PENERAPAN NEAREST NEIGHBOR CLUSTERING TERHADAP DATAWAREHOUSE TRANSAKSI PENJUALAN DENGAN PENDEKATAN PIVOTING

Skripsi

oleh
YOAS HERNANDA
71120095

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

PENERAPAN NEAREST NEIGHBOR CLUSTERING TERHADAP DATAWAREHOUSE TRANSAKSI PENJUALAN DENGAN PENDEKATAN PIVOTING

Skripsi

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

Disusun oleh

YOAS HERNANDA
71120095

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

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

PENERAPAN NEAREST NEIGHBOR CLUSTERING TERHADAP DATA WAREHOUSE TRANSAKSI PENJUALAN DENGAN PENDEKATAN PIVOTING

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, 24 Oktober 2016

YOAS HERNANDA
71120095

HALAMAN PERSETUJUAN

Judul Skripsi	:	PENERAPAN NEAREST NEIGHBOR CLUSTERING TERHADAP DATA WAREHOUSE TRANSAKSI PENJUALAN DENGAN PENDEKATAN PIVOTING
Nama Mahasiswa	:	YOAS HERNANDA
N I M	:	71120095
Matakuliah	:	Skripsi (Tugas Akhir)
Kode	:	TIW276
Semester	:	Gasal
Tahun Akademik	:	2016/2017

Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 27 September 2016

Dosen Pembimbing I	Dosen Pembimbing II
Budi Susanto, SKom.,M.T.	Antonius Rachmat C., S.Kom.,M.Cs.

HALAMAN PENGESAHAN

PENERAPAN NEAREST NEIGHBOR CLUSTERING TERHADAP DATA WAREHOUSE TRANSAKSI PENJUALAN DENGAN PENDEKATAN PIVOTING

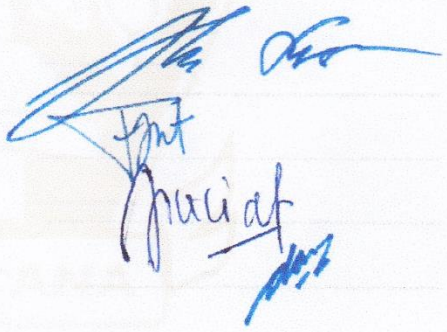
Oleh: YOAS HERNANDA / 71120095

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 12 Oktober 2016

Yogyakarta, 24 Oktober 2016
Mengesahkan,

Dewan Penguji:

1. Budi Susanto, SKom.,M.T.
2. Antonius Rachmat C., S.Kom.,M.Cs.
3. Lucia Dwi Krisnawati, Dr.
4. Danny Sebastian, S.Kom., M.M., M.T.



Dekan

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

Ketua Program Studi

(Gloria Virginia, Ph.D.)

UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan kehadirat Tuhan Yesus Kristus karena atas kasih dan karunia-Nya, skripsi yang berjudul "PENERAPAN *NEAREST NEIGHBOR CLUSTERING* TERHADAP *DATA WAREHOUSE* TRANSAKSI DENGAN PENDEKATAN PIVOTING" 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. Bapak Antonius Rachmat, S.Kom., M.Cs., selaku dosen pembimbing II yang juga telah banyak memberikan masukan dan arahan selama pembuatan skripsi.
3. Bapak Kristian Adi Nugraha, S.Kom., M.T., yang juga selalu memberikan kritik dan saran selama pembuatan skripsi.
4. Kepada salah satu sumber perusahaan ritel fashion dan sepatu di Indonesia yang telah bersedia memberikan data yang digunakan untuk sumber penelitian.

   Kepada keluarga terkasih, Bapak Abednego Heru Supartono, Ibu Yuli Lastini Wartiningsih, dan adik saya Rachel Herlinda yang memberikan dukungan, doa, nasehat, dan motivasi hingga sampai detik ini penulis menyelesaikan studi.

