

Volume VI 2011

ISSN 1907-3526



Proceeding

Seminar Nasional

Riset Teknologi Informasi 2011

**"Implementasi Mobile Computing di Dunia Pendidikan dan Industri:
Sebuah Peluang dan Tantangan"**

Yogyakarta, 17 September 2011

Komputasi
Teknologi Web
Keamanan Sistem
Kecerdasan Buatan
Teknologi Basis Data
Pemodelan dan Aplikasi
Pengolahan Citra, Grafika dan Multimedia
Komunikasi Data, Jaringan Komputer dan Sistem Kendali

Ditanggungjawab Oleh:





PERPUSTAKAAN AKAKOM

Tgl: 12 MAY 2015
No. Inv: 088/34/P/AW/1/2015
No. Kelas: BK 262 OM
Kode Bar: PRO 0008 E-12

Proceeding

Seminar Nasional

Riset Teknologi Informasi 2011

**"Implementasi Mobile Computing di Dunia Pendidikan dan Industri:
Sebuah Peluang dan Tantangan"**

Yogyakarta, 17 September 2011

Diselenggarakan Oleh :



YAYASAN PENDIDIKAN WIDYA BAKTI
STMIK
AKAKOM
YOGYAKARTA
Yang Pertama dan Utama



KATA PENGANTAR

Puji syukur marilah kita panjatkan ke hadirat Tuhan YME yang telah melimpahkan rahmat dan hidayahnya, sehingga dapat terselesaikannya penyusunan *Proceeding* SRITI 2011 ini. Buku ini memuat naskah hasil penelitian dari berbagai bidang kajian yang akan dipresentasikan pada Seminar Riset Teknologi Informasi (SRITI) 2011 ke-6 yang telah menjadi agenda tahunan dari Pusat Penelitian dan Pengembangan STMIK AKAKOM Yogyakarta dan sekaligus sebagai rangkaian dari peringatan 32 tahun STMIK AKAKOM.

Call for paper pada seminar SRITI 2011 naskah yang dikirimkan kepada Panitia sudah dalam bentuk *full paper*, sehingga naskah yang masuk ke panitia merupakan naskah final hasil penelitian yang siap dipublikasikan. Naskah yang masuk ke panitia selanjutnya di-review oleh para pakar di bidangnya. Atas kesediaan, kerjasama dan konsistensinya dalam me-review seluruh naskah yang dikirimkan, panitia mengucapkan banyak terima kasih.

Kegiatan SRITI 2011 mengambil tema tentang "*Implementasi Mobile Computing di Dunia Pendidikan dan Industri: Sebuah Peluang dan Tantangan*" direncanakan dapat menyidangkan secara paralel sesuai dengan kelompok kajian ilmu dalam waktu satu hari. Panitia menyadari bahwa, hingga saat ini masih banyak *paper content* yang belum mengacu pada tema, namun mengingat lingkup bidang kajian teknologi informasi yang sangat luas, maka ke depan diharapkan masih dapat ditingkatkan kesesuaian, kedalaman, maupun spektrum kajiannya.

Meskipun kegiatan seminar dan pendokumentasian naskah dalam *proceeding* ini telah dipersiapkan dengan baik, namun kami menyadari masih terdapat banyak kekurangannya. Untuk itu, panitia mohon maaf yang sebesar-besarnya dan terima kasih atas kepercayaan serta kerjasamanya dalam kegiatan ini. Kritik dan saran perbaikan sangat diharapkan untuk penyempurnaan di masa mendatang, yang dapat dikirimkan melalui e-mail sriti@akakom.ac.id.

Kepada semua pihak yang terlibat, baik langsung maupun tidak langsung dalam penyusunan *proceeding* SRITI 2011, panitia mengucapkan terima kasih.

Terima kasih.

Yogyakarta, 17 September 2011

Panitia SRITI 2011
Ketua Pelaksana,

Drs. Tri Prabawa, M.Kom.

DAFTAR ISI

KATA PENGANTAR	iii
DAFTAR ISI	v
Potret dan Potensi Pengembangan Mobile Software di Indonesia <i>Riadi Feruhana (UGM)</i>	ix
A. Komputasi	
Analisis Kinerja Pemecahan Persamaan Diffusi 2-D Menggunakan Modifikasi LU Dekomposisi dalam Komputer Kluster <i>Mike Sasmitanti (BATAN)</i>	1
Analisis Posisi Jarak Optimal Repeater terhadap Akurasi RADAR Transponder Tracking Roket 3 Dimensi <i>Wahyu Widada (LAPAN)</i>	5
Komputasi Ketidakpastian Probabilistik menggunakan Metode LHS dengan Pendekatan Permukaan Respon <i>Entin Hartini (PPIN-Badan Tenaga Nuklir Nasional)</i>	9
Pemodelan Barnsley Fern Menggunakan OpenGL <i>Isram Rasal, Yosfik Alqadri (Universitas Gunadarma)</i>	13
Penerapan Algoritma Greedy Kernel Principal Component Analysis Termodifikasi pada Ekstraksi Fitur Tak Terawasi <i>Victor Hariadi, Rully Soelaiman, Wimbi Perdana Putra (Institut Teknologi Sepuluh Nopember)</i>	21
Perbandingan Kinerja Kompresi File Menggunakan Metode Huffman dengan Adaptive Huffman <i>Sudarmanto (STMIK AKAKOM)</i>	27
B. Kecerdasan Buatan	
Aplikasi Fuzzy Analytical Hierarchy Process Dalam Pemilihan Dosen Teladan <i>Ariesta Damayanti (STMIK AKAKOM)</i>	35
Clustering terhadap Indeks Prestasi Mahasiswa STMIK AKAKOM menggunakan K-Means <i>Sri Redjeki, Andreas Pamungkas, Hastin Al-fatah, (STMIK AKAKOM)</i>	41
Diagnosa Penyakit dengan Gejala Utama Demam Menggunakan Jaringan Syaraf Tiruan Persempitan Balik <i>Sudharmadi Bayu Jati Wibowo, Syamsu Windarti (STMIK AKAKOM)</i>	49
Model Aturan Keterhubungan Data Mahasiswa Menggunakan Algoritma C4.5 untuk Meningkatkan Indeks Prestasi <i>Dedy Hartama (AMIK Tunas Bangsa Pematang Slantar), Muhammad Zarlis (FMIPA USU) Muhammad Saflil (AMIK Tunas Bangsa Pematang Slantar)</i>	57
Penentuan Jumlah Produksi Perusahaan menggunakan Metode Jaringan Saraf Tiruan <i>I Gede Santi Astawa (Universitas Udayana)</i>	65
Pengenalan Jenis Penyakit THT Menggunakan Jaringan Learning Vector Quantization <i>Enny Itje Sela (STMIK AKAKOM), Sri Hartati (Universitas Gadjah Mada)</i>	71

