

Analisis Perilaku Pengguna E-Learning BESMART Melalui Teknik *Clustering* dengan Algoritma K-Means

Neni Miswaningsih, Nur Insani

FMIPA, Universitas Negeri Yogyakarta

nenimiswaningsih@yahoo.co.id

Abstrak—Makalah ini membahas tentang analisis perilaku pengguna e-learning BESMART UNY khususnya oleh mahasiswa dengan teknik *clustering* menggunakan algoritma K-Means. Tujuan dari penelitian diharapkan dengan menggali data dari riwayat log pengguna BESMART, diperoleh informasi yang menarik dan bermanfaat bagi kemajuan BESMART dan hasil akademik mahasiswa. Hasil yang diperoleh yaitu mata kuliah yang paling banyak diakses oleh pengguna BESMART adalah mata kuliah dari Fakultas Teknik, Jurusan Pendidikan Teknik Elektronika yaitu Media Digital, sedangkan aktivitas atau modul yang paling banyak dilakukan adalah *course*. Pada umumnya jumlah pengguna yang mengakses BESMART didalam jam perkuliahan lebih banyak daripada diluar perkuliahan, kecuali pada hari libur dimana jumlah waktu akses relatif sama. Untuk mempersiapkan Ujian Tengah Semester, mahasiswa mempunyai kesadaran mengakses BESMART lebih rutin untuk mempersiapkan dan menghadapi ujian mereka.

Kata kunci:, *data mining*, *algoritma K-Means*, *teknik clustering*

I. PENDAHULUAN

E-learning merupakan sistem pembelajaran berbasis teknologi komputer dan komunikasi yang memberikan kondisi untuk mewujudkan prinsip belajar seumur hidup. *E-learning* yang memiliki dasar sistem pembelajar berbasis teknologi elektronik menjadi alat pendukung pembelajaran yang efisien. Hal tersebut dapat dilihat dari penggunaan *e-learning* sendiri, dimana pengguna tidak harus bertemu langsung dengan narasumber apabila memiliki kebutuhan tertentu. Akses tanpa batasan waktu yang diterapkan pada *e-learning* memberikan kemudahan bagi seluruh pengguna untuk dapat memenuhi kebutuhannya sesuai informasi yang tersedia di *e-learning* kapan saja. Melihat kemudahan dan keefektifan dari akses *e-learning* maka telah banyak perguruan tinggi yang menerapkan *e-learning* untuk menunjang perkuliahan. Salah satunya Universitas Negeri Yogyakarta (UNY) yang menerapkan BESMART sebagai *e-learningnya*.

Konten *e-learning* yang menyediakan berbagai fasilitas akan memberikan riwayat *log* dalam jumlah besar dan terus bertambah setiap waktunya. Namun sayangnya riwayat *log* dari perilaku pengguna baik dosen maupun mahasiswa tersebut jarang sekali diolah untuk menjadi suatu informasi yang bermanfaat. Dengan melihat banyaknya data yang terus bertambah disetiap tahunnya, maka amatlah sayang jika data-data tersebut tidak diolah menjadi suatu informasi yang lebih bermanfaat. Pengambilan keputusan yang tepat guna dan sasaran secara otomatis akan meningkatkan kualitas keputusan manajerial sehingga kualitas mutu organisasi/perguruan tinggi-pun tetap terjaga dan semakin baik [1]. Namun penggalian informasi dari data yang sangat besar pun bukan suatu pekerjaan yang mudah. *Knowledge Discovery in Database (KDD)* merupakan suatu teknik mengekstraksi informasi yang potensial, implisit dan tidak dikenal dari suatu kumpulan data. Salah satu step analisis pada KDD yang sangat penting adalah data mining. Data mining merupakan salah satu alat bantu untuk penggalian data pada basis data berukuran besar dan dengan spesifikasi tingkat kerumitan yang telah banyak digunakan pada banyak domain aplikasi seperti perbankan maupun bidang telekomunikasi [2]. Berbagai kemungkinan hubungan dapat terjadi didalam proses *data mining* karena kunci utama dalam *data mining* adalah penemuan sebuah hubungan pada data yang tersedia [3].

Data mining memiliki beberapa teknik pencarian, yaitu pengelompokan (*clusterings*), korelasi (*correlations*), urutan (*sequences*), episode (*episode*), klasifikasi (*classifications*), dan aturan asosiatif (*association rules*). Tujuan dari *clustering* adalah untuk mempartisi sekumpulan data yang diberikan untuk diatur kedalam sebuah kelompok berdasarkan ciri tertentu [4]. Salah satu algoritma yang digunakan dalam *clustering* yang paling populer adalah algoritma *K-Means*.

