

KLASTERING TINGKAT KESULITAN SOAL PEMROGRAMAN BERORIENTASI OBJEK MENGGUNAKAN K-MEANS

Mayang Sari, Mirza Yogy Kurniawan, dan Wagino

Fakultas Teknologi Informasi

E-mail: mayang.uniska@gmail.com, mirza.yogy@gmail.com, ginouniska@gmail.com

ABSTRAK

Evaluasi pada tingkat kesukaran soal bisa dilakukan dengan menghitung proporsi jawaban benar terhadap butir soal, namun evaluasi tersebut berlaku secara umum kepada setiap kelas maupun pengajar. Ketika kelas yang akan dievaluasi berbeda pengajar dan tipenya, tentu tidak bisa dipandang sama secara umum. Penelitian ini bertujuan untuk mengukur dan mengevaluasi tingkat kesulitan soal ujian dengan tipe kelas berbeda. K-Means yang merupakan salah satu algoritma dari klastering yang merupakan salah satu teknik data mining yang mampu membagi-bagi data menjadi kelompok yang natural tanpa diperlukan proses pelabelan pada data yang ingin dikelompokkan, atau biasa disebut unsupervised learning. K-Means menghasilkan tiga kelompok soal pada data, yaitu Mudah, Sedang, dan Susah. Implementasi K-Means pada penelitian ini mendapatkan nilai Davies-Boudin Index 0.198 yang berarti klaster yang dihasilkan sudah baik, dan proporsi soal ujian mudah, sedang, dan susah sebesar 4-3-3 yang berarti tingkat kesulitan soal sudah cukup proporsional.

ABSTRACT

Evaluation at the level of difficulty of the questions can be done by calculating the proportion of correct answers to items, but the evaluation applies generally to each class and teacher. When the class to be evaluated is different from the instructor and the type, of course it cannot be seen as the same in general. This study aims to measure and evaluate the level of difficulty of exam questions with different types of classes. K-Means which is one of the algorithms of clustering which is a data mining technique that is able to divide data into natural groups without the need for the labeling process on the data to be grouped, or commonly called unsupervised learning. K-Means produces three groups of questions on the data, namely Easy, Medium, and Difficult. The implementation of K-Means in this study obtained a Davies-Boudin Index score of 0.198, which means that the resulting cluster is good, and the proportion of exam questions is easy, moderate, and difficult at 4-3-3, which means that the difficulty level of the questions is quite proportional.

PENDAHULUAN

Mata kuliah pemrograman berorientasi objek merupakan salah satu mata kuliah yang paling penting dalam perkuliahan teknik informatika. Mata kuliah ini menjelaskan tentang salah satu paradigma dalam perancangan dan pembangunan aplikasi. Paradigma

ini diterapkan dalam teknologi terkini pembangunan aplikasi diantaranya pembangunan aplikasi *mobile* Android.

Evaluasi pada tingkat kesukaran soal ujian bisa dilakukan dengan melakukan pemetaan. Pemetaan dilakukan dengan cara mengelompokkan (*clustering*) soal Ujian Nasional tahun sebelumnya yang memuat Kompetensi dan Indikator berdasarkan persentase daya serap hasil simulasi ujian mata kuliah Pemrograman Berorientasi Objek untuk diketahui tingkat kesukarannya.

Data mining adalah proses ekstraksi informasi yang tersembunyi, belum pernah diketahui dan berpotensi untuk berguna dari data yang banyak. Pembangunan aplikasi komputer *data mining* dapat membantu manusia dalam masalah prediksi ataupun rekomendasi dalam pengambilan keputusan. (Witten, 2011)

Klastering adalah salah satu teknik *data mining* yang mampu membagi-bagi data menjadi kelompok yang natural. K-Means yang merupakan salah satu algoritma dari klastering.

