

CLUSTERING DAN VISUALISASI DATA MAGANG MAHASISWA POLITEKNIK NEGERI BATAM

Mir'atul Khusna Mufida¹, Cytra Sari Rewati Saragi², Florensa Evelina Siahaan³

^{1,2,3}Jurusan Teknik Informatika, Program Studi Teknik Informatika, Politeknik Negeri Batam
¹vda@polibatam.ac.id, ²florensaevelinas@gmail.com, ³cytrasari123@gmail.com

Abstrak

Dalam pelaksanaan kegiatan magang di Politeknik Negeri Batam masih ditemukan beberapa kendala, diantaranya mahasiswa selama ini mengalami kesulitan dalam menentukan posisi magang yang sesuai dengan minat dan kemampuannya. Hal ini dilatar belakangi oleh kurangnya informasi jenis pekerjaan yang dibutuhkan perusahaan atau instansi tujuan magang. Sebenarnya sudah ada data magang yang bisa dianalisis, akan tetapi data tersebut masih dalam bentuk tekstual sehingga sulit untuk dieksplor. Penelitian ini bertujuan untuk mengelompokkan tempat magang mahasiswa terdahulu sesuai dengan jenis pekerjaannya dengan menerapkan algoritma K-means sehingga menghasilkan suatu pola yang dapat mengidentifikasi mahasiswa yang berpotensi bermasalah. Selain metode k-means data magang yang tersedia juga diolah menggunakan metode visualisasi interaktif dengan aspek *overview, zoom and filter dan detail on demand*. Hasil visualisasi data ini mengilustrasikan jenis pekerjaan yang ditawarkan di instansi atau perusahaan tempat magang. Manfaat dari adanya aplikasi ini adalah mempermudah pihak Politeknik Negeri Batam dalam menentukan posisi magang mahasiswa dan mempermudah mahasiswa untuk mengetahui informasi jenis pekerjaan yang dibutuhkan oleh pihak perusahaan atau instansi. Manfaat dari sistem yang dibangun adalah kemampuan memetakan jenis pekerjaan di instansi atau perusahaan dan merekap persebaran mahasiswa magang. Sehingga dapat diketahui sebaran mahasiswa magang di perusahaan atau instansi serta jenis pekerjaan yang dibutuhkan serta catatan permasalahan yang dihadapi mahasiswa selama magang.

Kata kunci : pengelompokan, magang, visualisasi, clustering, k-means

1. Pendahuluan

Politeknik Negeri Batam memiliki suatu mata kuliah wajib yang diikuti oleh seluruh mahasiswa tingkat akhir yaitu magang. Magang merupakan suatu kegiatan pembelajaran dibidang akademis yang bersifat praktek kerja lapangan untuk menunjang wawasan mahasiswa dalam mengenal dunia kerja sesuai dengan pengetahuan yang dipelajarinya selama kuliah. Melalui proses magang tersebut, mahasiswa dapat mengimplementasikan ilmu yang diketuninya dalam bentuk kerja praktik di lapangan agar nantinya dapat menjadi sumber daya manusia yang bermutu dan mampu bersaing.

Kegiatan magang ini juga membawa keuntungan baik untuk pihak perusahaan atau instansi maupun civitas Politeknik Negeri Batam. Perusahaan mendapat tenaga bantuan dengan adanya mahasiswa magang. Politeknik Negeri Batam juga bisa menyalurkan alumni maupun calon lulusannya ke pihak perusahaan atau instansi terkait. Namun, dalam pelaksanaannya masih ditemukan beberapa kendala, diantaranya mahasiswa jurusan Teknik Informatika selama ini mengalami kesulitan dalam menentukan posisi magang yang sesuai dengan minat dan kemampuannya dan kurangnya informasi

jenis pekerjaan yang dibutuhkan dari pihak perusahaan atau instansi tujuan magang.