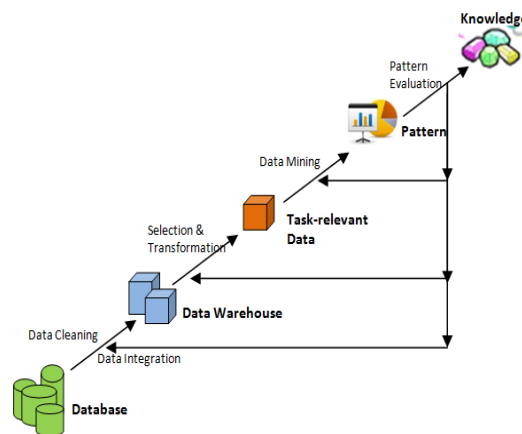
Penelitian ini bertujuan untuk menganalisa perilaku pengguna BESMART UNY dengan teknik *clustering* menggunakan algoritma K-Means. Hasil dari penelitian ini diharapkan memberikan informasi yang bermanfaat yang selama ini mungkin belum pernah tergalinya. Berdasar kecenderungan perilaku pengguna nantinya dapat dijadikan dorongan kepada mahasiswa dan dosen untuk dapat

memanfaatkan dan mengelola meningkatkan efektifitas penggunaan BESMART lebih efisien lagi, sedemikian sehingga dapat pula meningkatkan prestasi akademik mahasiswa. Untuk pihak management, diharapkan dari hasil penelitian ini dapat memberikan rekomendasi sekaligus evaluasi dalam perbaikan konten *e-learning*.

II. STUDI LITERATURE

A. Teknik Clustering

Menurut Larose [5], data mining adalah sebuah proses menemukan korelasi baru, pola, dan tren bermakna dengan memilah-milah sejumlah besar data yang tersimpan dalam gudang data, menggunakan teknologi pengenalan pola seperti statistik dan teknik matematika. Sedangkan menurut Han, J., et al [6], data mining adalah proses menemukan pola dan pengetahuan dari data yang berjumlah besar. Sumber data dapat berupa basis data, data warehouse, web, dan tempat penyimpanan informasi lainnya. *Data mining* merupakan salah satu langkah penting untuk menemukan sebuah pengetahuan pada proses *Knowledge Discovery from Data* (KDD). KDD merupakan proses ekstraksi informasi yang berguna dan tidak diketahui sebelumnya dari sebuah kumpulan data [7]. Adapun proses KDD ditunjukkan pada Gambar 1.



GAMBAR 1. PROSES *KNOWLEDGE DISCOVERY IN DATABASE*

Proses KDD pada Gambar 1 terdiri dari beberapa langkah iteratif berikut:

1. Pembersihan data (*Data cleaning*). Tahap ini merupakan tahap dimana data *outlier* adalah fase di mana *noise* data dan data yang tidak relevan akan dihapus dari koleksi.
2. Integrasi data (*Data Integration*). Pada tahap ini dilakukan penggabungan data dari berbagai sumber untuk dibentuk penyimpanan data yang koheren.
3. Pemilihan data (*Data selection*). Pemilihan data yang relevan untuk dimasukkan ke data warehouse dipilih dari kumpulan data.
4. Transformasi data (*Data transformation*). Pada tahap ini, data warehouse ditransformasi ke dalam bentuk yang siap dan sesuai untuk prosedur penambangan
5. Penambangan data (*Data mining*). Tahap ini merupakan langkah penting di mana teknik yang sesuai diterapkan untuk mengekstrak pola yang berguna.
6. Evaluasi pola (*Pattern evaluation*). Pada langkah ini, pola yang mewakili pengetahuan diidentifikasi berdasarkan langkah-langkah yang diberikan.
7. Representasi pengetahuan (*Knowledge representation*). Pada tahap ini penemuan pengetahuan direpresentasikan secara visual kepada pengguna untuk membantu dalam memahami hasil data mining juga disertakan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

Terdapat tiga teknik yang populer didalam *data mining*, yaitu: (1) Penambangan dengan Aturan Asosiatif (*Association Rule*), (2) Klasifikasi (*Classification*), dan (3) Pengelompokan (*Clustering*).

