

ANALISA DAN PEMANFAATAN ALGORITMA K-MEANS CLUSTERING PADA DATA NILAI SISWA SEBAGI PENENTUAN PENERIMA BEASISWA

Ari Muzakir

Fakultas Ilmu Komputer, Universitas Bina Darma Palembang
e-mail : ariemuzakir@gmail.com,

ABSTRACT

Education can be said is one of the key formation of qualified human resources. But in fact, there are various problems that exist in the world of education this country. The existence Scholarship is one of the form. Scholarship in question is a scholarship for further education to university level measurement using a favorite with the data value or achievement of students in the school. But not easy to measure these students to be able to obtain a scholarship. One way to measure the value of data on student achievement levels of students. Value is an important component in student learning in the school system, because the value of the student to be one measure of student mastery of the subject matter. Students also become a reference value for decision making. Data values students need to be grouped to distinguish good and bad value with a range of groups of a certain value. The result of grouping these values can be used to create a school policy to provide scholarships. To solve the problems in the above explanation is the utilization of the K-Means Clustering algorithm. K-Means algorithm is the simplest clustering algorithm over other clustering algorithms. This algorithm has the advantages of easy to implement and run, relatively fast, easy to adapt, and the most widely practiced in the data mining tasks. Expected results with k-means clustering method is to determine the data value corresponding student to get a scholarship to college recommendation by using some variables, such as the data rate of students from grade 1 to grade 2 and the data on parental income. The end result is that there is good value group (who will get scholarships) and low grades (which failed).

keywords : algorithms, clustering, k-means, scholarships

PENDAHULUAN

Pendidikan bisa dikatakan adalah salah satu kunci pembentukan sumber daya manusia yang berkualitas. Namun pada kenyataannya, masih terdapat berbagai persoalan yang ada di dunia pendidikan negeri ini. Adanya Beasiswa adalah salah satu wujudnya. Beasiswa adalah bantuan yang diberikan oleh pihak tertentu kepada perorangan yang digunakan demi keberlangsungan pendidikan yang ditempuh. Pada dasarnya banyak sekali manfaat dari beasiswa bagi seorang siswa salah satunya misalnya dapat memberikan bantuan kepada siswa yang kurang mampu untuk mendapatkan kesempatan dalam menempuh pendidikan selanjutnya. Namun timbul berbagai pertanyaan mengenai pemberian beasiswa ini, misalnya apakah sudah sesuai dengan sasaran dan kriteria dalam pemberiannya.

Kabupaten Musi Banyuasin merupakan kabupaten yang cukup potensial dan memiliki banyak sumber daya alam dan sumber daya manusia. Sehingga banyak siswa yang memang memiliki potensi untuk memperoleh beasiswa. Beasiswa yang dimaksud adalah beasiswa untuk pendidikan lanjutan ke universitas favorit dengan tingkat pengukuran menggunakan data nilai atau prestasi siswa di sekolah tersebut. Namun tidak mudah untuk mengukur siswa-siswa tersebut untuk dapat memperoleh beasiswa. Salah satu cara mengukur tingkat prestasi siswa data nilai siswa. Nilai siswa merupakan komponen penting dalam sistem pembelajaran di sekolah, karena nilai siswa menjadi salah satu tolok ukur penguasaan materi pelajaran oleh siswa. Nilai siswa juga menjadi acuan untuk pengambilan keputusan. Data nilai siswa perlu dikelompokkan untuk membedakan nilai yang baik dan buruk dengan jangkauan kelompok nilai tertentu. Hasil pengelompokan nilai ini dapat digunakan untuk membuat suatu kebijakan sekolah untuk memberikan beasiswa.