Faktor pendukung yang juga sangat perlu diperhatikan yaitu kesesuaian jenis pekerjaan dengan keahlian yang dimiliki mahasiswa. Berdasarkan hasil pengamatan di Politeknik Negeri Batam terutama jurusan Teknik Informatika, ketimpangan dalam menyesuaikan keahlian yang dimiliki dalam dunia kerja sering terjadi. Mahasiswa sering mendapatkan tempat magang di perusahaan maupun instansi yang memiliki lowongan magang namun tidak sesuai dengan keahlian yang dimiliki. Hal ini menyebabkan kegiatan magang mahasiswa tidak maksimal dan tidak sesuai dengan kompetensi yang dimiliki oleh mahasiswa.

Penyebab dari masalah tersebut adalah keterbatasan informasi mengenai distribusi magang, yang dapat membantu mahasiswa dalam memilih tempat magang yang sesuai dengan kompetensinya. Distribusi magang merupakan pemetaan jenis-jenis kegiatan mahasiswa magang yang tersebar dalam suatu instansi atau perusahaan yang akan dijadikan tempat magang oleh mahasiswa. Penyebab lainnya adalah tidak adanya rekap data jenis pekerjaan mahasiswa magang jurusan teknik informatika di Politeknik Negeri Batam. Padahal setiap mahasiswa

yang telah selesai magang wajib menyusun laporan magang. Laporan magang ini dibuat salah satunya untuk mengetahui jenis pekerjaan apa yang dilakukan mahasiswa di suatu perusahaan atau instansi, namun masalahnya belum ada pemanfaatan informasi jenis pekerjaan yang dilakukan mahasiswa magang. Sehingga, mahasiswa tidak mengalami kesulitan lagi untuk mengetahui informasi mengenai distribusi tempat magang dan jenis pekerjaan dari setiap tempat magang.

Data magang saat ini tersedia dalam bentuk arsip tekstual berupa dokumen Excel sehingga untuk mengeksplorasi dan mempelajarinya dengan lebih cermat, efisien dan efektif dibutuhkan sebuah metode analisis data salah satunya adalah visualisasi. Visualisasi merupakan sebuah metode untuk menampilkan informasi dalam bentuk visual atau gambar yang interaktif. Sehingga kedepannya, dengan metode ini mahasiswa dan dosen dapat lebih mudah lagi memahami informasi dari data yang ditampilkan.

Dari permasalahan tersebut, maka diperlukan analisis data berupa pengelompokan terhadap data yang ada. Data mining adalah metode analisis data yang berjumlah besar atau kompleks dengan tujuan menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya (Pramudiono, 2006). Metode data mining digunakan untuk menemukan pola yang menarik agar mahasiswa bisa menentukan posisi magang yang sesuai dengan kemampuannya. Pencarian tempat magang yang sesuai dilakukan berdasarkan data *job description* yang diisi oleh pihak perusahaan atau instansi. Variabel yang dipilih untuk melakukan pengelompokan mahasiswa adalah: *job description*, nilai akhir matakuliah magang, etika profesi dan pelaporan kerja.

Data tersebut diproses dengan menggunakan metode pengelompokan *clustering* K-Means. Metode K-Means digunakan untuk menyesuaikan pengelompokan mahasiswa ke beberapa *cluster* berdasarkan data tersebut, sehingga data yang memiliki kriteria yang sama dikelompokkan dalam satu *cluster* dan data yang memiliki kriteria berbeda dikelompokkan dalam *cluster* yang lain. Dengan adanya pengelompokan data, diharapkan dapat memberikan informasi kepada mahasiswa yang akan menentukan posisi magang dan mengidentifikasi mahasiswa magang yang berpotensi bermasalah, ditinjau dari ketidaksesuaian pemilihan tempat magang.

