

PENERAPAN DATA MINING DALAM MENENTUKAN PENERIMA BEASISWA UPZ (Unit Pengumpulan Zakat) MENGGUNAKAN ALGORITMA K-MEANS

Sri Astuti¹, Samsudin², Triase³

^{1,2,3}Sistem Informasi, Fakultas Saintek, Universitas Islam Negeri Sumatera Utara Medan
e-mail: tutisrii55@gmail.com¹⁾, samsudin@uinsu.ac.id²⁾, triase@uinsu.ac.id³⁾

Abstrak:

Data Mining merupakan teknik yang digunakan dalam pengolahan data. Teknik data mining ini mampu menggali informasi dari data yang sudah lama atau tumpukkan data menjadi informasi yang berguna seperti halnya data mahasiswa pada sebuah universitas. Data Mahasiswa Pada Perguruan Tinggi memiliki tambang data yang jika diolah dengan menggunakan teknik data mining maka akan menghasilkan informasi baru dan berguna salah satunya dapat menghasilkan informasi untuk kelayakan seorang Mahasiswa dalam mendapatkan beasiswa. Seperti beasiswa unit pengumpulan zakat yang ada pada perguruan tinggi di Sumatera Utara. Kesulitan yang dialami oleh para panitia penyeleksi beasiswa unit pengumpulan zakat ialah menentukan kelayakan bagi penerima beasiswa unit pengumpulan zakat tersebut dikarenakan proses seleksi yang masih manual, sehingga memerlukan waktu yang lama dalam penyeleksiannya. Berdasarkan masalah yang ada maka dibuatlah sebuah sistem data mining menggunakan algoritma k-means yang mampu mengelompokkan mahasiswa yang layak mendapatkan beasiswa unit pengumpulan zakat berbasis web dengan menggunakan bahasa pemrograman php. Tujuan dibuatnya sistem ini ialah agar mempermudah serta mempercepat pihak panitia penyeleksi unit pengumpulan zakat dalam menentukan kelayakan penerima beasiswa unit pengumpulan zakat).

Kata Kunci : Beasiswa; Data Mining; K-Means; PHP

Abstract:

Data Mining is a technology used in data processing. This data mining technique is able to dig up information from old data or piles of data into useful information such as student data at a university. Student Data in Higher Education has a student data mine which, if processed using data mining techniques, will produce new and useful information such as information on the eligibility of a student in getting a scholarship. One of them is the zakat collection unit scholarship at universities in North Sumatra. The difficulty faced by the zakat collection unit scholarship selection committee is determining the eligibility for the zakat collection unit scholarship recipients because the selection process is still manual, so the selection process takes a long time. Based on the existing problems, a data mining system was made using the k-means algorithm that was able to classify students who deserved a web-based zakat collection unit scholarship using the PHP programming language. The purpose of this system is to make it easier and faster for the UPZ selection committee to determine the eligibility of zakat collection unit scholarship recipients.

Keywords: Scholarship; Data Mining; K-Means; PHP

1. PENDAHULUAN

Perkembangan teknologi saat ini memiliki peran yang sangat penting dan dapat dimanfaatkan untuk sebuah kemudahan seperti dalam pengolahan data dan informasi. Penerapan teknologi ini dirasakan dalam berbagai bidang seperti bidang industri, ekonomi dan ilmu teknologi dan dengan berkembangnya dunia teknologi maka intansi

mulai menerapkan penyimpanan data yang terkomputerisasi yang dapat menyimpan data yang sangat besar atau tambang data dalam sebuah *data base*. Hal ini memicu dunia teknologi untuk menerapkan ilmu baru dalam teknologi untuk dapat menggali informasi dari tambang data untuk mendapatkan suatu informasi yang bermanfaat yaitu *data mining*.