Triwulandari melakukan analisis butir soal ujian akhir semester pertama mata pelajaran Bahasa Inggris untuk kelas XI SMAN Jombang 2012/2013 untuk mengetahui validitas isi, reliabilitas, tingkat kesukaran dan daya beda soal. Hasilnya bahwa soal ujian tersebut memiliki validitas isi dengan kriteria layak dengan prosentase 70%, reliabilitas 0,633 dengan kriteria layak, tingkat kesukaran masuk kriteria mudah dengan prosentase 48% dan daya beda soal yang termasuk kriteria tidak layak dengan prosentase 40%. (Triwulandari, 2013)

Sri Adi meneliti tentang efektifitas *Team Accelerated Instruction* (TAI) pada siswa kelas X SMK Tunas Harapan 2008/2009. Uji coba instrumen dilakukan menggunakan uji validitas, tingkat kesukaran, daya beda dan reliabilitas. Teknik analisis data yang digunakan adalah analisis kovariansi. Sebagai persyaratan analisis yaitu populasi berdistribusi normal yang diuji menggunakan uji chi square, populasi mempunyai variansi yang sama yang diuji menggunakan metode Barlett dan uji linieritas menggunakan uji F. Hasilnya adalah metode TAI belum efektif digunakan pada mata pelajaran matematika dengan hasil analisis kovariansi sebesar 1,464 sedangkan daerah kritik uji F dengan $V_1=1$ dan $V_2=36$ pada taraf signifikansi 5% adalah $\{F \geq 4,11\}$. (Adi, 2011)

Chrisnajanti dalam penelitiannya tentang pengaruh program remedial terhadap ketuntasan belajar siswa menyatakan bahwa program remedial berpengaruh positif terhadap hasil belajar siswa, akan tetapi data yang digunakan dalam penelitiannya hanya berupa data sekunder nilai siswa sehingga tidak diketahui validitas, reliabilitas, tingkat kesukaran, dan daya pembeda dari soal yang diberikan guru. (Chrisnajanti, 2002)

Slamet dalam thesisnya mengkaji pemetaan sebaran alumni Polines untuk mengukur tingkat kesesuaian bidang kerja dengan kompetensi masing-masing program studi. Metode yang digunakan adalah klastering K-Means dengan empat variabel yaitu jenis perusahaan, klasifikasi jabatan, bidang kerja dan kompetensi prodi sedangkan untuk klasternya ada tiga yaitu Sesuai, Kurang Sesuai, dan Tidak Sesuai. Hasilnya 51 alumni bekerja sesuai dengan kompetensinya, 23 alumni kurang sesuai dan sisanya 26 tidak sesuai. (Slamet, 2013)

Hugo Aprilianto dan Erfan Ramadhani dalam penelitiannya melakukan pengelompokan lulusan STMIK Banjarbaru berdasarkan IPK, lama studi, dan kompetensi lainnya menggunakan K-Means. Hasilnya kompetensi CTS dan Jarkom memiliki kelompok memuaskan paling banyak dengan presentase 36% sedangkan untuk kategori cukup memuaskan ada pada kompetensi Pemrograman Web dengan presentase 46%. (Hugo Aprilianto, Erfan Ramadahani, 2013)

Penelitian ini akan melakukan klastering terhadap rata-rata jumlah jawaban siswa yang benar pada tiap soal ujian mata kuliah Pemrograman Berorientasi Objek dengan menggunakan K-Means yang dikenal mampu membagi data kedalam kelompok yang natural.

METODE PENELITIAN

Penelitian ini melalui beberapa tahapan yang dilakukan agar dapat menghasilkan aplikasi yang mampu membagi sifat soal menjadi beberapa bagian :

1. Menyusun soal untuk ujian mata kuliah Pemrograman Berorientasi Objek dengan dikategorikan berdasarkan pokok bahasan.
2. Melaksanakan simulasi ujian kepada 3 (tiga) kelas, terdiri dari kelas reguler pagi, reguler malam, dan kelas online dengan menggunakan edmodo
3. Ekstrak data hasil ujian kedalam format csv
4. Data yang didapatkan diproses dengan menggunakan Rapidminer