Sesuai dengan pemaparan permasalahan tentang eksplorasi distribusi magang, maka dibuatlah “Visualisasi Distribusi Data Magang Jurusan Teknik Informatika Politeknik Negeri Batam” untuk membantu memetakan jenis pekerjaan dalam instansi atau perusahaan yang ditawarkan sebagai lowongan magang, merekap data distribusi mahasiswa magang, sehingga dapat diketahui

seberapa banyak mahasiswa magang di perusahaan atau instansi. Diharapkan dengan adanya sistem ini mahasiswa tidak mengalami kesulitan lagi dalam menentukan tempat magang sesuai keahliannya. Sehingga pihak yang berhubungan dengan pelaksana magang seperti koordinator magang, dapat memantau distribusi magang mahasiswa dan jenis pekerjaan apa yang banyak ditawarkan sebagai tugas mahasiswa ketika magang.

2. Landasan Teori

2.1 Magang

Magang merupakan kegiatan wajib yang diikuti oleh mahasiswa tingkat akhir dengan tujuan untuk memberikan gambaran tentang kondisi lingkungan kerja. Disamping itu pelaksanaan magang juga berguna untuk menerapkan ilmu yang telah dipelajari sambil menyesuaikan diri di instansi atau perusahaan, sehingga mampu menyerap berbagai pengalaman kerja yang sesungguhnya. Magang diartikan sebagai pelatihan kerja dengan bekerja langsung dibawah bimbingan dan pengawasan pekerja yang lebih berpengalaman di perusahaan untuk dapat menguasai keahlian atau keterampilan tertentu. Tujuan magang adalah sebagai berikut :

- a. Melatih sumber daya manusia menjadi tenaga kerja profesional di masa
- b. yang akan datang.
- c. Memberikan pengalaman nyata dan permasalahan yang dihadapi di dunia kerja untuk menumbuhkan rasa tanggungjawab.

Membantu industri dalam memenuhi sumber daya manusia sebagai kebutuhan tenaga kerja

2.2 Visualisasi

Visualisasi adalah sebuah produk dari kreasi, interpretasi dan refleksi gambar baik diatas kertas maupun dengan menggunakan teknologi, dengan tujuan menyampaikan informasi dan menggambarkannya untuk mempermudah pemahaman. Definisi lain mengenai visualisasi adalah metode untuk mengkonversi data kedalam format visual atau tabel sehingga karakteristik dari data dan relasi antar item data atau atribut dapat dianalisis serta dapat menangkap konsep data dari data yang ditampilkan, karena pada dasarnya manusia lebih mudah mengartikan sebuah gambar daripada teks. Secara umum, visualisasi ini dapat diterapkan di bidang sistem informasi dan aplikasi dengan koneksi database, sehingga data dapat diproses secara *real time* dan dinamis dengan tujuan memperluas pemanfaatan data. Pettersen, Eric F., et al, (2004).

Visualisasi Interaktif the seeking mantra :Shneiderman, Ben (1996)

- 1) Overview

Overview memberikan konteks umum secara keseluruhan untuk memahami kumpulan data, memberi gambaran dari seluruh entitas data yang mewakili visualisasi informasi. Komponen utama dan hubungan satu sama lain yang dibuat jelas. Bentuk keseluruhan data itu sendiri dapat memberikan bantuan dalam memahami informasi yang terkandung. Fitur yang signifikan dapat dilihat dan dipilih untuk pemeriksaan lebih lanjut.

2) Zoom and Filter

Zooming adalah salah satu teknik interaksi dasar visualisasi informasi. Dengan teknik ini dapat mengurangi kompleksitas dari data sehingga informasi asing dapat dihapus dan memungkinkan untuk pengelompokan lebih lanjut. Zooming terbagi dua, yaitu *zooming in* dan *zooming out*. Dengan *zooming in* maka dapat menghilangkan informasi asing dari tampilan, jadi hal ini memungkinkan pengolahan pusat yang lebih tinggi untuk mengatur informasi ataupun melihat isi informasi secara detail. Sedangkan dengan *zooming out* dapat mengungkapkan informasi yang tersembunyi, jadi hal ini memungkinkan pengguna untuk menemukan kembali lokasi data dalam ruang informasi.