Pada salah satu perguruan tinggi di Sumatera Utara medan terdapat sebuah lembaga amil zakat yaitu UPZ (Unit Pengumpulan Zakat). Lembaga ini didirikan untuk menghimpun dana zakat pada seluruh ASN maupun civitas akademik dilingkungan perguruan tinggi tersebut. Dana zakat ini dilokasikan untuk beasiswa bagi para Mahasiswa yang memiliki keinginan untuk menuntut ilmu namun terhalang oleh factor ekonomi. Metode pengambilan keputusan dalam menentukan penerima beasiswa Unit Pengumpulan Zakat yang saat ini dilakukan, belum mampu untuk menangani data dalam jumlah yang sangat besar dari data Mahasiswa di perguruan tinggi tersebut. Sehingga memerlukan waktu yang lama dalam tahap penyeleksiannya. Hal ini dikarenakan jumlah data mahasiswa yang terus meningkat setiap tahunnya sehingga menyulitkan pihak penyeleksi dalam menentukan kandidat-kandidat yang akan menerima beasiswa tersebut. Belum lagi banyaknya kriteria – kriteria dalam penerima beasiswa tersebut membuat metode pengambilan keputusan yang saat ini diterapkan semakin kurang optimal.

Dalam mengatasi masalah tersebut perlu diterapkannya pengembangan yaitu sebuah sistem penentuan kelayakan terkomputerisasi yang dapat mengelompokkan data Mahasiswa yang layak, tidak layak dan dipertimbangkan dalam mendapatkan beasiswa UPZ UIN Sumatera Utara Medan menggunakan bantuan teknik data mining. Salah satunya menggunakan algoritma *k-means clustering*. Algoritma *K-Means Clustering* ini merupakan Algoritma yang mempartisikan data ke dalam cluster sehingga data yang memiliki karakter yang sama dikelompokkan kedalam satu cluster. Seperti dalam penelitian Gustientiedina, (2019) yang berjudul Penerapan Algoritma *K-Means* Untuk Clustering Data ObatObatan Pada RSUD Pekan Baru menyimpulkan bahwa “algoritma *k-means* merupakan salah satu dari algoritma yang banyak digunakan dalam pengelompokkan karena sederhana serta telah diakui sebagai salah satu dari 10 algoritma data mining peringkat teratas oleh IEEE (*Institute of Electrical and Electronics Engineers*)”.

Berdasarkan latar belakang tersebut tujuannya untuk menerapkan teknik data mining menggunakan algoritma *k-means* dalam menentukan Mahasiswa yang layak, tidak layak serta dipertimbangkan untuk mendapatkan bantuan beasiswa UPZ UIN Sumatera Utara Medan dengan implementasi sistem menggunakan bahasa pemrograman PHP.

2. METODE PENELITIAN

Tempat penelitian ini dilakukan pada Unit Pengumpulan Zakat Pada salah satu universitas di Sumatera Utara Medan dan dimulai pada february sampai dengan juli 2020. Cara kerja dalam penelitian ini ialah menggunakan metode *Mixed Method Research* atau kombinasi metode penelitian kualitatif dan kuantitatif serta menggunakan metode pengembangan sistem yaitu metode *Prototyping*.

2.1 Metode Pengumpulan Data

Beberapa metode pengumpulan data pada penelitian ini ialah

1. Observasi

Observasi yang dilakukan pada penelitian ialah pengamatan langsung pada Unit Pengumpulan Zakat UIN Sumatera Utara Medan. dalam melihat data-data Mahasiswa yang lulus masuk tahap wawancara pada penerima beasiswa UPZ serta kuota Mahasiswa yang akan diterima oleh UPZ.

2. Wawancara

Wawancara dilakukan dengan Bapak sekretaris UPZ Universitas Islam Negeri Sumatera Utara Medan terkait dengan perkembangan Beasiswa UPZ dari awal dibuat kebijakan sampai pada tahap pelaksanaan yang telah dilakukan selama dua tahun terakhir. Serta menanyakan hal yang menjadi kesulitan dalam menentukan Penerima Beasiswa UPZ Universitas Islam Negeri Sumatera Utara Medan.

3. Studi Pustaka

Metode pengumpulan data melalui studi pustaka ini ialah dengan mempelajari banyak jurnal dan buku-buku literature yang berkaitan dengan *data mining*.