Rancang Bangun Elektrokardiogram (EKG) untuk Analisa Heart Rate Variability pada Domain Waktu <i>Rudi Urwarman, Dwi Ana Ratna Watt, Tito Yuwono (Universitas Islam Indonesia)</i>	77
Sistem Pembantu Pengambilan Keputusan Pemilihan Jurusan di Politeknik Informatika Del <i>Rosni Lumbantoran, Michel Mula Sibarsani, Hendra Sirait (Politeknik Informatika Del)</i>	85
Sistem Pendukung Keputusan dengan Metode Regresi Linear untuk Penentuan Kadar Lemak pada Tubuh Manusia <i>Y. Yohakin Marwanto (STMIK AKAKOM)</i>	97
Sistem Pendukung Keputusan Penentuan Lokasi Toko Swalayan Berbasis Sistem Informasi Geografis <i>Dara Kusumawati (STMIK AKAKOM)</i>	105
Strategi Pemasaran dengan Pendekatan Model Aturan Pohon Keputusan Menggunakan Algoritma ID3 <i>Ahhammad Safi, Dedy Hartama (AMIK Tunas Bangsa Pematang Siantar), Muhammad Zarlis (FMIPA USU)</i>	113

C. Teknologi Basis Data

Aplikasi Pencarian Data pada Database Menggunakan Multi-Kategori pada Tabel Tunggal <i>Thomas Edison Tarigan (STMIK AKAKOM)</i>	121
Pencarian pada Organisasi File Sekuen Berindeks <i>Pulut Suryati (STMIK AKAKOM)</i>	129
Penerapan Indexed View terhadap Performansi Basis Data Studi Kasus: Basis Data Sipketik Pemkab Taput <i>Imelda Rincawaty Simanjuntak, Managam Rajagukguk, R. Bernhard H. Saragih (Politeknik Informatika Del)</i>	137

D. Pemodelan dan Aplikasi

Aplikasi Penjualan Pulsa Elektronik Berbasis Mobile <i>Liliana Mustika Dewi, Indra Yatini B (STMIK AKAKOM)</i>	145
Interpretasi Mobile Learning Sebagai Media Pembelajaran Alternatif Menghadapi Ujian Akhir Nasional <i>Julius Hari, Yonatan Widianto (Universitas Widya Kartika Surabaya)</i>	149
Pembuatan Termometer Digital untuk Mengukur Suhu Ruang Casing Komputer <i>Bambang Eka Purnama (Universitas Surakarta)</i>	155
Pemodelan Distribusi Suhu di dalam Batang Elemen Bakar Reaktor <i>Elfrida Saragi, Ir. Henky P.R MSME (BATAN)</i>	165
Pemodelan Masalah Konduksi Panas dengan Pembangkit Energi pada Hollow Cilinder <i>Khairina Ns, Elfrida Saragih (BATAN)</i>	169
Pengembangan Alat Pembaca Data dari Transponder RFID pada Sistem Pendeteksi Barang Inventaris <i>Farid Thalib, Ali Yanuar, Ridha Iskandar (Universitas Gunadarma)</i>	175
Rancang Bangun Aplikasi Mobile Web Kampus pada Universitas Jenderal Soedirman <i>Laxmedi Afuan (Universitas Jenderal Soedirman)</i>	183
Regular Expression untuk Pencarian Kata Dasar Bahasa Indonesia <i>Yohanes Sayanto, Subanar, Agus Harjoko, Sri Hartati (UGM Yogyakarta)</i>	187

Simulasi Perhitungan Waktu Boarding Pada Pesawat dengan Membandingkan Beberapa Strategi Boarding <i>Bilqis Amaliah, Hanning Tri C., Candra Agus Sukirwan, Muhammad Najib, Virky Wiradharma S (Institut Teknologi Sepuluh Nopember)</i>	193
Sistem Aplikasi Penghitung Zakat Berbasis Sistem Operasi Android <i>Irfan Al Azhari, Tri Prabawa (STMIK AKAKOM)</i>	201
Sistem Deteksi Kerusakan pada Handphone Menggunakan Case Based Reasoning <i>FX Henry Nugroho (STMIK AKAKOM), Abriyono (STMIK WIDYA DHARMA)</i>	207
E. Komunikasi Data, Jaringan Komputer, dan Sistem Kendali	
Aplikasi Kompas HM55B sebagai Pemandu Gerak Robot <i>Supranto (Universitas Negeri Yogyakarta)</i>	215
Otomatisasi Sistem Pendeteksi Kecepatan untuk Membatasi Kecepatan Kendaraan <i>Semai Tjiharjadi, Marvin Chandra Wijaya (Universitas Kristen Maranatha)</i>	223
Pensaklaran Beban Elektronik Berbasis SMS <i>Miftakhu Huda, Wagito</i>	231
Protokol Secure Sealed Bid Auction Menggunakan SMS <i>Ricky Aji Pratama, Didik Utomo (Sekolah Tinggi Sandi Negara)</i>	237
Tapis Digital Chebyshev untuk Noise Canceller pada Suara Manusia <i>Dwi Nuri Putri Dharma, Riwaldi Pudja, Farid Thalib (Universitas Gunadarma)</i>	243
F. Pengolahan Citra, Grafika Komputer dan Multimedia	
Deteksi Abnormalitas pada Citra Mammogram dengan Gray-Level Co-Occurrence Matrix <i>Shofwatul Uyun (Universitas Islam Negeri Sunan Kalijaga), Agus Harjoko (UGM)</i>	249
Deteksi Tepi dengan Algoritma Sobel untuk Uji Berbagai Sumber Peta Bitmap (Vektorisasi Peta Bitmap untuk Sistem Informasi Geografi Dinamis) <i>M. Guntara (STMIK AKAKOM)</i>	255
Multimedia Interaktif Pembelajaran Berbasis Sistem Isyarat Bahasa Indonesia (SIBI) untuk Penyandang Tuna Rungu <i>Ana Heryana, Arif Lukman (LIPI)</i>	261
Restorasi Image dengan Metode Wiener dan Lucy-Richardson <i>Muhammad Kusban (Universitas Muhammadiyah Surakarta)</i>	267
Studi Awal Otomatisasi Deteksi Nuclei pada Citra Pap Smear <i>Izzaty Muhiimah, Rahadlan Kurniawan (Universitas Islam Indonesia)</i>	275
G. Teknologi Web	
Evaluasi Kualitas Website-Pemerintah Daerah dengan Menggunakan Webqual (Studi Kasus pada Kabupaten Ogan Ilir) <i>Candra Irawan (Departemen Inspektorat Pemerintah Kabupaten Ogan Ilir Propinsi Sumatera Selatan), Iswanti (Politeknik Negeri Semarang), Risamuri Hidayat, Sri Suning Kusumawardani (Universitas Gadjah Mada)</i>	281
Modifikasi Arsitektur MVC Codeigniter Menggunakan Web Service SOAP-RPC pada Sistem e-Learning untuk Memperluas Platform Aplikasi Smart Client <i>Wiharto, Wismi Widiarto, Didiek Winoyo (FMIPA UNS Surakarta)</i>	291