5. Pengujian Davies-Bouldin Index, pada variasi klustering.

PEMBAHASAN

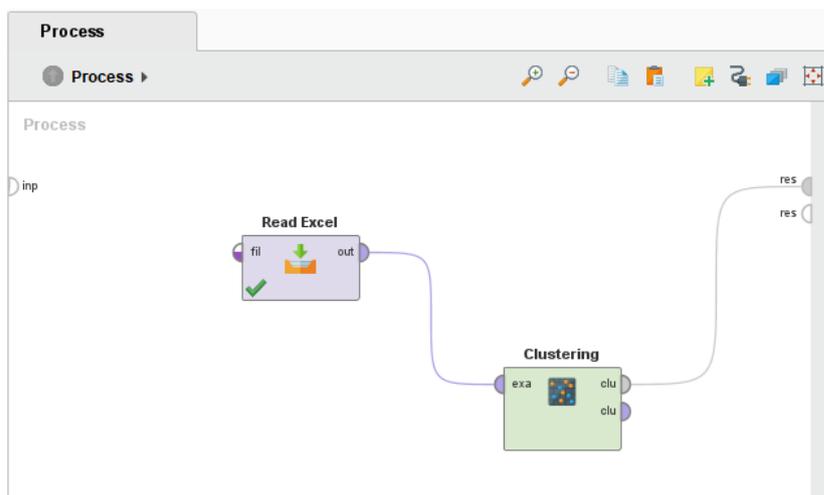
Table 1 yang dihasilkan dari pembuatan soal dan ujicoba ke mahasiswa, yang mana tiap soal dikategorikan sebagai soal teori atau soal koding. Nilai persentase jawaban benar mahasiswa dari tiap-tiap kelas dijadikan sebagai atribut.

Tabel 1 Persentase jawaban benar

No	TIPE	Pagi (%) 37	Malam (%) 23	Online (%) 35
1	Teori	94	86	71
2	Teori	84	60	51
3	Teori	94	86	5
4	Teori	71	69	4
5	Teori	71	52	82
6	Teori	39	65	57
7	Teori	31	65	60
8	Teori	7	43	45
9	Teori	7	39	51
10	Teori	39	56	60
11	Teori	18	52	74
12	Coding	26	39	45
13	Teori	84	60	14
14	Teori	89	91	54
15	Teori	65	73	54
16	Coding	84	82	37
17	Coding	28	78	20
18	Teori	81	69	28
19	Teori	71	82	77
20	Teori	65	69	14
21	Coding	86	82	68
22	Teori	89	91	60
23	Coding	36	73	42
24	Coding	94	86	88
25	Coding	42	56	82
26	Teori	89	73	48
27	Teori	86	73	22
28	Teori	55	47	22
29	Teori	94	86	62
30	Teori	86	86	48
31	Teori	76	69	65
32	Coding	89	86	8
33	Teori	68	73	71

34	Coding	71	60	60
35	Coding	97	78	48
36	Teori	92	69	37
37	Teori	97	100	85
38	Teori	92	78	31
39	Teori	94	82	42
40	Teori	55	65	68
41	Teori	84	86	45
42	Teori	63	52	28
43	Teori	26	60	28
44	Teori	10	39	68
45	Coding	57	52	22
46	Coding	13	52	28
47	Teori	31	60	71
48	Coding	55	69	65
49	Coding	36	56	60
50	Coding	63	73	37

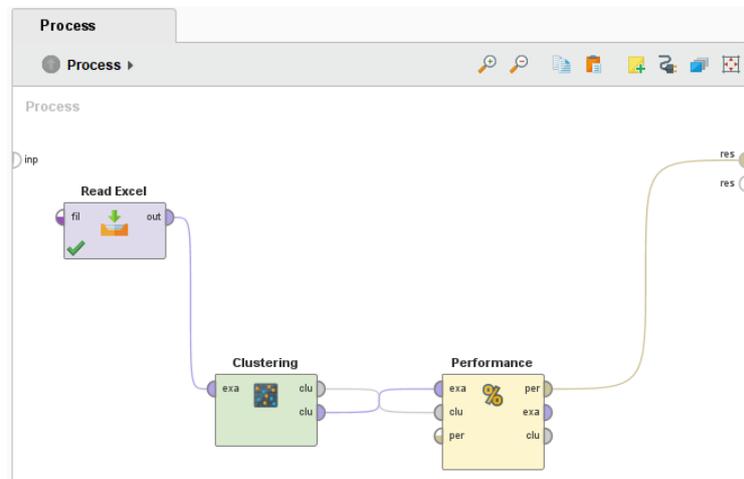
Data pada tabel 1 dioperasikan ke dalam Rapidminer dengan menggunakan clustering K-Means dan K-Medoids. Konfigurasi klastering yang digunakan adalah *numerical measures* dengan metode *Euclidian Distance*, sedangkan nilai K atau jumlah klaster yang dihasilkan akan menjadi salah satu parameter untuk pengujian dari klastering ini. Skema Rapidminer pada tahapan ini dapat dilihat pada gambar 1