2.2 Metode Pengembangan Sistem

Metode pengembangan sistem yang digunakan dalam penelitian ini ialah metode *prototyping*. Metode *prototyping* merupakan model pengembangan yang mempercepat dan menyederhanakan desain sistem serta pengujian dalam model kerja *prototype* ialah melalui proses interaksi dan dilakukan berulang-ulang yang dapat digunakan oleh ahli sistem. [1] Berikut adalah tahapan pada *prototype*:

1. Analisis Kebutuhan Sistem

Proses dalam analisis kebutuhan sistem memerlukan data dalam membangun dan mengembangkan sebuah sistem. Dalam hal ini data yang diperlukan ialah data Mahasiswa UIN Sumatera Utara Medan tahun 2019-2017 yang berjumlah 8543 Mahasiswa yang nantinya akan dilakukan tahapan pada *data mining*.

2. Desain Sistem

Pada proses ini mencakup beberapa tahapan seperti model proses, desain basis data dan desain *interface* sistem.

3. Pengujian Sistem

Pada proses ini dibuat pengujian sistem untuk menguji kesalahan-kesalahan apa saja yang terjadi ketika sistem berjalan.

4. Implementasi

Pada tahap ini implementasi sistem yang siap untuk dijalankan yang akan digunakan oleh *user*.

3. HASIL DAN PEMBAHASAN

3.1 Tahap *Data Mining*

Pada tahap ini yang secara langsung melibatkan teknik *data mining* dengan memilih data teknik penambangan dan penentuan algoritma yang akan digunakan. [2] Tahapan pada *data mining* diperlukan agar mempermudah proses pengambilan data yang akan dijadikan informasi yang bermanfaat dari tambang data yang besar. Berikut ini adalah tahapan pada *data mining* dalam membangun sistem data *mining* dalam menentukan penerima beasiswa UPZ pada penerapan algoritma *k-means*.

1. Tahap *Cleaning*

Proses *cleaning* meliputi antara lain membuang 4 duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data. [3] Dalam hal ini data yang akan dilakukan *cleaning* ialah data mahasiswa UIN Sumatera Utara Medan.

2. Tahap Transformasi

Proses Transformasi data ialah proses pengubahan data dari karakter menjadi data *numeric*. Dalam hal ini data yang telah dipilih sebagai kriteria untuk perhitungan *mining* akan ditransformasikan kedalam bentuk angka sehingga data tersebut dapat diterapkan pada algoritma *k-means clustering*. Metode *K-Means Clustering* hanya bisa mengolah data dalam bentuk angka, maka untuk data yang berbentuk nominal harus diinisialisasikan terlebih dahulu ke dalam bentuk angka. [4]

3. Tahap *Mining*

Merupakan satu proses utama dalam melakukan sebuah *data mining*. Penerapan algoritma *k-means* dilakukan pada tahap ini.[5]

K-Means merupakan salah satu metode pengelompokkan data *non-hierarki* (sekatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam kelompok yang lain.[6] Ide dasar algoritma *k-means* sangatlah sederhana, yaitu meminimalkan *Sum of Squared Error* (SSE) antara objek-objek data dengan sejumlah *k centroid*.

Algoritma *K-Means* sebenarnya melakukan 2 proses seperti proses mengetahui lokasi pusat *cluster* dan proses mencari anggota dari tiap-tiap *cluster*. Proses clustering dimulai dengan mengidentifikasi data yang akan dikluster, X_{ij} ($i=1, \dots, n; j=1, \dots, m$) dengan n adalah jumlah data yang akan dikluster dan m adalah jumlah variabel. Pada awal iterasi, pusat setiap kluster ditetapkan secara bebas.[7] Algoritma *k-means* bekerja dengan empat langkah, yang diilustrasikan dalam *pseudo code*. [8] Secara umum, cara kerja dari algoritma *K-Means clustering* adalah sebagai berikut:

1. Menentukan k sebagai jumlah kluster yang diinginkan.
2. Menentukan nilai *random* untuk pusat *centroid* (cluster awal) sebanyak k .
3. Menghitung jarak setiap data input terhadap masing-masing *centroid* menggunakan rumus jarak *Euclidean* (*Euclidean Distance*) sampai menemukan jarak terdekat dari setiap data dengan *centroid*. Berikut adalah persamaan pada rumus

Euclidean Distance :

$$d(a_i, b_j) = \sqrt{\sum (a_i - b_j)^2}$$

dimana :

a_i : data kriteria,

b_j : *centroid* pada *cluster* ke- j ,

4. Mengklasifikasikan data berdasarkan kedekatannya dengan *centroid* (jarak terdekat).
5. Memperbaharui nilai *centroid*. Nilai *centroid* baru diperoleh dari rata-rata *cluster* yang bersangkutan dengan menggunakan rumus:

$$b_j(t+1) = \frac{1}{NS_j} \sum_{j \in S_j} a_j$$

dimana :