Untuk menyelesaikan permasalahan pada penjelasan diatas adalah dengan pemanfaatan algoritma K-Means Clustering. Algoritma K-Means adalah algoritma klastering yang paling sederhana dibanding algoritma klastering yang lain. Algoritma ini mempunyai kelebihan mudah diterapkan dan dijalankan, relatif cepat, mudah untuk diadaptasi, dan paling banyak dipraktekkan dalam tugas data mining. Clustering merupakan suatu metode untuk pengelompokan dokumen dimana dokumen dikelompokkan

dengan konten untuk mengurangi ruang pencarian yang diperlukan dalam merespon suatu query (Grossman, David dan Ophir Frider, 2004, h.105).

Algoritma K-Means merupakan algoritma yang membutuhkan parameter input sebanyak k dan membagi sekumpulan n objek kedalam k cluster sehingga tingkat kemiripan antar anggota dalam satu cluster tinggi sedangkan tingkat kemiripan dengan anggota pada cluster lain sangat rendah. kemiripan anggota terhadap cluster diukur dengan kedekatan objek terhadap nilai mean pada cluster atau disebut sebagai centroid cluster (nango, Dwi Novianti, 2012).

Data nilai siswa hanya akan menjadi sekumpulan data yang tidak berguna jika tidak dilakukan penggalian data terhadapnya. Banyak informasi terpendam yang dapat diambil dari sekumpulan data tersebut sehingga dapat memberikan suatu pengetahuan untuk penentuan kebijakan. Penggalian data dapat dilakukan dengan cara pengelompokan data nilai siswa menjadi beberapa kelompok, kelompok nilai baik dan nilai buruk.

METODE PENELITIAN

Dalam melaksanakan proses seleksi penerimaan beasiswa ini, ada beberapa alur yang dilakukan dimulai dari input data, proses, sampai pada tahap akhir yaitu hasil yang diharapkan.

a. Alur Input Data

Data yang dibutuhkan dalam penelitian ini adalah data nilai siswa dari kelas 1 sampai kelas 2, penghasilan orang tua. Data-data tersebut dapat berupa dokumen yang telah terekam di SMA dan SMK se kabupaten Musi Banyuasin. Sedangkan untuk kebutuhan pengolahan data, maka Jenis file yang digunakan untuk input program adalah file Microsoft Excel (.xls atau .xlsx). selanjutnya data-data akan diolah dengan alat bantu sederhana memanfaatkan program yang dibangun dengan program berbasis PHP dan MySQL. File diimpor ke dalam program, kemudian ditampung pada tabel tampungan, menggunakan database MySQL. Sehingga data yang diolah adalah data yang sudah tertampung pada tabel MySQL. Banyak klaster ditentukan sebanyak dua, untuk kelompok nilai minimal dan kelompok nilai maksimal.

b. Proses

Data yang telah dimasukkan kemudian diproses. Proses yang ada pada program merupakan implementasi dari algoritma K-Means. Proses klastering dikatakan selesai apabila pusat klaster tidak lagi berubah.

c. Output yang diharapkan

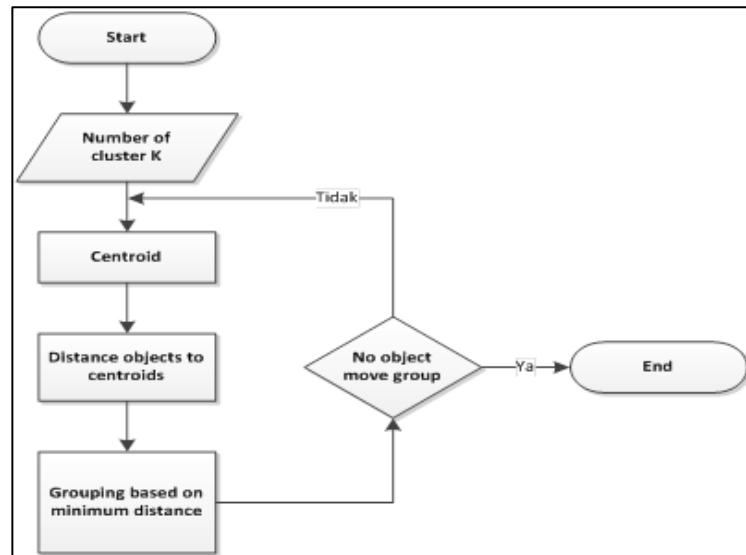
Keluaran yang diharapkan dari proses pengujian nantinya adalah tabel kelompok siswa dengan nilai baik (maksimal), tabel kelompok siswa dengan nilai buruk (minimal), tabel log proses pengujian.