5. Kepada sahabat penulis, Andriana Ripnasari dan Nehemia yang selalu mendukung dalam doa dan semangat.
6. Kepada teman-teman terdekat penulis, Jordan Efrata, Leo Agung Dandy, Johan Sutanto, Prakosa Ananta, Lukas Laksmana, Nanda Kurniawan, Anon Wibawa, Andono Swandaru, Dita Aprilia, Herman Yosef, Nilam Kakansing, Friska Arsha yang selalu memberikan semangat dan caci maki kebahagiaan setiap harinya.
7. Kepada teman-teman jurusan Teknik Informatika 2012, yang senantiasa ada untuk memberikan dukungan, dan sama-sama berjuang untuk menyelesaikan tugas akhir.
8. 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.
9. Kepada Big Data Skripsweet Squad yaitu Constantius Damar Wicaksono, Michael Abadi Santoso, Klaudius Jemly Naban, dan Widnyana Santika yang seperjuangan bersama bahu membahu dalam suka dan duka menyelesaikan topik Big Data.
10. Terakhir, penulis hendak menyapa setiap nama yang tidak dapat penulis cantumkan satu per satu, terima kasih atas doa yang senantiasa mengalir tanpa sepengetahuan penulis, juga kepada spotify.com yang setia menemani penulis dengan musik-musik terkini dalam penulisan penelitian ini.

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 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, September 2016

INTISARI

PENERAPAN *NEAREST NEIGHBOR CLUSTERING* TERHADAP *DATA WAREHOUSE* TRANSAKSI DENGAN PENDEKATAN PIVOTING

Nearest Neighbor Clustering merupakan salah satu metode klasterisasi dalam *Data Mining*. Metode ini melakukan klasterisasi secara bertahap untuk menentukan klaster data dengan melakukan perhitungan jarak data baru dengan semua data lama yang telah dipetakan pada klaster sebelumnya. Hasil perhitungan jarak data baru dengan semua data lama akan dipilih jarak dengan tingkat kemiripan paling besar. Apabila jarak tersebut memenuhi nilai *threshold*, titik baru akan dipetakan pada klaster yang sama sesuai data dengan jarak terdekat. Perhitungan jarak antar data pada penelitian ini menerapkan Metode *Euclidean Distance* dan *Cosine Similarity*, yang didukung Metode Normalisasi *Min-Max* dan *Altman ZScore*. Seluruh metode ini dikombinasi dan di uji menggunakan Data Uji Transaksi Bon dan Tunai Retail Fashion dan Sepatu dengan kurun waktu 3 Tahun, terhitung dari Tahun 2010 sampai Tahun 2012. Evaluasi berdasarkan penentuan nilai *threshold* dan kombinasi metode menghasilkan Kombinasi Metode Normalisasi *Min-Max* dan Metode *Similarity Euclidean Distance* dengan rata-rata nilai *purity* paling tinggi, yaitu 0.299 di atas kombinasi lain yang mencapai nilai paling tinggi tidak terpaut jauh yaitu 0.290 pada Kombinasi Metode Normalisasi ZScore dan Metode *Cosine Similarity*. Berdasarkan hasil evaluasi tersebut, *threshold* dengan nilai 0.35 dan Kombinasi Metode yang paling optimal berdasarkan data uji yang tersedia, yaitu Metode *Min-Max* dan *Euclidean Distance* dipakai pada akhir penelitian untuk meneliti perilaku belanja pelanggan pada setiap kuartal pada keseluruhan data transaksi.