$b_j(t+1)$: *centroid* baru pada *iterasi* ke (t+1)

NS_j : banyak data pada *cluster* S_j . [9]

Langkah 2 hingga 5 diulangi sampai cluster bertemu. Konvergensi dapat didefinisikan secara berbeda tergantung pada implementasinya. ketika langkah 2 dan 3 adalah berulang atau ada perubahan maka akan mengubah klaster. Sampai klaster atau kedudukan tidak berubah kemudian iterasi dihentikan dan didapatlah hasil akhirnya. [10]

Berikut adalah langkah-langkah nya:

1) Menentukan jumlah *cluster*

Jumlah cluster pada penelitian ini ialah 3 yaitu (C1) layak, (C2) dipertimbangkan, (C3) tidak layak.

2) Menentukan Pusat *Cluster (centroid)*

Berikut ini adalah pusat *centroid* pada penelitian ini yang diambil secara *random*:

Tabel 1. Pusat *Centroid* Awal Yang Diambil Secara Acak

<i>Centroid</i>	Pk (Pekerjaan Ayah)	Pb (Pekerjaan Ibu)	Pa (Penghasilan ayah)	Pi (Penghasilan Ibu)	Js (Jumlah Saudara)	Ka (Kondisi anak)	<i>Cluster</i>
1	5	5	6	6	3	1	C1 (Layak)
2	3	3	4	5	2	0	C2 (Dipertimbangkan)
3	1	1	2	3	1	0	C3 (Tidak Layak)

3) Perhitungan Jarak Data dengan *Centroid*

Perhitungan jarak data dengan centroid awal yang akan dihitung dengan rumus *Euclidean distance* yang nantinya akan ditemukan jarak terdekat atau perhitungan terkeci. Berikut ini adalah data Mahasiswa yang akan dijadikan sample perhitungan:

Tabel 2. Data Mahasiswa UIN Sumatera Utara Medan

Data Ke-	Pekerjaan Ayah	Pekerjaan Ibu	Penghasilan ayah	Penghasilan ibu	Jumlah saudara	Kondisi anak
1.	3	5	7	9	2	0
2.	3	5	7	9	1	0
3.	6	3	9	8	1	0
4.	5	5	7	9	2	0
5.	3	5	7	9	1	0

6.	1	5	5	9	2	0
7.	4	5	7	9	2	0
....						
....						

Berikut ini adalah perhitungan jarak data mahasiswa dengan centroid menggunakan rumus *Euclidean distance*:

$$d(ai,bj) = \sqrt{\sum (ai - bj)^2}$$

dimana :

ai : data kriteria,

bj : *centroid* pada *cluster* ke- j ,

Jarak data mahasiswa 1 dengan *centroid* 1, 2, 3 :

$$d(ai,bj) = \sqrt{(3 - 5)^2 + (5 - 5)^2 + (7 - 6)^2 + (9 - 6)^2 + (2 - 3)^2 + (0 - 1)^2} = 4$$

$$d(ai,bj) = \sqrt{(3 - 3)^2 + (5 - 3)^2 + (7 - 4)^2 + (9 - 5)^2 + (2 - 2)^2 + (0 - 0)^2} = 5,38$$

$$d(ai,bj) = \sqrt{(3 - 1)^2 + (5 - 1)^2 + (7 - 2)^2 + (9 - 3)^2 + (2 - 1)^2 + (0 - 0)^2} = 9,05$$

Jarak data mahasiswa 2 dengan *centroid* 1, 2, 3 :

$$d(ai,bj) = \sqrt{(3 - 5)^2 + (5 - 5)^2 + (7 - 6)^2 + (9 - 6)^2 + (1 - 3)^2 + (0 - 1)^2} = 4,35$$

$$d(ai,bj) = \sqrt{(3 - 3)^2 + (5 - 3)^2 + (7 - 4)^2 + (9 - 5)^2 + (1 - 2)^2 + (0 - 0)^2} = 5,47$$

$$d(ai,bj) = \sqrt{(3 - 1)^2 + (5 - 1)^2 + (7 - 2)^2 + (9 - 3)^2 + (1 - 1)^2 + (0 - 0)^2} = 9$$