Tahapan dalam penelitian dapat dilihat pada Gambar 1 ini sebagai berikut:



Gambar 1. Alur dalam tahapan penelitian

Pada penelitian ini melakukan analisis dan menerapkan algoritma K-means clustering untuk menentukan data nilai siswa yang sesuai untuk mendapatkan rekomendasi beasiswa ke perguruan tinggi. Menurut Santosa (2007), adapun langkah-langkah dalam penerapan algoritma K-means ini seperti berikut: 1. Menentukan data centroid, pada sistem ini, ditentukan bahwa centroid pertama adalah n data pertama dari data-data yang akan di-cluster. 2. Menghitung jarak antara centroid dengan masing-masing data. 3. Mengelompokkan data berdasarkan jarak minimum. 4. Jika penempatan data sudah sama dengan sebelumnya, maka stop. Jika tidak, kembali ke cara yang ke-2.



Gambar 2. Tahapan perhitungan pada algoritma K-Means

PEMBAHASAN

Pengujian sistem merupakan elemen kritis dalam pengembangan sebuah perangkat lunak (software) karena akan merepresentasikan hasil akhir dari spesifikasi kebutuhan dari aplikasi nantinya, yaitu perancangan dan implementasi. Tujuan utama dari pengujian sistem adalah untuk memastikan bahwa hubungan antar modul aplikasi telah memenuhi spesifikasi kebutuhan dan berjalan sesuai dengan skenario yang telah dideskripsikan sebelumnya. Pada pembahasan ini, pengujian menggunakan bahasa pemrograman PHP dan MySQL. Data inputan akan disimpan dalam database MySQL yang kemudian di olah memanfaatkan algoritma clustering. Jumlah iterasi yang dilakukan dalam pengujian ini sebanyak 3 (tiga) kali dengan jumlah centroid ada 3. Dalam ujicoba dalam penelitian ini, ada 3 komponen dalam penentuan kelayakan dalam penerimaan beasiswa yaitu nilai matematika, bahasa inggris, dan komputer. Jika dari ketiga nilai tersebut di nilai baik maka siswa akan mendapat rekomendasi untuk mendapatkan beasiswa tersebut. Berikut isi dari database siswa pada Gambar 3.