Pemanfaatan Teknologi Web3D untuk Pembuatan Situs Web Promosi Perumahan <i>Ahsaid W Hamanto (Universitas Ahmad Dahlan)</i>	297
Pengembangan Sistem Pelayanan Moda Transportasi Darat Berbasis Web <i>Wibridha Bambang Triodi Handaya, Ratih Dwi Purwatingsih (Universitas Kristen Maranatha)</i>	303
Perancangan dan Pembuatan Aplikasi untuk Pencarian Web Service Menggunakan Lucene <i>Sarwozri, Umi Laili Yuhana, Olga Ceria Sari (Institut Teknologi Sepuluh Nopember)</i>	309

H. Keamanan Sistem

Analisis Perbandingan Kualitas Transformasi pada Teknik Obfuskasi <i>Arlinda Christy Baras, Fina Manurung, Harisen Ong, Rony Simamora (Politeknik Informatika Del)</i>	315
Aplikasi Enkripsi SMS Dengan Metode Kombinasi Vigenere Cipher Dan RC4 Berbasis J2ME <i>Sepatiawan Dwi Pria Sulisty, Indra Yatini B (STMIK AKAKOM)</i>	325
Aplikasi Secure Chat Client-Server berbasis OpenVPN dengan Menggunakan Dongle Flashdisk sebagai Otentifikasi Tambahan <i>Innan Rahayu MT, Yoyok Darmanto (Sekolah Tinggi Sandi Negara)</i>	331
Enkripsi E-mail untuk End-to-End User yang Terdaftar pada Pop Mail Server dengan Menggunakan Dongle <i>Nurman Yohan Sopanji, S.ST (Lembaga Sandi Negara)</i>	341
Enkripsi File Video MPEG dengan Teknik Selektif menggunakan Algoritma Serpent <i>Elisa M Sibarani, Krityna Simanjuntak, Tri Artha Siringoringo, Dita Simanjuntak (Politeknik Informatika Del)</i>	347
Metode Pengundian Grup secara Mobile menggunakan Fungsi Hash <i>Ardaya Suryadinata, Eka Rahayu Melaniwulandari (Sekolah Tinggi Sandi Negara)</i>	359
Perancangan dan Implementasi Algoritma Enkripsi Arefour pada Perangkat Kriptografi Berbasis FPGA <i>Mohamad Syahral, Agus Mahardika A. L. (Lembaga Sandi Negara)</i>	365
Sistem Keamanan Data Teks Menggunakan Algoritma RC6 <i>Indra Yatini, F. Wiewiek Nurwiyati (STMIK AKAKOM)</i>	373
Sistem Login Otentifikasi Dua Arah pada Smartphone Menggunakan Gambar Acak <i>Rikson Gultom (Lembaga Sandi Negara)</i>	379

DAFTAR SUSUNAN PANITIA	385
-------------------------------------	-----

Clustering Terhadap Indeks Prestasi Mahasiswa STMIK Akakom Menggunakan K-Means

Sri Redjeki¹⁾, Andreas Pamungkas²⁾, Hastin Al-fatah³⁾

¹⁾²⁾³⁾STMIK AKAKOM YOGYAKARTA
e-mail : dzeky@akakom.ac.id

Abstrak

Clustering k-means merupakan salah satu metode *Data Mining* yang bersifat tanpa arahan (*unsupervised*). *Clustering* merupakan cara memasukkan suatu pola yang diamati ke suatu kelas pola yang belum diketahui dan disebut sebagai kluster pola. Ada dua jenis data *clustering* yang sering dipergunakan dalam proses pengelompokan data yaitu *hierarchical* (hirarki) data *clustering* dan *non-hierarchical* (non hirarki) data *clustering*. Pada penelitian ini menggunakan pendekatan hirarki dengan *partitioning clustering*.

Obyek dari penelitian ini menggunakan data indeks prestasi mahasiswa yang berupa IPS dan IPK. Nilai IPS yang digunakan yaitu semester ganjil dan semester genap pada T.A 2007/2008 dan 2008/2009, sedangkan nilai IPK yang digunakan adalah IPK pada saat semester ganjil 2009/2010. Hasil dari penelitian ini menunjukkan bahwa cluster akhir untuk T.A 2007/2008 mempunyai kecenderungan yang sama. Sedangkan untuk cluster akhir untuk T.A. 2008/2009 mempunyai kecenderungan yang cukup berbeda. Proses *clustering* data tiap semester dilakukan dalam dua kali percobaan agar dapat dibandingkan nilai tengah terbaiknya.

