IQR* digunakan untuk mencari *outlier*. Data yang merupakan *outlier* kemudian tidak digunakan. Hal yang sama dilakukan untuk memproses data pelanggan yang masih banyak terlihat *field* bernilai *NULL*.
2. Dari hasil pengujian ditemukan algoritma *Complete Linkage* memiliki nilai *purity* yang lebih baik dibandingkan dengan algoritma *Single Linkage* dengan hasil rata-rata *purity* 0,4 seperti terletak pada halaman 117. Selain itu dari 7 data uji, ditemukan bahwa metode normalisasi dan *similarity* yang paling baik adalah dengan menggunakan metode *Min-Max* dan *Euclidean Distance* yang 100% menghasilkan nilai *purity* terbaik di semua data uji. Namun *purity* terbaik yang dihasilkan oleh sistem masih di bawah 0,5. Hal ini disebabkan oleh faktor data

sampling manual menggunakan metode *binning* yang dinilai tidak cocok menjadi data yang dibandingkan. Sehingga dapat dikatakan juga bahwa metode *binning* tidak tepat untuk melakukan pengelompokan secara manual untuk klusterisasi.

3. Dari hasil analisis perilaku ditemukan bahwa sistem tidak dapat memprediksi klaster yang mempunyai nilai dominan pembelian barang yang sama untuk setiap kuartal setiap tahunnya. Hal ini dikarenakan dalam hasil penelitian setiap tahunnya pada transaksi bon hanya 25% yang memiliki pola yang sama dan 50% pada transaksi tunai. Selain itu pola perilaku nota tunai dan bon dalam pembelian barang pada setiap kuartalnya juga berbeda-beda.
4. Data uji yang digunakan yaitu data transaksi penjualan pada perusahaan adalah data yang buruk dan tidak cocok digunakan untuk proses klusterisasi dikarenakan persebaran nilai data yang masih kotor serta tidak adanya karakteristik khusus yang menggambarkan kedekatan masing-masing nota sesuai *feature* yang ditentukan. Oleh karena data uji yang buruk sistem juga tidak dapat digunakan untuk membantu analisis perilaku pelanggan piutang.
5. Dari hasil sistem, penulis juga berhasil membuat library baru untuk perhitungan *Single Linkage* serta *Complete Linkage* yang dapat digunakan oleh masyarakat umum.

## 5.2 Saran

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

1. Data yang digunakan untuk analisis transaksi diperbesar jangka waktunya menjadi 10 tahun supaya hasil dari perilaku pelanggan dalam klaster lebih detail.
2. Data yang dihitung kedekatannya berdasarkan kedekatan antar kode barang bukan kelompok jenis barang supaya hasil pengelompokan klaster terlihat lebih bagus.

3. Penghitungan *outlier* dilakukan oleh sistem supaya dapat mengantisipasi kemungkinan data yang lebih besar.
4. Sampling manual yang dibandingkan untuk perhitungan *purity* yang didapat berasal dari perusahaan yang dianalisa supaya hasil yang didapat lebih presisi.
5. Arsitektur *backend* untuk *query* data pada ekosistem *Hadoop* diubah menggunakan aplikasi yang didesain untuk pengolahan data *realtime* misalnya seperti penggunaan *Apache Spark*.
6. Sistem dapat menganalisis efektifitas and efisiensi waktu yang dihasilkan oleh algoritma untuk mendapatkan hasil dari efisiensi penggunaan algoritma.

## DAFTAR PUSTAKA

- Chen, M., Mao, S., & Liu, Y. (2014). *Big Data A Survey*. New York: Springer Science and Bussiness Media.
- Chen, Y., Alspaugh, S., & Katz, R. (2012). *Interaactive Analytical Processing in Big Data Systems: A Cross-Industry Study of MapReduce Workloads*. Berkeley: University of California.
- Dhillon, I.S., Modha, D.S., & Spangler, W.S. (2001). *Visualizing Class Structure of Multidimensional Data*. San Jose: IBM Almaden Research Center.
- Eckerson, & Wayne, W. (2011). *Performance Dashboards: Measuring, Monitoring, and Managing Your Bussiness*. Indianapolis: Wiley Publishing, Inc.
- Hall, M., & Brown, L. (2004). *Core SERVLETS and JAVASERVER PAGES*. Santa Clara, California.
- Handoyo, R., Rumani, M. R., & Nasution, M. S. (2014). *Perbandingan Metode Clustering Menggunakan Metode Single Linkage Dan K-Means Pada Pengelompokkan Dokumen*. Bandung: Universitas Telkom.
- Hidayat, G. Z., & Supriyanto, C. (2011). *Pengelompokan Citra Rambu Lalu Lintas Dengan Hierarchical Aggloromorative Clustering Berbasis Scale Invariant Feature Transform*. Semarang: Universitas Dian Nuswantoro.
- Imbar, R.V., Adelia, Ayub M., & Rehatta A. (2014). *Implementasi Cosine Similarity dan Algoritma Smith-Waterman Untuk Mendeteksi Kemiripan Teks*. Bandung: Universitas Kristen Maranatha.
- Junaedi, H., Budianto H., Maryati I. & Melani Y. (2011). *Data Transformation Pada Data Mining*. Surabaya: Sekolah Tinggi Teknik Surabaya.
- Laeli, S. (2014). *Analisis Cluster Dengan Average Linkage Method dan Ward's Method Untuk Data Responden Nasabah Asuransi Jiwa Unit Link*. Yogyakarta: Universitas Negeri Yogyakarta.
- 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*. USA: Mckinsey Global Institute.

- Marsudi E.K. (2014). *Sistem Pencarian Komunitas Menggunakan Hierarchical Clustering Berbasis Modularity Pada Studi Kasus Twitter*. (Undergraduate thesis, Duta Wacana Christian University, 2014). Retrieved from <http://sinta.ukdw.ac.id>
- Patro, S. G. K., Kumar K., (2015). *Normalization: A Processing Stage*. India: Department of CSE & IT, VSSUT, Burla.
- Simamora, B. (2005). *Analisis Multivariant Pemasaran Edisi Pertama*. Jakarta: PT. Gramedia Pustaka Utama.
- Soraya, Y. (2011). *Perbandingan Kinerja Metode Single Linkage, Metode Complete Linkage Dan Metode K-Means Dalam Analisis Cluster*. Semarang: Universitas Negeri Semarang.