Clustering adalah proses pengelompokan satu set objek data ke dalam beberapa kelompok atau *cluster* sehingga objek dalam sebuah *cluster* memiliki jumlah kemiripan yang tinggi, tetapi sangat berbeda dengan objek di *cluster* lain. Prinsip dari *clustering* adalah memaksimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar *cluster*. *Clustering* dapat dilakukan pada data yang memiliki beberapa atribut yang dipetakan sebagai ruang multidimensi. Ilustrasi dari *clustering* dapat dilihat di Gambar 2.2 dimana lokasi, dinyatakan dengan bidang dua dimensi, dari pelanggan suatu toko dapat dikelompokkan menjadi beberapa *cluster* dengan pusat *cluster* ditunjukkan oleh tanda positif (+). Beberapa algoritma pada teknik *clustering* memerlukan fungsi perhitungan jarak untuk mengukur kemiripan antar data, diperlukan juga metode untuk normalisasi beberapa atribut yang dimiliki data [8].

B. Algoritma K-Means

Algoritma *K-Means* adalah algoritma klasik untuk menyelesaikan masalah *clustering*, algoritma ini relatif sederhana dan cepat [9]. Algoritma *K-Means* lebih sering dikenal karena kecerdasannya dalam mengelompokkan data dalam jumlah besar dengan cepat dan efisien. *K-Means* merupakan algoritma untuk mengklasifikasi atau mengelompokkan objek berdasarkan atribut/fitur ke dalam k kelompok. k adalah bilangan positif. Pengelompokan ini dilakukan dengan meminimalkan jumlah kuadrat dari jarak antara data dan *cluster centroid* yang bersangkutan. Dengan demikian, tujuan dari *k-means clustering* adalah untuk mengelompokkan data. Efisiensi keaslian algoritma *K-Means* sangat bergantung pada *centroid* awal [4]. Langkah kerja dari algoritma *K-Means* adalah sebagai berikut:

1. Menentukan jumlah *cluster* k dan *centroid* awal
2. Menghitung jarak masing-masing objek ke *centroid*
3. Mengelompokkan objek berdasarkan jarak minimum objek ke *centroid*
4. Menentukan *centroid* kembali apabila terjadi perpindahan kluster dengan menghitung nilai rata-rata dari seluruh anggota pada masing-masing kluster
5. Ulangi langkah 1 sampai 4 sampai tidak terjadi perpindahan kluster lagi.

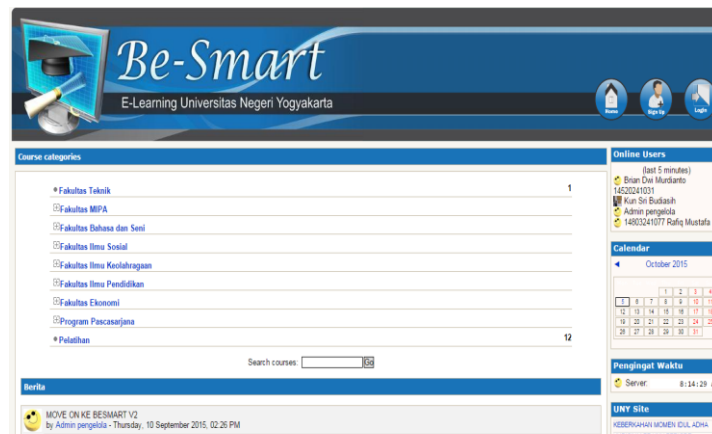
Algoritma berakhir ketika semua *record* yang dimiliki oleh masing-masing pusat *cluster* tetap dalam *cluster* itu. Atau, algoritma dapat berhenti ketika beberapa kriteria konvergensi terpenuhi, seperti ada penyusutan yang tidak signifikan dalam jumlah kuadrat *error* (*sum of squared errors*):

$$SSE = \sum_{i=1}^k \sum_{p \in C_i} d(p, m_i)^2 \quad (1)$$

dimana $p \in C_i$ melambangkan setiap titik dalam *cluster* i dan m_i merupakan pusat massa *cluster* i [5].

C. BESMART UNY

E-learning UNY diimplementasikan dengan paradigma pembelajaran pembelajaran *online* terpadu menggunakan *Learning Management System* (LMS) Moodle. Sistem *e-learning* UNY yang telah berfungsi sejak tahun 2006 kini dikenal dengan nama BESMART. Beberapa kegiatan yang dapat dilakukan di BESMART, yaitu *user*, *course*, *assignment*, *upload*, *resource*, *quiz*, *message*, *survey*, *chat*, *glossary*, *forum*, *vicon*, *scorm*. Berikut diberikan tampilan utama BESMART yang dapat dilihat pada Gambar 2 [10].