Kata Kunci : *Clustering k-means, nilai tengah, indeks prestasi mahasiswa, STMIK AKAKOM.*

1. Pendahuluan

Sebuah perguruan tinggi yang baik sangat dipengaruhi oleh prestasi mahasiswa yang ada didalamnya. Mahasiswa merupakan obyek dari proses pembelajaran yang ada pada perguruan tinggi sehingga apabila prestasi mahasiswa baik maka ukuran ini dapat dijadikan indikator bahwa proses pada perguruan tinggi tersebut telah berjalan dengan baik. Perlunya analisa mengenai prestasi mahasiswa dalam hal ini nilai indeks prestasi sangat bermanfaat agar kualitas sebuah perguruan tinggi dapat dipertahankan.

STMIK AKAKOM menyadari bahwa analisa mengenai indeks prestasi mahasiswa secara periodik sangatlah bermanfaat untuk mengendalikan proses yang ada pada tiap periodik. STMIK AKAKOM menyadari bahwa input mahasiswa baru setiap tahunnya mempunyai kualitas yang berbeda-beda sehingga akan mempengaruhi proses pembelajaran yang ada agar dapat mempertahankan kualitas yang telah ada sehingga diperlukan sebuah analisa dengan melakukan pengelompokan (cluster) indeks prestasi kedalam kelompok-kelompok tertentu. Clustering

indeks prestasi ini dapat dilakukan secara periodik sehingga dapat dibandingkan hasil cluster secara periodiknya.

Clustering merupakan salah satu metode *Data Mining* yang bersifat *unsupervised* (tidak terawasi). Terdapat dua jenis data *clustering* yang sering digunakan untuk pengelompokan data yaitu hirarki data *clustering* dan non-hirarki data *clustering*. K-means merupakan salah satu metode data *clustering* non-hirarki yang berusaha mempartisi data yang ada kedalam kelompok. Dari permasalahan diatas peneliti akan melakukan penelitian untuk melakukan *clustering* terhadap indeks prestasi mahasiswa STMIK AKAKOM menggunakan metode K-means.

2. Tinjauan Pustaka

Penelitian ini banyak mengacu pada tulisan-tulisan mengenai Clustering antara lain:

- Yudi Agusta, Ph.D dengan judul K-Means, penerapan, permasalahan dan metode terkait.
- Noor Rindho & Suzuki Syofian dengan judul Implementasi Data Mining dengan metode

Clustering untuk melakukan *Competitive Intelligence* Perusahaan.

- c. Vidya Ayuningtias dengan judul Pengkategorian Hasil Pencarian Dokumen dengan Clustering.

3. Landasan Teori

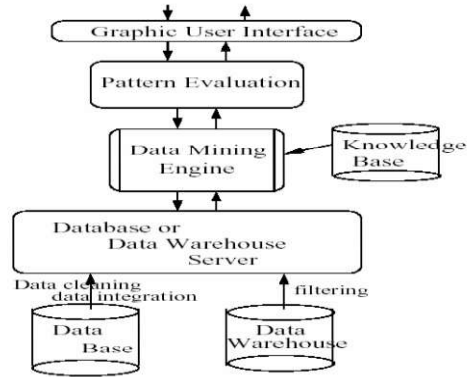
a. Data Mining

Data Mining memang salah satu cabang ilmu komputer yang relatif baru. Dan sampai sekarang orang masih memperdebatkan untuk menempatkan *data mining* di bidang ilmu mana, karena *data mining* menyangkut *database*, kecerdasan buatan (*artificial intelligence*) dan statistik. Ada pihak yang berpendapat bahwa *data mining* tidak lebih dari *machine learning* atau analisa statistik yang berjalan di atas *database*. Namun pihak lain berpendapat bahwa *database* berperan penting di *data mining* karena *data mining* mengakses data yang ukurannya besar (bisa sampai terabyte) dan disini terlihat peran penting *database* terutama dalam optimisasi *query*-nya. Definisi *data mining* (Jiawei, 2000) adalah proses mengekstraksi pola-pola yang menarik (tidak remeh-temeh, implisit, belum diketahui sebelumnya, dan berpotensi untuk bermanfaat) dari data yang berukuran besar. Terdapat beberapa istilah yang mempunyai kemiripan dengan *data mining*, yaitu ekstraksi pengetahuan, analisis pola, pengerukan data, dan lain-lain. Beberapa buku menulis bahwa *data mining* merupakan sinonim dari istilah *knowledge discovery in database* (KDD) (Jiawei,2000).

Data mining muncul berdasarkan fakta bahwa pertumbuhan data yang sangat pesat, tetapi miskin dengan pengetahuan. Alasan memilih *data mining* dibanding analisis data secara tradisional adalah :

- a. *Data mining* mampu menangani jumlah data kecil sampai data yang berukuran terabyte,
- b. Mampu menangani data yang mempunyai banyak dimensi, yaitu puluhan sampai ribuan dimensi,
- c. Mampu menangani data dengan kompleksitas yang tinggi, misalnya data stream, data sensor, data *spasial*, teks, data web, dan lain-lain.

Proses data mining dari basis data sampai pada user dapat dilihat pada gambar 1.



Gambar 1. Arsitektur data mining (Jiawei Han, 2000)

Contoh aplikasi data mining paling banyak digunakan dalam melakukan analisis dan manajemen pasar, manajemen keuangan, industri telekomunikasi, *text mining* dan *web mining*, *stream data mining*, analisis bioinformatika dan biodata, dan masih banyak lagi.

b. Clustering

Data *Clustering* merupakan salah satu metode *Data Mining* yang bersifat tanpa arahan (*unsupervised*). Ada dua jenis data *clustering* yang sering dipergunakan dalam proses pengelompokan data yaitu *hierarchical* (hirarki) data *clustering* dan *non-hierarchical* (non hirarki) data *clustering*. *Clustering* merupakan cara memasukkan suatu pola yang diamati ke suatu kelas pola yang belum diketahui dan disebut sebagai kluster pola. Tujuan dari *clustering* (*unsupervised classification*) adalah berusaha untuk mengelompokkan data dalam ruang ciri (*feature space*) secara natural ke dalam sejumlah cluster (Pedrycz, witold 2005).