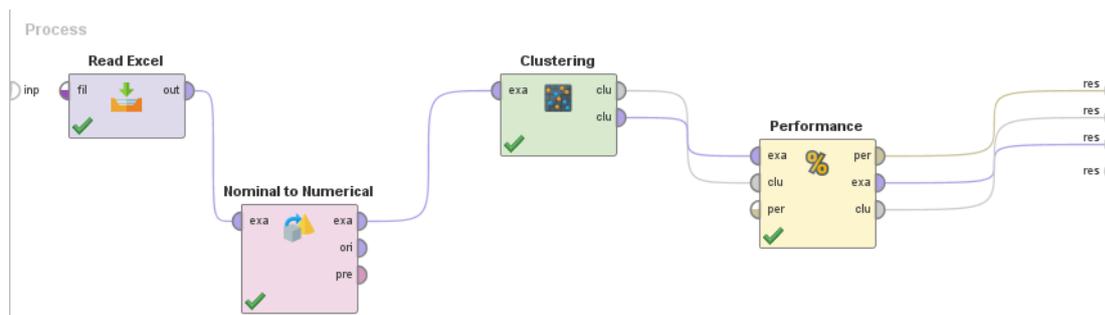
Gambar 1 Clustering di Rapidminer

Performa dari hasil klastering dapat diuji dengan Davies-Bouldin *Index*, yang memvalidasi seberapa baik klastering dilakukan berdasarkan kuantitas dan *inherent features* dari dataset. Hasil validasi ini dinormalisasi dan maksimasi sehingga

menghasilkan indeks antara 0-1 dengan semakin rendah nilainya semakin baik. Implementasi validasi ini diterapkan pada Rapidminer seperti gambar 2



Gambar 2 Mengukur Performa Klastering



Gambar 3 Konversi nominal ke numerik

Pada awalnya dataset yang digunakan hanya pada atribut yang bersifat numerik (presentase nilai benar), karena secara bawaannya K-Means hanya dapat dilakukan pada data numerik. Jumlah klaster yang diinginkan ditentukan pada nilai K. Pada pelaksanaannya DBI yang dihasilkan masih terlalu tinggi, sehingga ditambahkan atribut nominal (tipe soal) yang ternyata mampu menurunkan nilai DBI meskipun klaster yang dihasilkan tetap sama.

Tabel 2 Pengukuran DBI

K	Tippe Data	Pengukuran Centroid	DBI
3	Numeric	Euclidian Distance	0.299
3	Numeric	Corelation Similarity	0.362
3	Numeric	Cosine Similarity	0.330
5	Numeric	Euclidian Distance	0.330

5	Numeric	Corelation Similarity	0.831
5	Numeric	Cosine Similarity	0.379
7	Numeric	Euclidian Distance	0.272
7	Numeric	Corelation Similarity	0.401
7	Numeric	Cosine Similarity	0.416
3	Numeric + Nominal	Euclidian Distance	0.198
3	Numeric + Nominal	Corelation Similarity	0.214
3	Numeric + Nominal	Cosine Similarity	0.228
5	Numeric + Nominal	Euclidian Distance	0.198
5	Numeric + Nominal	Corelation Similarity	0.214
5	Numeric + Nominal	Cosine Similarity	0.228
7	Numeric + Nominal	Euclidian Distance	0.163
7	Numeric + Nominal	Corelation Similarity	0.273
7	Numeric + Nominal	Cosine Similarity	0.250

Berikutnya hasil klastering diambil dengan K=3. K=3 diambil karena lebih mudah mengkategorikannya dibandingkan dengan K=7 meskipun nilai DBI-nya lebih baik. Klastering menggunakan nilai centroid sebagai patokan kluster dan mengkluster data tersebut sesuai dengan jarak terhadap centroid. Tabel 3 menyajikan data centroid untuk K=3.

Tabel 3 Centroid

	Cluster_0	Cluster_1	Cluster_2
P 37	81.38	76.86	25.93
M 23	77.62	68.93	55.53
NR 35	62.48	22.07	52.73
Rerata	73.83	55.95	44.73

Nilai rerata digunakan untuk mengurutkan centroid sehingga dapat diasumsikan cluster_0 = mudah, cluster_1 = sedang, dan cluster_2 = susah. Tabel 4.4 menyajikan data hasil kluster K-Means dengan K=3 dan nama kluster menggunakan asumsi sebelumnya.