GAMBAR 2. TAMPILAN MUKA BESMART UNY

III. METODE PENELITIAN

A. Sumber Data

Sumber data yang diperoleh dari basis data *log history* BESMART UNY pada tanggal 1 April 2015 sampai dengan 30 April 2015 dengan sejumlah 198.948 dataset.

B. Kerangka Kerja Penelitian

Sumber data kemudian diolah melalui beberapa tahap yaitu *preprocessing* yang termasuk *data cleaning*, *data integration*, *data selection*, *data transformation*. Setelah tahap-tahap tersebut terpenuhi baru dilakukan proses *data mining*. Teknik *data mining* yang akan digunakan pada penelitian ini, yaitu teknik *clustering* dengan algoritma *K-Means*. Pada tahap akhir, dari hasil penambangan data, dilakukan proses evaluasi pola perilaku pengguna BESMART untuk menggali informasi yang mungkin akan bermanfaat bagi pengguna dan pengelola.

IV. HASIL PENELITIAN

A. Pre-processing

Langkah awal pada analisis ini yaitu menganalisis data *log* untuk mengetahui semua aktivitas yang terdapat di data riwayat *log* BESMART UNY. Riwayat *log* BESMART sebagai data mentah memiliki 8 atribut, yaitu *userid*, *firstname*, *lastname*, *email*, tanggal, pukul, makul, modul. Pada analisis ini akan diterapkan *data mining* dengan teknik *clustering*. Sumber data diperoleh dari basis data riwayat *log* BESMART UNY pada tanggal 1 April sampai dengan 30 April tahun 2015 sejumlah 198.948 dataset.

Proses pembersihan data (*data cleaning*) dilakukan pada data yang tidak memiliki atribut lengkap, dan data yang tidak konsisten dan data yang tidak lengkap, sehingga diperoleh data siap olah sejumlah 187.857. Data tersebut kemudian melewati proses *data integration*, *data selection*, *data transformation* dan siap menjadi sebuah *data warehouse*.

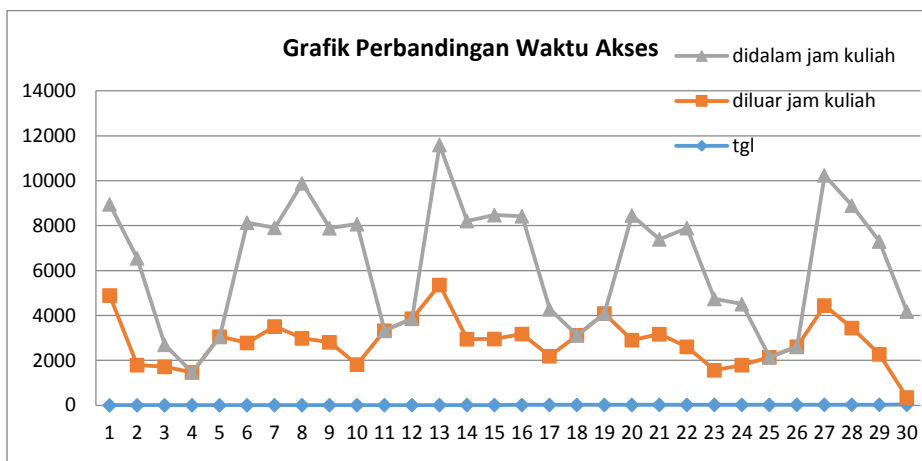
B. Proses Clustering

Setelah didapatkan *data warehouse* sejumlah 187.857 data *log* dan disimpan dalam format *CSV* (*Comma delimited*), selanjutnya dilakukan *clustering* dengan *software* WEKA. Adapun hasil dari proses *clustering* ditunjukkan pada Tabel 1.