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

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

INTISARI

PENERAPAN *NEAREST NEIGHBOR CLUSTERING* TERHADAP *DATA WAREHOUSE* TRANSAKSI DENGAN PENDEKATAN PIVOTING

Nearest Neighbor Clustering merupakan salah satu metode klasterisasi dalam *Data Mining*. Metode ini melakukan klasterisasi secara bertahap untuk menentukan klaster data dengan melakukan perhitungan jarak data baru dengan semua data lama yang telah dipetakan pada klaster sebelumnya. Hasil perhitungan jarak data baru dengan semua data lama akan dipilih jarak dengan tingkat kemiripan paling besar. Apabila jarak tersebut memenuhi nilai *threshold*, titik baru akan dipetakan pada klaster yang sama sesuai data dengan jarak terdekat. Perhitungan jarak antar data pada penelitian ini menerapkan Metode *Euclidean Distance* dan *Cosine Similarity*, yang didukung Metode Normalisasi *Min-Max* dan *Altman ZScore*. Seluruh metode ini dikombinasi dan di uji menggunakan Data Uji Transaksi Bon dan Tunai Retail Fashion dan Sepatu dengan kurun waktu 3 Tahun, terhitung dari Tahun 2010 sampai Tahun 2012. Evaluasi berdasarkan penentuan nilai *threshold* dan kombinasi metode menghasilkan Kombinasi Metode Normalisasi *Min-Max* dan Metode *Similarity Euclidean Distance* dengan rata-rata nilai *purity* paling tinggi, yaitu 0.299 di atas kombinasi lain yang mencapai nilai paling tinggi tidak terpaut jauh yaitu 0.290 pada Kombinasi Metode Normalisasi ZScore dan Metode *Cosine Similarity*. Berdasarkan hasil evaluasi tersebut, *threshold* dengan nilai 0.35 dan Kombinasi Metode yang paling optimal berdasarkan data uji yang tersedia, yaitu Metode *Min-Max* dan *Euclidean Distance* dipakai pada akhir penelitian untuk meneliti perilaku belanja pelanggan pada setiap kuartal pada keseluruhan data transaksi.

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

BAB I

PENDAHULUAN

1.1. Latar Belakang

Konsumsi, Finansial, semakin menjadi hal yang tidak dapat dipisahkan seiring terus berkembangnya suatu negara. Transaksi, jual, beli, sudah menjadi kata yang sangat *familiar* di kehidupan sehari-hari. Di tengah pesatnya perkembangan dunia transaksi, tidak heran bila semakin banyak ritel dan pasar yang semakin menjamur di seluruh penjuru negeri. Para pembeli pun seakan juga terus bertambah hari demi hari karena tuntutan hidup dan satu atau lain hal. Banyaknya calon pembeli yang terus bertambah dari ke hari membuat retail dan toko-toko bahkan berlomba-lomba untuk menyediakan pelayanan yang terbaik dan tidak jarang dikaji secara berkala untuk para pelanggannya.

Didasarkan pada hal tersebut, penulis akan melakukan penelitian terhadap salah satu ritel yang tengah berkembang di Indonesia. Perusahaan yang berfokus pada fashion dan sepatu ini merupakan ritel yang sudah memiliki banyak cabang di berbagai wilayah dan terus membuka cabang baru di berbagai wilayah baru. Sebagai perusahaan yang tidak lagi kecil dengan cabang yang tidak lagi sedikit, tentu saja tidaklah sedikit transaksi yang terjadi setiap harinya. Akan sangat melelahkan apabila pihak ritel terus menerus mengamati secara *manual* perilaku pembeli. Dengan data transaksi pengunjung hari demi hari yang telah terkumpul dalam jangka waktu tertentu, akan terdapat pola yang bisa ditarik dan dipelajari dari sekumpulan data tersebut. Dari analisis pola yang di dapat, perusahaan yang bersangkutan dapat mengambil keputusan tepat berdasar data riil yang telah terkumpul tersebut. Analisis pola ini akan dilakukan oleh peneliti menggunakan teknik *data mining.*

Data yang akan digunakan peneliti dalam kasus ini adalah data transaksi dari pihak Ritel Fashion dan Sepatu. Data transaksi tersebut merupakan hasil rekapitulasi data dalam kurun waktu selama dua tahun, terhitung sejak tahun 2010

hingga tahun 2012, sehingga dapat digolongkan sebagai *Big Data* karena jumlahnya yang tidak lagi sedikit. Keseluruhan data yang berhasil dikumpulkan nantinya akan diolah dengan teknik *Data Warehouse*. Konsep dari *Data Warehouse* sendiri adalah mampu melakukan proses query terhadap *Big Data* dengan cepat. Data selanjutnya akan diproses dan diolah kembali dengan konsep *Business Intelligence*.