Filter yaitu menyaring item yang tidak menarik. Penerapan kueri dinamis pada item adalah salah satu ide kunci dalam visualisasi informasi. Dengan memungkinkan pengguna untuk mengontrol konten pada layar, pengguna dapat dengan cepat fokus pada minat mereka dengan menghilangkan item yang tidak diinginkan.

3) Details on demand

Teknik interaktif yang memungkinkan memilih bagian dari data yang akan divisualisasikan, secara lebih rinci sambil memberikan seluruh gambaran konsep informasi dan data bersifat jelas sehingga bisa dimengerti oleh user yang melakukan teknik visualisasi tersebut. Teknik rincian-*on-demand* menyediakan informasi tambahan ini secara *point-by-point*, tanpa memerlukan perubahan tampilan.

4) View Relationship

Hal yang dapat memungkinkan pengguna untuk dapat melihat hubungan antar item data. Dalam membuat data menjadi saling terhubung salah satunya adalah menemukan kesaamaan antar item data. Item data tersebut dapat mengungkapkan item tertentu dengan perubahan item yang tersedia.

5) History

Proses mengeksplorasi data oleh pengguna yang bersifat umum agar dapat membandingkan keadaan saat representasi ke keadaan sebelumnya, dan dapat menghasilkan pemahaman yang lebih baik dari data itu sendiri. Sebuah antarmuka yang optimal memungkinkan

kegiatan ini mudah diakses dari perintah yang dikeluarkan atau mengembalikan tampilan ke keadaan sebelumnya.

6) Extract

Proses menggali data lebih dalam yang bersifat spesifik terhadap penemuan penting agar digunakan pada sistem lainnya. Tujuan dari proses ini untuk mencegah kebutuhan mengulang data di proyek yang akan datang. Dengan kata lain, sebagai sarana menabung kerja karena proses ini menangani operasi yang panjang dan kompleks.

2.3 Data Mining, Clustering dan K-Means

Data mining adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya (Pramudiono, 2006). Data mining merupakan gabungan dari berbagai bidang ilmu, antara lain basis data, *information retrieval*, statistika, *machine learning* dan sebagainya. Data mining dapat diterapkan diberbagai bidang, seperti bisnis, kesehatan, asuransi, pemasaran dan perbankan. Data mining merupakan bagian dari proses *Knowledge Discovery in Database (KDD)* untuk menemukan informasi dan pola yang berguna dalam data.

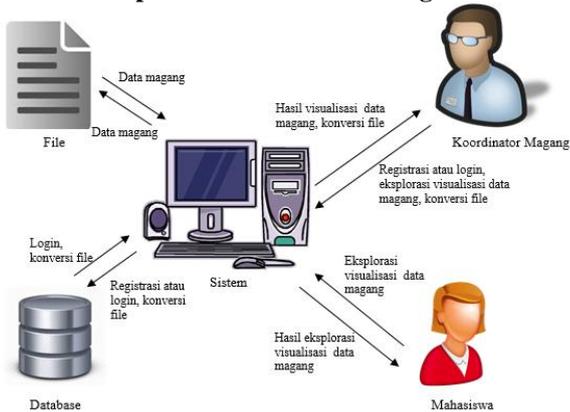
Proses pengelompokan sekumpulan obyek kedalam kelas-kelas obyek yang sama disebut clustering atau pengelompokan. Pengklasteran merupakan satu dari sekian banyak fungsi proses data mining untuk menemukan kelompok atau identifikasi kelompok obyek yang hampir sama. Analisis cluster merupakan usaha untuk mengidentifikasi kelompok obyek yang mirip-mirip dan membantu menemukan pola penyebaran dan pola hubungan dalam sekumpulan data yang besar. Hal penting dalam proses pengklasteran adalah menyatakan sekumpulan pola ke kelompok yang sesuai yang berguna untuk menemukan kesamaan dan perbedaan sehingga dapat menghasilkan kesimpulan.