Cluster merupakan suatu kelompok yang homogen, dimana tiap unit di dalamnya memiliki kemiripan satu sama lain. Untuk membentuk *clustering* dari sekumpulan data, maka kriteria dari kluster harus mempunyai kumpulan data yang homogen dan tidak serupa dengan kumpulan data lainnya, sedangkan *cluster* yang berbeda secara umum akan mengarah kepada kluster yang berbeda pula. Aplikasi dari *Clustering* antara lain: *engineering*, *bioinformatics*, *social sciences* (*sociology*, *archeology*), *medicine sciences* (*psychiatry*, *pathology*), data dan *web mining*.

c. K-Means

K-Means merupakan metode klusterisasi yang paling terkenal dan banyak digunakan di berbagai bidang karena sederhana, mudah diimplementasikan, memiliki kemampuan untuk mengkluster data yang besar, mampu menangani data outlier, dan kompleksitas waktunya linear $O(nKT)$ dengan n adalah jumlah dokumen, K adalah jumlah kluster, dan T adalah jumlah iterasi. *K-Means* merupakan salah satu metode data *clustering* non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*/kelompok yang berbeda. Dengan *partitioning* secara iteratif, K-Means mampu meminimalkan rata-rata jarak setiap data ke klusternya. Metode ini dikembangkan oleh Mac Queen pada tahun 1967. Metode ini mempartisi data ke dalam *cluster*/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain.

Adapun tujuan dari data *clustering* ini adalah untuk meminimalisasikan *objective function* yang diset dalam proses *clustering*, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu *cluster* dan memaksimalkan variasi antar *cluster*.

d. Algoritma K-Means

Dasar algoritma K-means adalah sebagai berikut (Budi Santoso, 2007):

1. Tentukan nilai k sebagai jumlah kluster yang ingin dibentuk.
2. Bangkitkan k centroid (titik pusat kluster) awal secara random.
3. Hitung jarak setiap data ke masing-masing centroid menggunakan rumus antar dua objek yaitu Euclidean Distance.
4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan centroidnya.
5. Tentukan posisi centroid baru (k C) dengan cara menghitung nilai pusat dari data-data yang ada pada centroid yang sama. Dimana k n adalah jumlah dokumen dalam *cluster* k dan i d adalah dokumen dalam *cluster* k .
6. Kembali ke langkah 3 jika posisi centroid baru dengan centroid lama tidak sama.

Adapun karakteristik dari algoritma K-Means salah satunya adalah sangat sensitif dalam penentuan titik pusat awal kluster, karena K-Means membangkitkan titik pusat kluster awal secara random. Pada saat pembangkitan awal titik pusat yang random tersebut mendekati solusi akhir pusat kluster, K-Means mempunyai posibilitas yang tinggi untuk menemukan titik pusat kluster yang tepat. Sebaliknya, jika awal titik pusat tersebut jauh dari solusi akhir pusat kluster, maka besar kemungkinan ini menyebabkan hasil pengelompokan yang tidak tepat. Akibatnya K-Means tidak menjamin hasil pengelompokan yang unik. Inilah yang menyebabkan metode K-Means sulit untuk mencapai optimum global, akan tetapi hanya minimum lokal. Selain itu, algoritma K-Means hanya bisa digunakan untuk data yang atributnya bernilai numerik.

e. Permasalahan K-Means

Beberapa permasalahan yang sering muncul pada saat menggunakan metode *K-Means* untuk melakukan pengelompokan data adalah:

1. Ditemukannya beberapa model *clustering* yang berbeda
2. Pemilihan jumlah *cluster* yang paling tepat
3. Kegagalan untuk *converge*
4. Pendeteksian *outliers*
5. Bentuk masing-masing *cluster*
6. Masalah *overlapping*

Keenam permasalahan ini adalah beberapa hal yang perlu diperhatikan pada saat menggunakan *K-Means* dalam mengelompokkan data. Permasalahan 1 umumnya disebabkan oleh perbedaan proses inialisasi anggota masing-masing *cluster*. Proses inialisasi yang sering digunakan adalah proses inialisasi secara random. Dalam suatu studi perbandingan (Pena, 1999), proses inialisasi secara random mempunyai kecenderungan untuk memberikan hasil yang lebih baik dan independent, walaupun dari segi kecepatan untuk *converge* lebih lambat. Permasalahan 2 merupakan masalah laten dalam metode *K-Means*. Beberapa pendekatan telah digunakan dalam menentukan jumlah *cluster* yang paling tepat untuk suatu *dataset* yang dianalisa termasuk di antaranya *Partition Entropy (PE)* dan *GAP Statistics* (Tibshirani, 2000). Satu hal yang patut diperhatikan mengenai metode-metode ini adalah pendekatan yang digunakan dalam mengembangkan metode-metode tersebut tidak

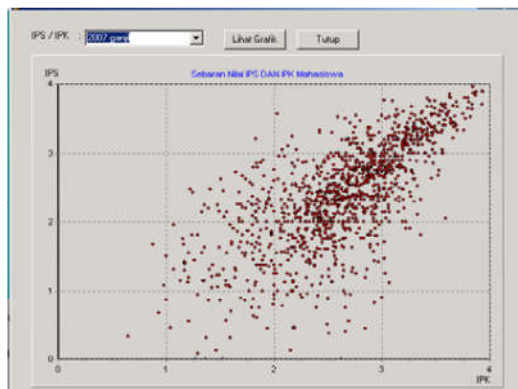
sama dengan pendekatan yang digunakan oleh *K-Means* dalam mempartisi data items ke masing-masing *cluster*.

Perpindahan suatu data ke suatu *cluster* tertentu dapat mengubah karakteristik model *clustering* yang dapat menyebabkan data yang telah dipindahkan tersebut lebih sesuai untuk berada di *cluster* semula sebelum data tersebut dipindahkan. Demikian juga dengan keadaan sebaliknya. Kejadian seperti ini tentu akan mengakibatkan pemodelan tidak akan berhenti dan kegagalan untuk *converge* akan terjadi.