Penelitian sejenis yang pernah dilakukan adalah Perancangan Data Warehouse dan Penerapan Data Mining pada Data Penjualan PD XYZ (Wanajaya, 2011). Penelitian tersebut mempunyai fokus yang hampir sama dengan penelitian penulis, yakni merancang Data Warehouse untuk kemudian diterapkan Data Mining (*Clustering*) untuk mendapatkan informasi-informasi yang dapat membantu perusahaan dalam mengambil keputusan bisnis. Penelitian ini juga mengambil data transaksi penjualan sebagai fokus penelitian utama, namun tidak menerapkan *pivoting* data seperti yang akan diterapkan penulis terhadap Data Transaksi.

Penulis akan menerapkan algoritma yang akan mengelompokkan data berdasarkan kedekatan atributnya, dimana data akan dikelompokkan dengan data lain yang terdekat dengan data tersebut. Hal ini akan sangat penting untuk pembentukan klaster yang optimal untuk diteliti. Berdasarkan kebutuhan tersebut penulis memilih Algoritma *Nearest Neighbor* yang memiliki prinsip mencari jarak kedekatan terbesar antara titik baru dengan titik lama untuk menentukan pemetaan klaster titik baru tersebut.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, dapat dirumuskan bahwa pokok bahasan dalam tugas penelitian ini adalah:

- Bagaimana penerapan *Nearest Neighbor Clustering* terhadap Data Transaksi Pelanggan pada salah satu perusahaan retail fashion dan sepatu?
- Bagaimana perbandingan kualitas *clustering* berdasarkan nilai *purity* dengan kombinasi metode normalisasi *Altman ZScore & Min Max*

serta metode perhitungan jarak *Euclidean Distance* & *Cosine Distance?*

1.3 Batasan Sistem

Adapun batasan sistem dalam penelitian ini adalah sebagai berikut:

1. 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 Tahun 2010 – 2012.
2. Big Data akan disimpan dalam *Hadoop* dengan piranti *Hive,* dan proses *back-end* akan dijalankan di *Java Server Pages* menggunakan *Tomcat Apache Server*.
3. Penelitian akan membandingkan antara metode normalisasi *Altman ZScore* & *Min Max*, serta metode perhitungan jarak *Cosine Distance* & *Euclidean Distance.*
4. Hasil klasterisasi data akan divisualisasikan ke dalam grafik *Scatter Plot.*
5. Hasil akhir dari serangkaian analisa tersebut adalah terbentuknya *cluster* transaksi pengunjung sejenis sesuai sejarah transaksi belanja pengunjung selama dua tahun.

1.4 Tujuan Penelitian

Berdasarkan permasalahan yang diteliti, maka tujuan diadakannya penelitian ini adalah untuk membuat sistem yang dapat menerapkan *Nearest Neighbor Clustering* terhadap Data Transaksi Pelanggan salah satu Retail Fashion & Sepatu dengan metode normalisasi *ZScore* & *Min Max* serta metode perhitungan jarak *Cosine Distance* & *Euclidean Distance.*

1.5 Metodologi Penelitian

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

1. **Metode Pengumpulan Data**
   a. **Studi Pustaka**

      Penulis melakukan studi pustaka dengan cara mempelajari teori-teori dan literatur yang mendukung penyelesaian penelitian yang berhubungan dengan *Data Mining, Data Warehouse, Big Data, Pivoting, Nearest Neighbor Clustering* serta perangkat lunak yang akan dipakai untuk membangun sistem

   b. **Pengambilan Data**

      Penulis mendapatkan *Big Data* dari salah satu Perusahaan Ritel Fashion dan Sepatu sebagai bahan dalam pengolahan *data warehouse* yang digunakan dalam penelitian

2. **Metode Perancangan Sistem**

   Tahap ini merupakan tahap peneliti melakukan perancangan dimulai dari sistem yang akan dibangun dan perancangan antarmuka.