Jarak data mahasiswa 3 dengan *centroid* 1, 2, 3 :

$$d(ai,bj) = \sqrt{(6 - 5)^2 + (3 - 5)^2 + (9 - 6)^2 + (8 - 6)^2 + (1 - 3)^2 + (0 - 1)^2} = 4,79$$

$$d(ai,bj) = \sqrt{(6 - 3)^2 + (3 - 3)^2 + (9 - 4)^2 + (8 - 5)^2 + (1 - 2)^2 + (0 - 0)^2} = 6,63$$

$$d(ai,bj) = \sqrt{(6 - 1)^2 + (3 - 1)^2 + (9 - 2)^2 + (8 - 3)^2 + (1 - 1)^2 + (0 - 0)^2} = 10,14$$

adapun hasil dari keseluruhan perhitungan jarak data mahasiswa dengan centroid awal ialah sebagai berikut:

Tabel 3. Hasil Perhitungan Jarak Data Mahasiswa Dengan *Centroid* awal

No	C1	C2	C3	Cluster
1	4,0000	5,3852	9,0554	C1
2	4,3589	5,4772	9,0000	C1
3	4,7958	6,6332	10,1489	C1
4	3,4641	5,7446	9,6954	C1
5	4,3589	5,4772	9,0000	C1
6	5,2915	5,0000	7,8740	C2

7	3,6056	5,4772	9,3274	C1
8	2,4495	5,1962	9,2736	C1
9	4,4721	5,9161	9,4868	C1
10	4,0000	4,7958	8,2462	C1
...				
...				

Setelah dilakukan beberapa iterasi maka perhitungan berhenti pada iterasi ke-5 dimana posisi cluster sudah tidak berubah lagi. Berikut ini adalah hasil iterasi ke-5 dan hasil cluster terakhir:

Tabel 4. Hasil Perhitungan Iterasi Ke-5

No	C1	C2	C3	cluster
1	1,96566	2,217635	5,642908	C1
2	1,964069	2,160063	5,616419	C1
3	2,547316	5,752272	5,301469	C1
4	1,096047	3,594149	5,945686	C1
5	1,090329	3,462813	5,322935	C1
6	2,502762	4,362537	3,980253	C1
...				

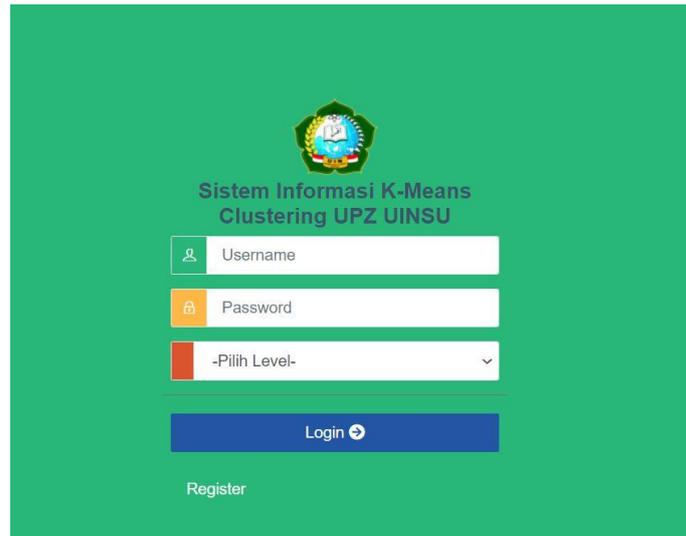
4) Tahap Evaluasi Pola (*Pattern Evaluation*)

Tahap ini merupakan tahap kesimpulan yang dihasilkan dari perhitungan mining. Kesimpulan pada tahap ini ialah bahwa perhitungan ini dihentikan pada iterasi ke-5 dikarenakan pada iterasi ke-6 posisi cluster sudah tidak berubah lagi.