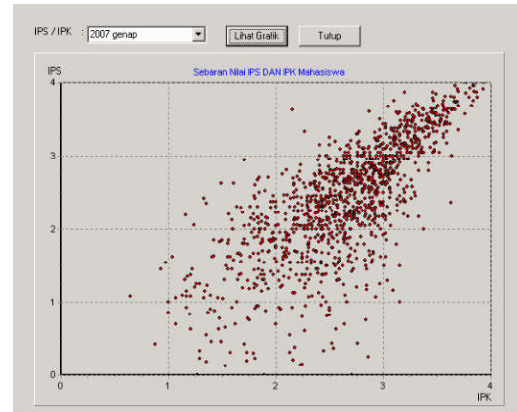
4. Hasil dan Pembahasan

a. Plotting Data Indeks Prestasi Mahasiswa

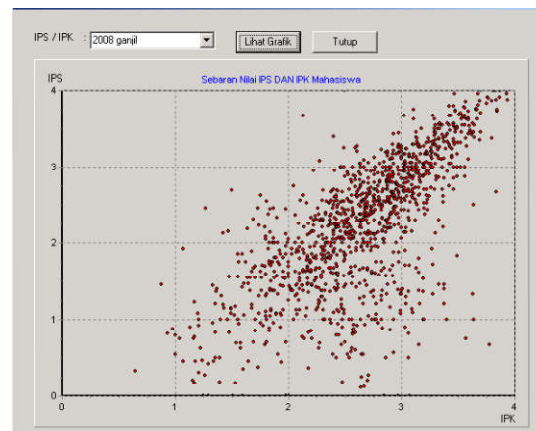
Data indeks prestasi yang digunakan pada analisa cluster k-means meliputi indeks prestasi semester (IPS) ganjil dan genap T.A 2007/2008 dan T.A 2008/2009, sedangkan nilai indeks prestasi kumulatif (IPK) yang digunakan merupakan indeks prestasi kumulatif pada saat semester Ganjil T.A 2009/2010. Total data yang digunakan pada penelitian ini sebanyak 1227 mahasiswa. Gambar 2 sampai gambar 5 menunjukkan sebaran data nilai indeks prestasi mahasiswa selama empat semester .



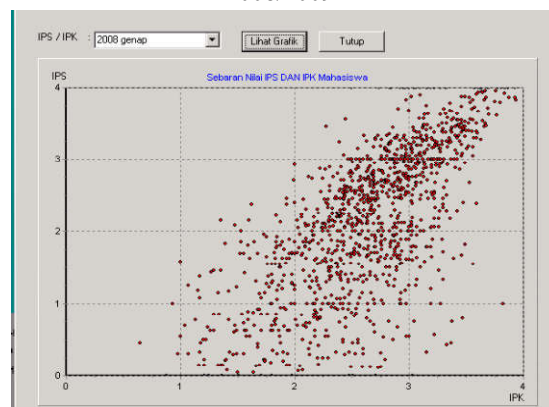
Gambar 2. Data Indeks Prestasi Semester Ganjil 2007/2008



Gambar 3. Data Indeks Prestasi Semester Genap 2007/2008



Gambar 4. Data Indeks Prestasi Semester Ganjil 2008/2009



Gambar 5. Data Indeks Prestasi Semester Genap 2008/2009

Dari hasil plotting data untuk empat semester dapat terlihat bahwa pada T.A 2007/2008 untuk semester ganjil dan genap mempunyai pola yang sama, hal ini dapat dikatakan bahwa indeks prestasi IPS terhadap IPK untuk T.A 2007/2008 relatif sama dan data cenderung mengumpul pada garis linier antara IPS dan IPK (gambar 1 dan gambar.2). Hal ini menunjukkan proses yang terjadi masih relatif normal karena hubungan antara data IPS terhadap IPK mempunyai kecenderungan garis yang linier.

Sedangkan untuk data indeks prestasi pada tahun 2008/2009 terdapat perbedaan yang cukup signifikan antara semester ganjil dan semester genap, hal ini terlihat pada sebaran data yang terjadi pada semester genap yang cenderung lebih banyak berada dibawah garis linier (gambar 5). Plotting data seperti ini diasumsikan bahwa nilai IPS mahasiswa pada semester genap 2008/2009 kurang baik dibandingkan semester ganjil 2008/2009. Banyak faktor yang menentukan mengenai kondisi data pada periode ini, antara lain beban mahasiswa yang cukup besar, proses penilaian kurang representatif dan proses pembelajaran belum maksimal.

Data indeks prestasi mahasiswa pada semester ganjil 2008/2009 dibandingkan dengan indeks prestasi T.A 2007/2008 masih cukup baik T.A 2007/2008, hal ini dapat dilihat pada sebaran data pada semester ganjil 2008/2009 yang cukup banyak berada dibawah garis linier (gambar 4).

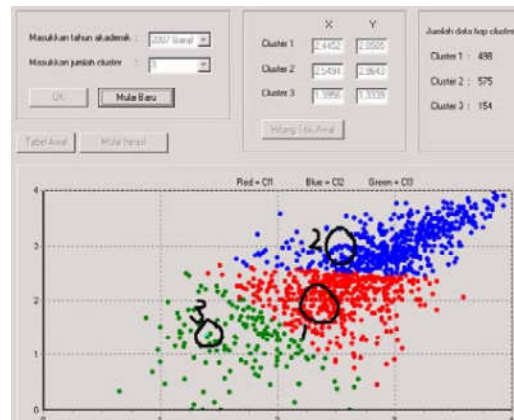
b. Hasil Analisa K-Means

Setelah melakukan plotting data terhadap nilai IPS dan IPK untuk masing-masing semester, maka dilakukan pengelompokan untuk masing-masing semester menjadi 3 cluster. Jumlah cluster sebanyak 3 dikarenakan rata-rata range indeks prestasi mahasiswa berada pada nilai 1 sampai 4. Dari range tersebut akan dibagi menjadi 3 kelompok, sehingga ditentukan cluster sebanyak 3. Untuk masing-masing semester akan dilakukan dua kali perhitungan k-means sehingga kita dapat membedakan hasil clustering

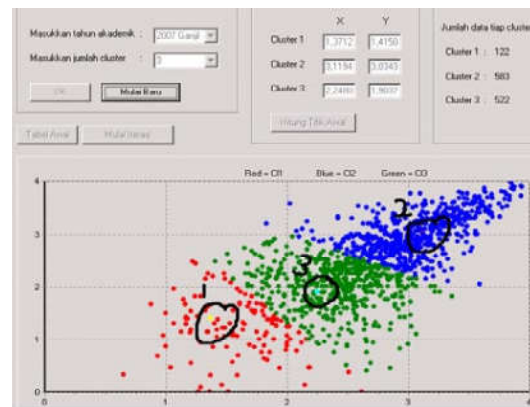
terutama nilai akhir dari masing-masing cluster. Untuk setiap semester dilakukan percobaan dua kali, sehingga dapat dibandingkan hasil untuk masing-masing percobaan.

b.1. Data Semester Ganjil 2007/2008

Untuk masing-masing percobaan hasil akhir cluster dapat dilihat pada gambar 5a dan 5b.