TABEL 1. HASIL CLUSTERING UNTUK MATAKULIAH, AKTIVITAS (MODUL), DAN WAKTU AKSES

Matakuliah yang paling banyak diakses	Media Digital (her 2015)
Aktivitas (modul) yang paling sering dilakukan	Course
Waktu akses yang sering dilakukan	Pukul 13-14.00 WIB

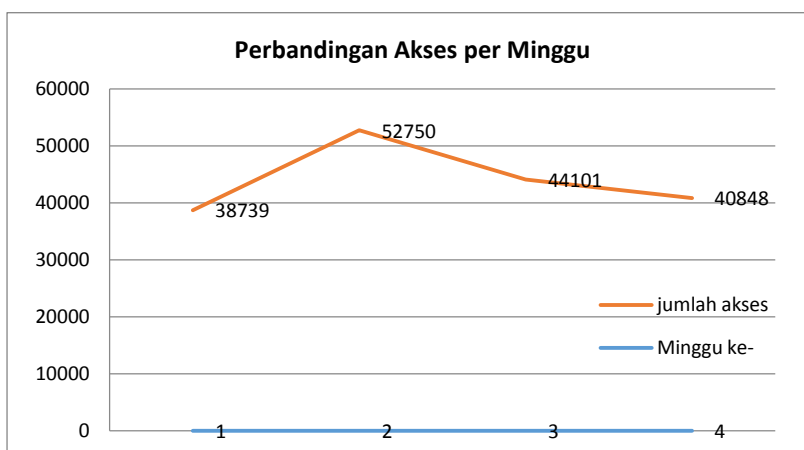
Berdasarkan Tabel 1 dapat diketahui bahwa matakuliah yang paling sering diakses adalah Media Digital (her 2015). Aktivitas (modul) yang paling sering dilakukan pada adalah aktivitas perkuliahan *online* (*course*) sedangkan untuk waktu akses yang paling sering dilakukan terjadi pada pukul 13.00 – 14.00 WIB. Waktu akses diluar jam perkuliahan dan tidak dalam jam perkuliahan akan menjadi objek analisis *clustering* selanjutnya. Pada bulan April 2015 waktu perkuliahan yang masih berlaku di UNY yaitu pukul 07.00-17.00 WIB.



GAMBAR 3. PERBANDINGAN WAKTU AKSES DIDALAM DAN DILUAR JAM PERKULIAHAN

Berdasarkan Gambar 3 dapat diketahui bahwa waktu akses terbanyak terjadi pada jam perkuliahan. Pengaksesan terbanyak pada saat jam perkuliahan terjadi pada tanggal 13 April 2015, dimana hari tersebut menepati hari Senin. Faktanya didapatkan informasi bahwa pada satu minggu pertama akses di dalam jam perkuliahan paling banyak terjadi pada hari Rabu sedangkan tingga minggu selanjutnya terjadi pada hari Senin. Pada umumnya jumlah pengguna yang mengakses BESMART didalam jam perkuliahan lebih banyak daripada diluar perkuliahan, kecuali pada hari libur dimana jumlah waktu akses relatif sama. Pada tanggal 4, 5, 11, 12, 18, 19, 25 dan 26 April 2015, jumlah pengguna menempati posisi yang hampir sama. Hal ini dikarenakan pada tanggal-tanggal tersebut, merupakan hari Sabtu dan Minggu. Hal ini dapat diartikan juga walaupun merupakan hari libur, pengguna BESMART masih berminat untuk menggunakan dan mengakses sistem tersebut.

Apabila dilihat dari jumlah akses per minggu, akses pengguna terbanyak terjadi pada minggu kedua, yang ditunjukkan oleh Gambar 4. Lebih khususnya, pada minggu kedua April 2015, jumlah pengguna menempati peringkat pertama. Jika ditilik dari kalender akademik, pelaksanaan Ujian Tengah Semester memang diadakan pada minggu tersebut, sehingga dapat disimpulkan bahwa mahasiswa mempunyai kesadaran untuk mengakses BESMART lebih rutin untuk mempersiapkan dan menghadapi ujian mereka.



GAMBAR 4. JUMLAH AKSES BESMART PER MINGGU

Selanjutnya setelah memperoleh beberapa informasi mengenai waktu akses, peneliti bermaksud untuk menggali informasi mengenai program studi yang paling banyak melakukan aktivitas di BESMART UNY. Dilakukan dilakukan analisis *cluster* pada program studi dan berikut program studi yang menduduki 7 teratas.

TABEL 2. TUJUH PROGRAM STUDI YANG MEMILIKI JUMLAH AKSES TERBANYAK

Program studi	Jumlah akses
Pendidikan Teknik Informatika	51.241
Akuntansi	33.970
Pendidikan Akuntansi	10.841
Matematika	9.880
Pendidikan Administrasi Perkantoran	9.737
Bahasa dan Sastra Inggris	9.334
Teknik Elektronika	5.561

Berdasarkan Tabel 2 diketahui bahwa program studi (prodi) Pendidikan Teknik Informatika memiliki jumlah akses terbanyak. Dari hasil ini, karena prodi Pendidikan Teknik Informatika mempunyai jumlah akses terbanyak, maka dari prodi ini, peneliti bermaksud untuk menganalisis lebih lanjut tentang perilaku pengguna pada prodi tersebut. Tabel 3 menunjukkan jumlah akses terbanyak pada mahasiswa yang berbeda angkatan.