K-Means merupakan salah satu algoritma clustering. Tujuan algoritma ini yaitu untuk membagi data menjadi beberapa kelompok. Algoritma K-means menerima masukan berupa data tanpa label kelas, komputer mengelompokkan sendiri data-data yang menjadi masukannya tanpa mengetahui terlebih dulu target kelasnya. Algoritma K-means termasuk dalam *unsupervised learning*. Masukan yang diterima adalah data atau objek dan k buah kelompok (cluster) yang diinginkan. Algoritma ini akan mengelompokkan data atau objek ke dalam k buah kelompok tersebut. Pada setiap cluster terdapat titik pusat (*centroid*) yang merepresentasikan cluster tersebut. Ada beberapa kelebihan pada algoritma k-means, yaitu :a) Mudah diimplementasikan dan dijalankan. b) Waktu yang

dibutuhkan untuk menjalankan relatif cepat. c) Mudah untuk diadaptasi. d) Umum digunakan. Algoritma K-means juga memiliki kekurangan yaitu, sebelum algoritma dijalankan k buah titik diinisialisasi secara random sehingga pengelompokan data yang dihasilkan dapat berbeda-beda.

3. Perancangan dan Implementasi

3.1 Deskripsi Umum dan Perancangan Sistem

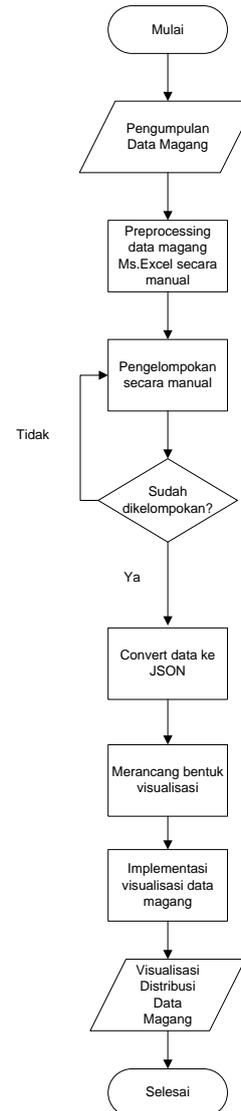


Gambar 1 Arsitektur sistem pengolahan data magang

Aplikasi *clustering* data magang dibangun berbasis web dinamis. Terdapat 2 hak akses terhadap aplikasi yang menggunakan aplikasidengan cara memasukkan data magang dalam format excel. Data di-*impor* kedalam program kemudian ditampung menggunakan database MySQL. Data yang telah dimasukkan kemudian diproses menggunakan algoritma k-means. Proses *clustering* selesai bila jumlah anggota dari tiap kluster tidak berpindah dan berubah lagi atau mencapai global optimum. Sehingga keluaran yang dihasilkan dari aplikasi *clustering* data magang ini adalah berupa data pengelompokan mahasiswa magang yang di representasikan dalam bentuk visualisasi.

Visualisasidistribusi data magang Jurusan Teknik Informatika Politeknik Negeri Batam digunakan untuk memetakan jenis pekerjaan mahasiswa magang yang tersebar pada instansi dan perusahaan tempat pelaksanaan magang mahasiswa. berdasarkan nama perusahaan atau instansi, jenis pekerjaan dan nama mahasiswa. Data magang yang tersedia akan melewati tahap *preprocessing* untuk disimpan dalam bentuk *file* JSON kemudian, sistem dapat menampilkan data perusahaan dalam bentuk visualisasi. Pengguna yang bisa mengakses sistem ini adalah mahasiswa dan koordinator magang. Mahasiswa dapat melihat visualisasi data dan mengeksplor data magang tanpa login sedangkan koordinator magang dapat melakukan registrasi terlebih dahulu jika belum terdaftar, tetapi jika sudah pernah melakukan registrasi sebelumnya maka

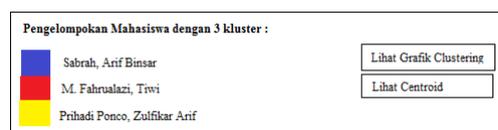
langsung melakukan login, kemudian dapat melakukan eksplorasi data sesuai Gambar 2.