Gambar 5a. Hasil Cluster (percobaan 1)



Gambar 5b. Hasil Cluster (percobaan 2)

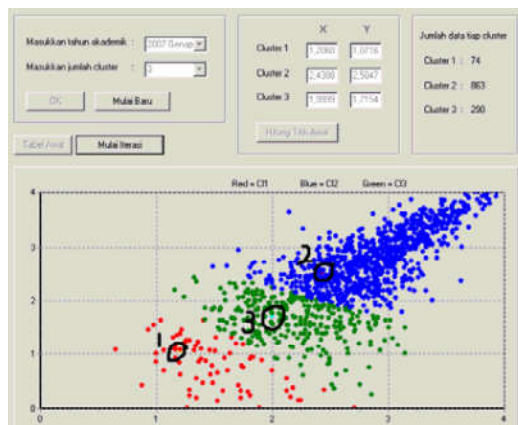
Perbedaan kedua percobaan untuk semester ganjil 2007/2008 dapat dilihat pada tabel 1. Pada tabel tersebut terlihat bahwa nilai tengah terbaik ditunjukkan oleh percobaan kedua.

Tabel 1. Perbandingan K-Means Semester Ganjil 2007/2008

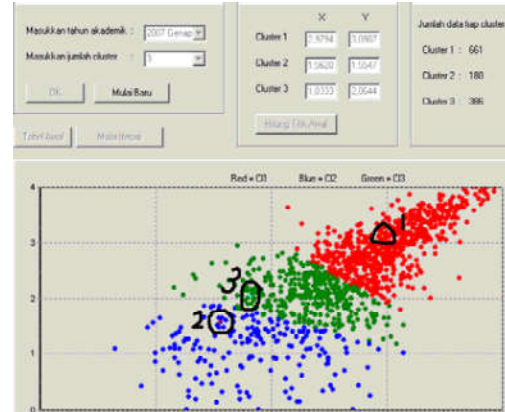
Hasil K-means	CL1		CL2		CL3	
	IPK	IPS	IPK	IPS	IPK	IPS
Nilai tengah AWAL1	1.3	2.7	2.8	3.3	3	1.2
Nilai tengah AKHIR1	2.4	2.05	2.5	2.96	1.39	1.33
Jumlah Data awal 1	173		695		359	
Jumlah Data akhir 1	498		575		154	
Nilai tengah AWAL 2	0.1	2.3	3	0.1	2	2.5
Nilai tengah AKHIR2	1.37	1.42	3.12	3.03	2.25	1.9
Jumlah Data awal 2	10		77		1140	
Jumlah Data akhir 2	122		583		522	

b.2. Data Semester Genap 2007/2008

Hasil akhir dari cluster untuk semester genap dengan dua kali percobaan dapat dilihat pada gambar 6a dan 6b.



Gambar 6a. Hasil akhir (percobaan 1)



Gambar 6b. Hasil akhir (percobaan 2)

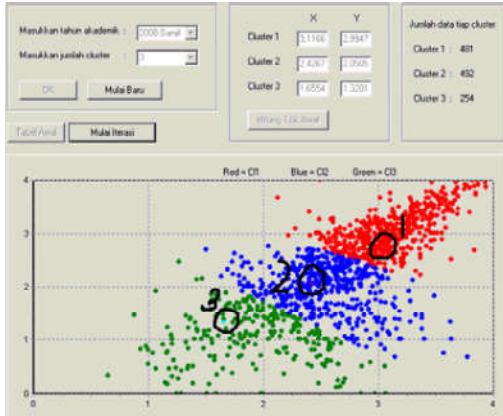
Perbedaan kedua percobaan untuk semester genap 2007/2008 dapat dilihat pada tabel 2. Dari tabel ini dapat dilihat bahwa nilai tengah terbaik ditunjukkan oleh percobaan kedua.

Tabel 2. Perbedaan Percobaan

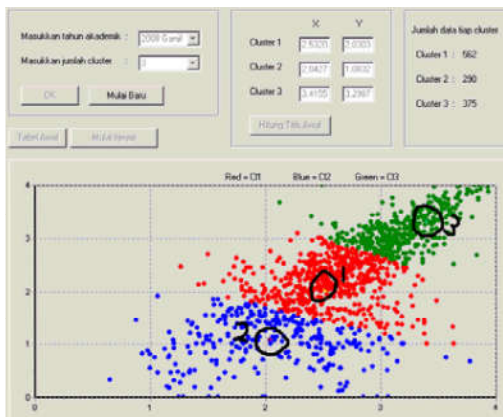
Hasil K-means	CL1		CL2		CL3	
	IPK	IPS	IPK	IPS	IPK	IPS
Nilai tengah AWAL1	0.6	1.1	2.2	0.7	0.7	1.9
Nilai tengah AKHIR1	1.21	1.07	2.44	2.5	1.9	1.72
Jumlah Data awal 1	27		957		243	
Jumlah Data akhir 1	74		863		290	
Nilai tengah AWAL 2	2.7	1.3	2.7	0.1	3.7	3.8
Nilai tengah AKHIR2	2.97	3.09	1.56	1.55	1.83	2.06
Jumlah Data awal 2	741		38		448	
Jumlah Data akhir 2	661		180		386	

b.3. Data Semester Ganjil 2008/2009

Hasil akhir dari cluster untuk semester genap dengan dua kali percobaan dapat dilihat pada gambar 7a dan 7b.