3.2 Implementasi Sistem

1. Tampilan *Login*

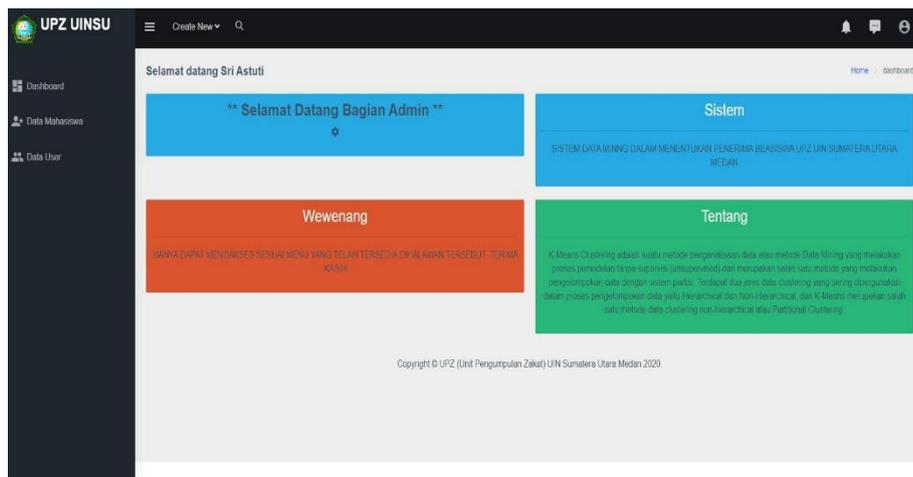
Tampilan ini merupakan tampilan pertama ketika user ingin menjalankan program dan mengoperasikan perintah didalamnya. Pada tampilan login terdiri dari menu *user*, *password*, dan *level*. Berikut ini adalah gambar dari tampilan login :



Gambar 1. Tampilan Login Administrasi dan Panitia

2. Tampilan Beranda

Berikut ini adalah tampilan halaman beranda administrasi



Gambar 2. Tampilan Halaman Beranda Administrasi

3. Tampilan Data Mahasiswa

Berikut ini adalah tampilan halaman data mahasiswa pada sistem informasi *k-means clustering* :

selanjutnya dapat juga ditambahkan dengan jumlah kriteria yang lebih banyak dan dengan jumlah cluster yang berbeda.

DAFTAR PUSTAKA

- [1] D. Purnomo, "Model Prototyping Pada Pengembangan Sistem Informasi," *J. Inform. Merdeka Pasuruan*, vol. 2, no. 2, pp. 54–61, 2017.
- [2] Yuda Irawan, "Implementation Of Data Mining For Determining Majors Using K-Means Algorithm In Students Of SMA Negeri 1 Pangkalan Kerinci," *J. Appl. Eng. Technol. Sci.*, vol. 1, no. 1, pp. 17–29, 2019, doi: 10.37385/jaets.v1i1.18.
- [3] G. W. N. Dicky Nofriansyah, *Algoritma Data Mining dan Pengujian*. Yogyakarta: CV Budi Utama, 2015.
- [4] F. L. Sibuea, A. Sapta, S. Informasi, and S. Royal, "Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering," *Tekno. dan Sist. Inf.*, vol. IV, no. 1, pp. 85–92, 2017.
- [5] M. Hasan, "Prediksi Tingkat Kelancaran Pembayaran Kredit Bank Menggunakan Algoritma Naive Bayes Berbasis Forward Selection," vol. 9, pp. 317–324, 2017.
- [6] E. Prasetyo, *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi Offset, 2012.
- [7] A. K. Wardhani, "Implementasi Algoritma K-Means untuk Pengelompokan Penyakit Pasien pada Puskesmas Kajen Pekalongan," *J. Transform.*, vol. 14, no. 1, pp. 30–37, 2016.
- [8] Suryanto, *Data Mining Untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika, 2019.
- [9] R. N. S. Pramana, B. Yuniarto, S. Mariyah, I. Santoso, *Data Mining dengan R Konsep Serta Implementasi*. Bogor: In Media, 2018.
- [10] K. Singh, D. Malik, and N. Sharma, "Evolving limitations in K-means algorithm in data mining and their removal," *IJCEM Int. J. Comput. Eng. Manag. ISSN*, vol. 12, no. April, pp. 2230–7893, 2011, [Online]. Available: www.IJCEM.org%5Cnwww.ijcem.org.