Gambar 2 Algoritma visualisasi data magang

3.2 Implementasi dan Hasil Pengujian

Sebelum mengakses aplikasi pengguna harus login terlebih dahulu.

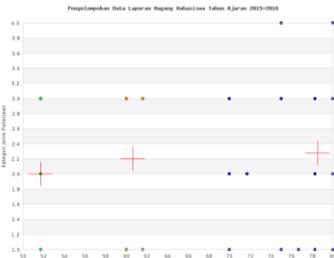


Gambar 3 Perancangan Clustering K-Means



Gambar 4 Implementasi Hasil Clustering

Untuk hasil cluster pada gambar diatas data yang ditampilkan berjumlah 3 cluster. Setiap warna mewakili 1 cluster dan seterusnya. Pengelola dapat menekan tombol “Lihat Grafik Clustering” untuk melihat representasi data tulisan dalam bentuk grafik tombol “Lihat Centroid” berguna untuk menampilkan titik centroid algoritma k-means dari data yang ada. Selanjutnya data di visualisasikan sesuai Gambar 5



Gambar 5 Implementasi visualisasi cluster

Gambar 5 merupakan representasi dari hasil data magang berupa *job description* dan nilai magang mahasiswa.



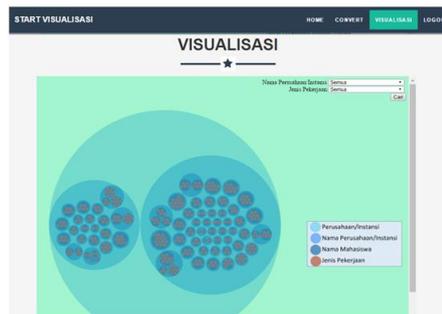
Gambar 6 Implementasi Visualisasi

Sesuai Gambar 6 Terdapat 3 gambar yang masing-masing jika ditekan akan menjelaskan variabel interaktif yang digunakan sistem.



Gambar7 Implementasi Menu Convert

Pada menu *convert* ini menjelaskan proses memasukkan data baru berupa file excel. Pada menu *convert* ini memberikan fasilitas untuk mengunduh format file yang telah disediakan terlebih dahulu. Setelah file diunduh maka dapat dilakukan pengisian data baru dalam file tersebut, kemudian *upload* file dengan menekan tombol *choose file* kemudian ketika data yang diinginkan telah dimasukkan maka tekan tombol submit agar data masuk ke *database* dan dikonversi ke dalam format JSON.



Gambar 8 Implementasi Menu Visualisasi

Halaman visualisasi ini merupakan halaman eksplorasi data. Data yang dimasukkan dalam format JSON, ditampilkan dan diterjemahkan dalam bentuk visualisasi. Dalam visualisasi ini juga disediakan fitur filter untuk membantu pengguna mengeksplorasi data lebih cepat. Jika pengguna memilih salah satu filter atau keduanya lalu menekan tombol cari, maka data akan berubah sesuai dengan hasil filter yang diminta.

4. Kesimpulan dan Saran

Kesimpulan dan saran penelitian sesuai hasil implementasi dan pengujian aplikasi adalah:

1. Aplikasi dapat menerapkan metode *clustering-k-means* untuk mengelompokkan mahasiswa berdasarkan nilai magang dan *job description*
2. Berdasarkan data yang diolah, didapatkan 3 kelompok dengan hasil sebagai berikut :
 - a. Cluster 1 merupakan kelompok mahasiswa magang yang berpotensi berprestasi dengan jenis pekerjaan berupa pemrograman
 - b. Cluster 2 merupakan kelompok mahasiswa magang yang berpotensi mengalami kesulitan dengan jenis pekerjaan sebagai *IT Support*.
 - c. Cluster 3 merupakan kelompok mahasiswa magang yang berpotensi bermasalah dengan jenis pekerjaan berhubungan dengan *survey* pemetaan.
3. Sistem berhasil memvisualisasikan distribusi data magang jurusan Teknik Informatika Politeknik Negeri Batam.