Gambar 7a. Hasil cluster (percobaan 1)

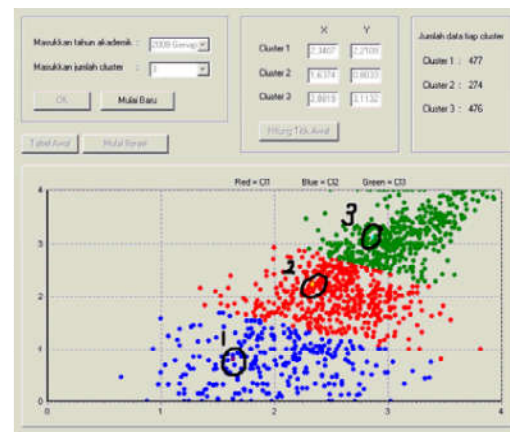


Gambar 7b. Hasil cluster (percobaan 2)

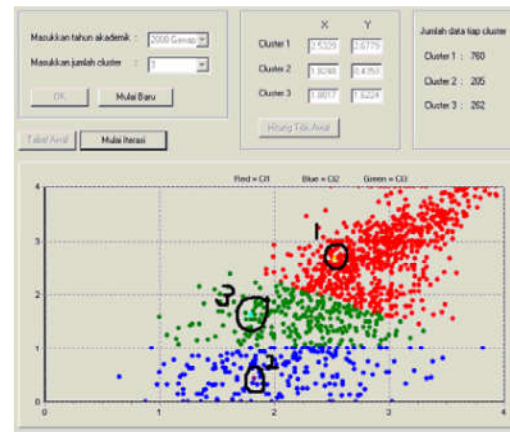
Nilai tengah AWAL 2	0.6	0.4	0.3	0.3	3.2	0.8
Nilai tengah AKHIR2	2.53	2.03	2.04	1.08	3.42	3.29
Jumlah Data awal 2	103		1		1123	
Jumlah Data akhir 2	562		290		375	

b.4. Data Semester Genap 2008/2009

Hasil akhir dari cluster untuk semester genap dengan dua kali percobaan dapat dilihat pada gambar 8a dan 8b.



Gambar 8a. Hasil cluster (percobaan 1)



Gambar bb. Hasil cluster (percobaan 2)

Perbedaan kedua percobaan untuk semester ganjil 2008/2009 dapat dilihat pada tabel 3. Pada tabel tersebut menunjukkan bahwa percobaan pertama memberikan nilai tengah yang lebih baik dibandingkan percobaan kedua.

Tabel 3. Perbandingan Hasil K-Means Semester Ganjil 2008/2009

Hasil K-means	CL1		CL2		CL3	
	IPK	IPS	IPK	IPS	IPK	IPS
Nilai tengah AWAL1	0.6	1.1	2.2	0.7	0.7	1.9
Nilai tengah AKHIR1	3.11	2.99	2.42	2.05	1.66	1.3
Jumlah Data awal 1	31		1053		143	
Jumlah Data akhir 1	481		492		254	

Perbedaan kedua percobaan untuk semester genap 2008/2009 dapat dilihat pada tabel 4. Dari tabel 4 terlihat bahwa nilai tengah terbaik ditunjukkan oleh hasil percobaan pertama.

Tabel. 4. Perbandingan Hasil K-Means Semester Genap 2008/2009

Hasil K-means	CL1		CL2		CL3	
	IPK	IPS	IPK	IPS	IPK	IPS
Nilai tengah AWAL1	0.1	2.3	3	0.1	2	2.5
Nilai tengah AKHIR1	2.34	2.2	1.64	0.8	2.88	3.11
Jumlah Data awal 1	10		238		979	
Jumlah Data akhir 1	477		274		476	
Nilai tengah AWAL 2	1.3	2.7	2.8	3.3	3	1.2
Nilai tengah AKHIR2	2.52	2.67	1.83	0.43	1.8	1.62
Jumlah Data awal 2	86		585		556	
Jumlah Data akhir 2	760		205		262	

5. Kesimpulan dan Saran

a. Kesimpulan

Dari penjelasan pada bab-bab sebelumnya maka hasil penelitian ini dapat disimpulkan antara lain:

1. Nilai tengah untuk masing-masing cluster pada tahun akademik 2007/2008 dan 2008/2009 mempunyai hasil yang berbeda
2. Hasil cluster 2007/2008 untuk semester ganjil dan genap mempunyai perbedaan tetapi tidak terlalu signifikan, hal ini dapat dikatakan bahwa proses pembelajaran yang terjadi pada semester ganjil dan genap pada T.A 2007/2008 relatif sama.
3. Hasil cluster 2008/2008 untuk semester ganjil dan genap mempunyai perbedaan yang cukup signifikan, hal ini menunjukkan bahwa proses pembelajaran yang terjadi pada T.A 2008/2009 tersebut tidak sama.
4. Hasil akhir dari clustering menggunakan K-Means sangat dipengaruhi oleh nilai awal titik tengah masing-masing cluster.

b. Saran

Peneliti menyadari bahwa hasil penelitian ini masih jauh dari sempurna sehingga perlu banyak masukan dari peneliti lain. Hal-hal yang dapat disarankan peneliti agar penelitian ini menjadi sempurna antara lain:

1. Menggunakan aturan baku untuk menentukan banyaknya jumlah cluster awal.
2. Menggunakan perhitungan nilai tengah cluster dengan ukuran pusat yang lain, misalnya rata-rata
3. Banyaknya cluster bisa dipilih secara interaktif

Daftar Pustaka

- Andrew W Moore, K-means and Heirarchical Clustering, School of Computer Science, Carniege Melon University, www.cs.cmu.edu/~awm.
- Budi Santoso, “ Data Mining: Teknik Pemanfaatan Data Untuk Keperluan Bisnis”, Graha Ilmu, Yogyakarta, 2007.
- Jiawei Han, Micheline Kamber, Data Mining : Concepts and Techniques, Morgan Kaufmann Publisher, Microsoft research, 2007.
- Pedrycz, witold, Knowledge base Clustering, John Wiley&Sons, Inc., 2005.
- Pena, J. M., Lozano, J. A. and Larranaga, P. (1999). An empirical comparison of four initialization methods for the k-means algorithm. *Pattern Recognition Lett.*, 20:1027-1040.
- Tibshirani, R., Walter, G. and Hastie, T. (2000). Estimating the Number of Clusters in a Dataset using the Gap Statistics, *Technical Report 208*, Department of Statistics, Stanford University, Standford, CA 94305, USA.
- <http://people.revoledu.com/kardi/tutorial/kMean>
- <http://www.autonlab.org/tutorials/kmeans11.pdf>