kode	nama	nilai	asal	mtk	ing	kom
MB005	ANISA RIZKI RAMADANA	237	SMAN 2 LALAN	90	60	87
MB115	ANNASH ZAENUN MUHENDRA	200	SMAN 1 SUNGAI KERUH	70	50	80
MB074	BUDI SAIFUL ROHMAN	177	SMA Bina Bangsa PP Al - Falah Sukamaju	60	60	57
MB043	MUNIROTUN NAIMAH	183	SMA Muhammadiyah Sekayu	60	60	63
MB102	MOHAMMAD AGUSTIAN	203	SMAN 2 BABAT TOMAN	60	50	93
MB106	FITRI SULISTIYANI	190	SMAN 1 PLAKAT TINGGI	50	60	80
MB002	YANIZAR SURYA DRATAMA	193	SMAN 1 SANGA DESA	50	70	73
MB004	PUTRI AYU SURYA CITRA	207	SMAN 2 BAYUNG LENCIR	50	70	87
MB062	SANDI SETIAWAN	210	SMAN 2 LAIS	50	80	80
MB064	AGUS SETIAWAN	177	SMA Bhakti Ibu 2 Sekayu	50	50	77
MB076	ARIFATUS SHOLEHA	163	SMAN 1 lais	50	40	73
MB019	CITRA PUSPITA SARI	187	SMAN 1 KELUANG	40	70	77
MB008	IIS TIANAH	163	SMKN 1 Sekayu	40	50	73
MB100	LISA WARANITA	183	SMA PGRI 268 Pangkalan Kersik	40	70	73
MB096	KURNITA SARI	160	SMK N 2 Sungai Lilin	40	40	80
MB001	MOH. HAIRUL ANAM	183	SMA Bina Bangsa Sejati	40	60	83
MB050	DEFFA ERLIAN SARI	167	MAN 1 Sekayu	40	50	77
MB029	NAZILA QORIATUN NISA	170	SMKN 1 Sungai Lilin	40	50	80
MB018	AYU AZKY SALSABYLA	177	MA AL FALAH SUKAJAYA	40	70	67
MB121	RULLI BAGUS PRATAMA	173	MA AL MUBAROQAH	40	50	83
MB003	TIWI ROBIANINGSIH	173	SMKN 1 Babat Toman	40	50	83
MB092	INTAN PRASTIKA	160	SMK PGRI Lais	30	70	60

Gambar 3. Struktur data Masukan

Dari data pada Gambar 3 tersebut, langkah selanjutnya adalah melakukan proses menggunakan algoritma k-means sehingga akan didapatkan hasil nilai yang masuk dalam kriteria baik. Dalam pembahasan ini yang dikatakan nilai baik adalah nilai yang diatas 70, jika nilai 60 kebawah maka akan langsung di tandai dengan angka 0 pada kolom C. berikut diperlihatkan pada Gambar 4 hasil dari iterasi proses 1 sampai proses ke 3.

Proses Iterasi Selanjutnya

Kode	Nama	Matematika	Inggris	Komputer	Centroid 1			Centroid 2			Centroid 3			C1	C2	C3
					81	65	65	65	81	65	65	65	65			
MB053	MOHAMMAD FAHRUR ROZI	30	50	80	55.235857918566	49.101934788764	38.091993909482	0	0	1						
MB080	MAULIDA NURSANTI	20	80	60	63.015871016753	45.287967496897	51.874849397372	0	1	0						
MB092	INTAN PRASTIKA	30	70	60	51.487862647424	37.027017163147	41.121770389904	0	1	0						
MB096	KURNITA SARI	40	40	80	50.309044912421	50.309044912421	35.369478367655	0	0	1						
MB104	LILING APRISTIANA	20	60	80	63.015871016753	51.874849397372	45.287967496897	0	0	1						
MB008	IIS TIANAH	40	50	73	44.384682042344	40.62019202318	30.232432915662	0	0	1						
MB076	ARIFATUS SHOLEHA	50	40	73	40.62019202318	44.384682042344	30.232432915662	0	0	1						
MB110	IKRIMATUS SHOLEHA	30	60	73	51.86520991956	41.593268686171	36.249137920784	0	0	1						
MB050	DEFFA ERLIAN SARI	40	50	77	45.278925690687	41.593268686171	29.427877939124	0	0	1						
MB029	NAZILA QORIATUN NISA	40	50	80	46.16275550335	42.55845661906	29.17190429163	0	0	1						
MB112	DEWI LUTFIANA	30	60	80	53.39475629685	43.485629810318	35.369478367655	0	0	1						
MB003	TIWI ROBIANINGSIH	40	50	83	47.22287581247	43.703546766824	29.223278392405	0	0	1						
MB114	AMALIYAH	30	60	83	54.313902456001	44.609416046391	35.411862419252	0	0	1						