4. Sistem berhasil menerapkan metode visualisasi interaktif yaitu *overview, zoom and filter* dan *detail on demand*.

Daftar Pustaka:

- Asroni dan Adrian, Ronald, 2015, "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang", *Jurnal Ilmiah Semesta Teknik* Vol.18, No. 1, 76-82, Mei 2015.
- Bayu, Muhammad, "Laporan Produktif", 2013
- Jogiyanto, HM, 2005, "Analisis dan Desain Sistem Informasi : Pendekatan Terstruktur Teori dan Praktik Aplikasi Bisnis", Penerbit Andi, Yogyakarta.
- B. Craft, and P. Cairns, 2005, *Beyond guidelines: what can we learn from the visual information seeking mantra?*, Ninth International Conference on Information Visualisation (IV'05). IEEE.
- B. Shneiderman, 2008, *Extreme visualization: squeezing a billion records into a million pixels*, Proceedings of the 2008 ACM SIGMOD international conference on Management of data. ACM.
- D.A. Keim, F. Mansmann, D. Oelke, and H. Ziegler, 2008, *Visual analytics: Combining automated discovery with interactive visualizations*, International Conference on Discovery Science. Springer Berlin Heidelberg.
- Junaedi, F., 2005, *Panduan Lengkap Pemrograman HTML*, Yogyakarta: Percetakan Negeri. November 2016.
- Hardjono D. (Ed), 2006, *Seri Panduan Lengkap Menguasai Pemrograman Web dengan PHP 5*, ANDI Yogyakarta
- Kusrini dan Luthfi, Emha Taufiq, 2009 "Algoritma Data Mining". Yogyakarta. C.V Andi
- Lee S, Finn dan Santana, Juan, 2010 "Data Mining". I. Jakarta. PT. Gramedia
- Luthfi, Emha Taufiq and Kusrini, 2009, "Algoritma Data Mining". I. Yogyakarta. C.V Andi
- M. Bostocks., 2013, *D3 Zoomable Circle Packing Visualization*, Website: jeromefroe.github.io/circlepacker, diakses tanggal: 4 November 2016.
- M. Drapeu., 2014, *CSVJSON*, Website: www.csvjson.com, diakses tanggal: 4
- Nugroho, B., 2008, *Aplikasi Pemrograman Web Dinamis Dengan PHP dan MySQL*, Gava Media, Yogyakarta.
- Ong, Johan Oscar, 2013, "Implementasi K-Means Clustering Untuk Menentukan Strategi Marketing President University" *Jurnal Ilmiah Teknik Industri*, Vol.12, No.1, Juni 2013.
- Pettersen, Eric F., et al., 2004, "UCSF Chimera—avisualization system for exploratory research and analysis." *Journal of computational chemistry* 25.13 :1605-1612.
- Rachmansyah, dkk, "Penerapan Algoritma K-Means Untuk Clustering Dokumen E-jurnal STMIGI MDP".
- S. Mulyana, E. Winarko, 2009, *Teknik Visualisasi dalam Data Mining*, Seminar Nasional Informatika UPN "Veteran" Yogyakarta, 1979-2328
- Shneiderman, Ben. "The eyes have it: A task by data type taxonomy for information visualizations." *Visual Languages*, 1996. Proceedings., IEEE Symposium on. IEEE, 1996.
- Simanjuntak, Elvis, "Sistem Informasi Pemasaran Berbasis Website Pada PT. Trimitra Sebagai Sarana Penjualan Produk Secara Online", 2014
- Syafii, M., 2004, *Membangun Aplikasi Berbasis PHP dan MySQL*, ANDI: Yogyakarta
- Zikry, A.F., 2015, *Visualisasi Interaktif Pengelompokan dan Persebaran Perusahaan*, Politeknik Negeri Batam, Batam