3. **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*.

4. **Metode Pengembangan Sistem**

   Metode pengembangan system yang dipakai oleh peneliti adalah *Nearest Neighbor Clustering*, yang akan dipadukan dengan data *pivoting*.

5. **Metode Evaluasi**

   Evaluasi akan dilakukan menggunakan metode Purity untuk menganalisis tingkat optimalitas clustering yang telah dijalankan sistem.

1.6 Sistematika Penulisan

Tulisan ini disusun dalam sebuah laporan dengan sistematika penulisan yang terdiri dari lima bab, sebagai berikut:

Bab 1 yang merupakan pendahuluan tersusun atas latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, metode dan sistematika penulisan Skripsi.

Tinjauan pustaka, dan landasan teori akan dibahas pada Bab 2. Bab ini menguraikan teori maupun hasil dari penelitian yang telah dilakukan sebelumnya untuk dipergunakan sebagai referensi dalam perancangan sistem. Landasan teori diperoleh dari berbagai sumber, berisi konsep untuk pemecahan masalah.

Perancangan Sistem yang akan dibahas pada Bab 3 mencakup analisis teori-teori yang digunakan, materi dan data yang akan dikumpulkan, serta berisi perincian rancangan aplikasi program yang akan dibuat.

Berdasarkan perancangan sistem yang telah dibahas pada Bab 3, Bab 4 memuat hasil implementasi dan analisis dari penelitian mengenai tiap proses yang ada, beserta contoh-contoh hasil program yang telah diimplementasikan.

Bagian Terakhir atau Bab 5 berisi kesimpulan dari hasil analisis implementasi dan penyusunan skripsi dan saran untuk kegiatan pengembangan penelitian di masa mendatang.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dalam penelitian ini, *Nearest Neighbor Clustering* dengan pendekatan *Pivoting* pada *Datawarehouse* Transaksi berhasil diterapkan dan dikembangkan menjadi sebuah program. Berdasarakan hasil pengujian dan analisis yang telah dilakukan oleh penulis, dapat disimpulkan:

1. Berdasarkan hasil pengujian ditemukan bahwa Kombinasi Metode yang menghasilkan rata-rata nilai *purity* paling baik adalah Metode IV (*Min Max Normalization* dan *Euclidean Distance*) dan pada urutan berikutnya Metode II (*Altman ZScore & Euclidean Distance*), Metode I (*Altman ZScore & Cosine Distance*), serta Metode III (*Min Max Normalization & Cosine Distance*) dengan nilai *purity* paling rendah.
2. Berdasarkan hasil analisis *purity* per-klaster, kombinasi *threshold* dan metode yang paling optimal belum dapat menghasilkan hasil klasterisasi yang optimal dengan Data Uji yang digunakan. Hasil klaster yang memiliki data dominan >= 75% dari keseluruhan hasil klaster hanya dapat mencapai nilai paling baik 71.43%, dan dapat turun hingga mencapai angka 50% pada nilai paling rendah.
3. Data Transaksi Bon dan Data Transaksi Tunai yang dipergunakan sebagai Data Uji pada penelititan ini merupakan data yang tidak baik. Ini terlihat dari tidak adanya karakteristik khusus yang menggambarkan kedekatan antar nota satu dengan lainnya sesuai dengan *feature* yang ditentukan. Hal ini menyebabkan sistem tidak dapat dipergunakan untuk memprediksi perilaku pelanggan untuk membantu perusahaan.