Proses Iterasi Selanjutnya

Kode	Nama	Matematika	Inggris	Komputer	Centroid 1			Centroid 2			Centroid 3			C1	C2	C3
					75	60	72	38.333333333333	73.333333333333	66.833333333333	42.083333333333	57.916666666667	79.083333333333			
MB053	MOHAMMAD FAHRUR ROZI	30	50	80	46.786750261158	28.057975693196	14.474834483797	0	0	1						
MB080	MAULIDA NURSANTI	20	80	60	59.741108124975	20.670026608594	36.599464932336	0	1	0						
MB092	INTAN PRASTIKA	30	70	60	47.634021455258	11.280514172678	25.616157010761	0	1	0						
MB096	KURNITA SARI	40	40	80	41.097445176069	35.87826640182	18.060661671157	0	0	1						
MB104	LILING APRISTIANA	20	60	80	55.578772926361	26.215453457836	22.200319066776	0	0	1						
MB008	IIS TIANAH	40	50	73	36.414282912066	24.191940906806	10.199060414241	0	0	1						
MB076	ARIFATUS SHOLEHA	50	40	73	32.031234756094	35.850383540487	20.510667955969	0	0	1						
MB110	IKRIMATUS SHOLEHA	30	60	73	45.011109739708	16.889345754054	13.687737821373	0	0	1						
MB050	DEFFA ERLIAN SARI	40	50	77	36.742346141748	25.506535110307	8.4471395552974	0	0	1						
MB029	NAZILA QORIATUN NISA	40	50	80	37.269290307168	26.843683304147	8.2373640605883	0	0	1						
MB112	DEWI LUTFIANA	30	60	80	45.705579528106	20.508128469788	12.295832627358	0	0	1						
MB003	TIWI ROBIANINGSIH	40	50	83	38.026306683663	28.435599753361	9.0749196507007	0	0	1						
MB114	AMALIYAH	30	60	83	46.324939287602	22.551792242155	12.871965661856	0	0	1						

Proses Iterasi Selanjutnya

Kode	Nama	Matematika	Inggris	Komputer	Centroid 1			Centroid 2			Centroid 3			C1	C2	C3
					70	57.5	71.75	35.555555555556	74.444444444444	72.666666666667	40	53.529411764706	80.294117647059			
MB053	MOHAMMAD FAHRUR ROZI	30	50	80	41.524841962372	26.11843968814	10.608640468748	0	0	1						
MB080	MAULIDA NURSANTI	20	80	60	56.074169632728	20.815473826394	38.891429037709	0	1	0						
MB092	INTAN PRASTIKA	30	70	60	43.523700440105	14.527963669939	27.984486557542	0	1	0						
MB096	KURNITA SARI	40	40	80	35.697513919039	35.495783091572	13.532608318032	0	0	1						
MB104	LILING APRISTIANA	20	60	80	50.737683234456	22.458741321107	21.02272621001	0	0	1						
MB008	IIS TIANAH	40	50	73	30.94854600785	24.847435717361	8.1031413448108	0	0	1						
MB076	ARIFATUS SHOLEHA	50	40	73	26.604745817241	37.352012522836	18.337097233426	0	0	1						
MB110	IKRIMATUS SHOLEHA	30	60	73	40.097537330864	15.479576349197	13.966841602877	0	0	1						
MB050	DEFFA ERLIAN SARI	40	50	77	31.36578549949	25.220264241182	4.827831653808	0	0	1						
MB029	NAZILA QORIATUN NISA	40	50	80	32.004862440028	25.904859165705	3.5416454643508	0	0	1						
MB112	DEWI LUTFIANA	30	60	80	40.918363848033	17.12553504616	11.914487706194	0	0	1						
MB003	TIWI ROBIANINGSIH	40	50	83	32.90611645272	26.908395128566	4.4473078050437	0	0	1						
MB114	AMALIYAH	30	60	83	41.627064513367	18.608706312296	12.21434858757	0	0	1						