TABEL 3. MATA KULIAH YANG PALING BANYAK DIAKSES BERDASAR TAHUN ANGGKATAN MAHASISWA DI PRODI TEKNIK INFORMATIKA

Tahun Angkatan	Matakuliah yang paling banyak diakses
2010	Algoritma Pemograman ADI
2011	Jaringan Terdistribusi ekoaji 2015
2012	Animasi Sigit Pambudi
2013	Media Digital (her 2015)
2014	Jaringan Komputer

Berdasarkan Tabel 3 diperoleh informasi bahwa berdasarkan tahun angkatan mahasiswa yang berbeda, diperoleh matakuliah yang paling banyak diakses juga berbeda. Berdasarkan informasi tersebut dapat dikemungkinan bahwa untuk mahasiswa dengan tahun lebih lama mengulang matakuliah tersebut untuk memperbaiki nilainya.

V. SIMPULAN DAN SARAN

Berdasarkan pembahasan dalam penelitian ini dapat disimpulkan mata kuliah yang paling banyak diakses oleh pengguna BESMART adalah mata kuliah dari Fakultas Teknik, Jurusan Pendidikan Teknik Elektronika yaitu Media Digital, sedangkan aktivitas atau modul yang paling banyak dilakukan adalah *course*. Pada umumnya jumlah pengguna yang mengakses BESMART didalam jam perkuliahan lebih banyak daripada diluar perkuliahan, kecuali pada hari libur dimana jumlah waktu akses relatif sama. Untuk mempersiapkan Ujian Tengah Semester, mahasiswa mempunyai kesadaran mengakses BESMART lebih rutin untuk mempersiapkan dan menghadapi ujian mereka. Sesuai dengan minat areanya, matakuliah pada prodi Pendidikan Teknik Informatika diakses paling banyak oleh mahasiswa.

VI. UCAPAN TERIMA KASIH

Ucapan terima kasih diberikan kepada Fakultas Matematika dan Ilmu Pengeahuan Alam UNY yang telah memberi kesempatan baik moril dan materil kepada peneliti untuk melakukan penelitian ini.

VII. DAFTAR PUSTAKA

- [1] J. Arief, "Memprediksi Kinerja Mahasiswa Menggunakan Teknik Data Mining (Studi kasus data akademik mahasiswa UNISBANK)", Tesis Tidak Terpublikasi, Yogyakarta: Universitas Gajah Mada, 2010
 - [2] K. Hernawati, dkk, "Application of Association Rules With Apriori Algorithm to Determine The Pattern of The Relationship Between SBMPTN Database and Student's Grade Point Average", Prociding International Seminaron Innovation in Mathematics and Mathematics Education, FMIPA UNY, 2014
 - [3] P. Ponniah, "Datawarehouse Fundamentals: A comprehensive Guide for IT Professional", New York: John Willey & Sons. Inc, 2001
 - [4] Y. Madhu, P. Srinivasa, "Enhancing K-means Clustering Algorithm eith Improved Initial Center, International Journal of Computer Science and Information Technologies, Vol. 1(2). Hlm. 121-125, 2010
 - [5] D. T. Larose, "Discovering Knowledge in Data: An Introduction to Data mining, New Jersey: John Willey & Sons. Inc.", 2005
 - [6] J. Han, M. Kamber, & J. Pei., "Data Mining: Concept and Techniques, Second Edition", Waltham: Morgan Kaufmann Publishers, 2006
 - [7] B. Max, "Principles of Data Mining", London: Springer, 2007
 - [8] Kusnawi, "Pengantar Solusi Data Mining", Seminar Nasional Teknologi 2007 (SNT), Yogyakarta: STMIK AMIKOM Yogyakarta, 2007
 - [9] C. Zhang dan F. Fang, "An Improved K-means Clustering Algorithm", Journal of International & Computational Science, Vol. 10 (1), Hlm. 193-199, 2013
- Puskom UNY, "Praktek E-Learning", Yogyakarta: Universitas Negeri Yogyakarta, 2012