5.2 Saran

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

1. Data transaksi yang diteliti memiliki rentang waktu yang lebih lama sehingga bisa mendapatkan hasil pengujian metode, *threshold*, dan analisis yang lebih akurat.
2. Akan lebih baik apabila penelitian dilakukan dengan data sampling dari perusahaan yang bersangkutan ataupun *expert* sehingga hasil pengujian akurasi klaster dapat lebih presisi dan sesuai dengan kasus riilnya.
3. Arsitektur *backend* untuk *query* data pada ekosistem *Hadoop* diubah menggunakan aplikasi yang didesain untuk pengolahan data *realtime* misalnya seperti penggunaan *Apache Spark*.
4. Sistem dapat menganalisis dan menampilkan efisiensi waktu yang dibutuhkan oleh setiap algoritma untuk mendapatkan algoritma yang paling efisien.

DAFTAR PUSTAKA

Ballard, C., Farrell, D. M., Gupta, A., Mazuela, C., & Vohnik, S. (2006). *Dimensional Modeling: In a Business Intelligence Environment.* International Business Machines Corporation.

Dhillon I.S., Modha D.S., & Spangler W.S. (2001). *Visualizing Class Structure of Multidimensional Data.* San Jose: IBM Almaden Research Center.

Gartner. (2011, 6 27). *Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data*. Retrieved from Gartner: http://www.gartner.com/newsroom/id/1731916.

Hermawan, Yudi. (2005). *Konsep OLAP dan Aplikasinya Menggunakan Delphi.* Yogyakarta: Andi Offset.

Imbar, R.V., Adelia, Ayub M., & Rehatta A. (2014). *Implementasi Cosine Similarity dan Algoritma Smith-Waterman Untuk Mendeteksi Kemiripan Teks.* Bandung: Universitas Kristen Maranatha.

Kusrini & Taufiq Luthfi, E. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Offset.

Manning, C., Raghavan, P., & Schütze, H. (2009). *An Introduction to Information Retrieval.* Cambridge: Cambridge University Press.

McKinsey Global Institute. (2011). *Big Data: The Next Frontier for Innovation, Competition, and Productivity.* McKinsey Global Institute.

Ponniah & Paulraj. (2001). *Data Warehouse Fundamentals: a Comprehensive Guide for IT Professional*. New York : John Wiley & Sons.

Patro, S. G. K., Kumar K., (2015). *Normalization: A Processing Stage.* India: Department of CSE & IT, VSSUT, Burla.

Ramadhana, Cakra & Lulu W, Yohana Dewi & Diah K. W., Kartina. (2013). Data Mining dengan Algoritma Fuzzy C-Means Clustering Dalam Kasus Penjualan di PT Sepatu Bata. *Semantik 2013*, 54-60.

Sasirekha, K., Baby, P. (2013). Agglomerative Hierarchical Clustering Algorithm-A Review. *International Journal of Scientific and Research Publications*, 1-3.

Turban, E., Aronson, J. E., & Liang, T. (2007). *Decision Support Systems* (7th Editio). New Jersey: Prentice-Hal.

Wanajaya, A., & Sabloak, R. (2011). *Perancangan Data Warehouse dan Penerapan Data Mining pada Data Penjualan PD XYZ* (Undergraduate Thesis, Sekolah Tinggi Manajemen Informatika Multi Data Palembang) Retrieved from eprints.mdp.ac.id: http://eprints.mdp.ac.id.

Wardoyo, H., Pragantha, J. & Christanti, Viny M. (2013). Penentuan Kelas dengan Nearest Neighbor Clustering dan Penggunaan Metode Naïve Bayes untuk Klasifikasi Dokumen. *Jurnal Ilmu Komputer dan Sistem Informasi*, 79-85.