Gambar 4. Proses Iterasi 1 sampai 3

Setelah melalui proses iterasi sebanyak 3 kali, maka proses selanjutnya adalah pengelompokan dan penentuan hasil dari clustering. Dari proses tersebut, data akan langsung disimpan di dalam tabel hasil di Gambar 4, tabel hasil centroid di Gambar 5, nilai rata-rata di Gambar 6 berikut.

kode	predikat	d1	d2	d3	d4	d5
MB005	Baik	19	1	18	31	68
MB074	Cukup	39	21	2	11	48
MB001	Cukup	37	19	0	13	50
MB043	Cukup	37	19	0	13	50
MB100	Cukup	37	19	0	13	50
MB017	Cukup	36	18	1	14	51
MB019	Cukup	36	18	1	14	51
MB103	Cukup	36	18	1	14	51
MB106	Cukup	35	17	2	15	52
MB002	Cukup	34	16	3	16	53
MB115	Cukup	32	14	5	18	55
MB102	Cukup	31	13	6	19	56
MB004	Cukup	29	11	8	21	58
MB062	Cukup	28	10	9	22	59
MB064	Cukup	39	21	2	11	48
MB018	Cukup	39	21	2	11	48
MB050	Cukup	43	25	6	7	44
MB029	Cukup	42	24	5	8	45
MB112	Cukup	42	24	5	8	45
MB003	Cukup	41	23	4	9	46
MB114	Cukup	41	23	4	9	46

Gambar 4. Hasil dari proses Iterasi

nomor	c1a	c1b	c1c	c2a	c2b	c2c	c3a	c3b	c3c
1	75	60	72	38.33333333333333	73.33333333333333	66.83333333333333	42.08333333333333	57.91666666666667	79.08333333333333
2	70	57.5	71.75	35.55555555555556	74.44444444444444	72.66666666666667	40	53.529411764706	80.294117647059
3	70	57.5	71.75	35.55555555555556	74.44444444444444	72.66666666666667	40	53.529411764706	80.294117647059

Gambar 5. Hasil Centroid dari proses Iterasi

kode	rata_rata
MB005	79
MB062	70
MB004	69
MB102	67
MB115	66
MB002	64
MB106	63
MB017	62
MB019	62
MB103	62
MB100	61
MB043	61
MB001	61
MB074	59
MB064	59
MB018	59
MB121	57
MB114	57
MB003	57
MB112	56
MB029	56
MB050	55

Gambar 6. Hasil nilai rata-rata dari proses Iterasi

KESIMPULAN

Dari proses perancangan, implementasi dan pengujian dapat diambil beberapa kesimpulan sebagai berikut:

1. Algoritma K-Means dapat melakukan pengelompokan dokumen dalam jumlah yang banyak akan tetapi belum efisien dalam mengelompokan dokumen secara tepat.
2. Penentuan centroid (titik pusat) pada tahap awal Algoritma K-Means sangat berpengaruh pada hasil cluster seperti pada hasil pengujian yang dilakukan dengan menggunakan 100 record dengan centroid yang berbeda menghasilkan hasil cluster yang berbeda juga.

UCAPAN TERIMA KASIH

Terima kasih saya ucapkan kepada segenap panitia penyelenggara Seminar Nasional Aplikasi Sains & Teknologi 2014 Yogyakarta atas kesempatan yang telah diberikan dalam mengikuti seminar ini.

DAFTAR PUSTAKA

- Grossman, David A. dan Ophir Frieder. 2004. *Information Retrieval Algorithms and Heuristics* Second Edition. Springer, The Netherlands.
- Ken .2009. *Clustering Analysis, Part II: K-Means Clustering*. <http://www.centerspace.net>. Diakses pada tanggal 10 September 2014
- Nango, Dwi Novianti .2012. *Penerapan Algoritma K-means untuk Clustering Data Anggaran Pendapatan Belanja Daerah di Kabupaten XYZ*. <http://sro.web.id>. Diakses pada tanggal 16 Agustus 2014.
- Santosa, B. 2007. *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.