Tabel 4 Hasil klastering

No	TIPE	Pagi (%) 37	Malam (%) 23	Online (%) 35	K=3	K=3
					(Numeric)	(Numeric & Nominal)
1	Teori	94	86	71	mudah	mudah
2	Teori	84	60	51	mudah	mudah
3	Teori	94	86	5	sedang	sedang
4	Teori	71	69	4	sedang	sedang
5	Teori	71	52	82	mudah	mudah
6	Teori	39	65	57	susah	susah

7	Teori	31	65	60	susah	susah
8	Teori	7	43	45	susah	susah
9	Teori	7	39	51	susah	susah
10	Teori	39	56	60	susah	susah
11	Teori	18	52	74	susah	susah
12	Coding	26	39	45	susah	susah
13	Teori	84	60	14	sedang	sedang
14	Teori	89	91	54	mudah	mudah
15	Teori	65	73	54	mudah	mudah
16	Coding	84	82	37	sedang	sedang
17	Coding	28	78	20	susah	susah
18	Teori	81	69	28	sedang	sedang
19	Teori	71	82	77	mudah	mudah
20	Teori	65	69	14	sedang	sedang
21	Coding	86	82	68	mudah	mudah
22	Teori	89	91	60	mudah	mudah
23	Coding	36	73	42	susah	susah
24	Coding	94	86	88	mudah	mudah
25	Coding	42	56	82	susah	susah
26	Teori	89	73	48	mudah	mudah
27	Teori	86	73	22	sedang	sedang
28	Teori	55	47	22	sedang	sedang
29	Teori	94	86	62	mudah	mudah
30	Teori	86	86	48	mudah	mudah
31	Teori	76	69	65	mudah	mudah
32	Coding	89	86	8	sedang	sedang
33	Teori	68	73	71	mudah	mudah
34	Coding	71	60	60	mudah	mudah
35	Coding	97	78	48	mudah	mudah
36	Teori	92	69	37	sedang	sedang
37	Teori	97	100	85	mudah	mudah
38	Teori	92	78	31	sedang	sedang
39	Teori	94	82	42	mudah	mudah
40	Teori	55	65	68	mudah	mudah
41	Teori	84	86	45	mudah	mudah
42	Teori	63	52	28	sedang	sedang
43	Teori	26	60	28	susah	susah
44	Teori	10	39	68	susah	susah
45	Coding	57	52	22	sedang	sedang
46	Coding	13	52	28	susah	susah
47	Teori	31	60	71	susah	susah
48	Coding	55	69	65	mudah	mudah
49	Coding	36	56	60	susah	susah
50	Coding	63	73	37	sedang	sedang

Disimpulkan bahwa meskipun dengan DBI yang berbeda, namun dengan tipe data numerik saja menghasilkan klaster yang sama.

Tabel 5 Rangkuman hasil klastering

	Mudah	sedang	susah	total
Teori	16 (32%)	10 (20%)	9 (18%)	35 (70%)
Coding	5 (10%)	4 (8%)	6 (12%)	15 (30%)
total	21 (42%)	14 (28%)	15 (30%)	

KESIMPULAN

Metode klastering K-Means dengan perhitungan *Euclidian Distance* mampu mengelompokkan data persentase jawaban benar dari 3 (tiga) kelas berbeda dengan hasil $K=3$, yaitu mudah, sedang, dan susah dengan Davies-Boudin Index 0.198 yang menunjukkan klastering yang dihasilkan sudah baik.

Soal matakuliah PBO telah terbagi menjadi 3 klaster dengan proporsi mudah, sedang, susah sebesar 4-3-3 yang ini cukup baik dan proporsional

DAFTAR PUSTAKA

- Adi, W. S. (2011). Efektifitas Model Pembelajaran Team Accelerated Instruction pada Siswa Kelas X SMK Tunas Harapan 2008-2009. *Pemantapan Keprofesionalan Peneliti, Pendidik, dan Praktisi MIPA Untuk Mendukung Pembangunan Karakter Bangsa*.
- Chrisnajanti, W. (2002). Pengaruh Program Remedial terhadap Ketuntasan Belajar Siswa. *Jurnal Pendidikan Penabur* , 81-86.
- Hugo Aprilianto, Erfan Ramadahani. (2013). Implementasi Metode K-Means Clustering untuk Pengelompokan Lulusan STMIK Banjarbaru Berdasarkan Kompetensi. *Jurnal Progresif STMIK Banjarbaru*, 907-918.
- Slamet, H. (2013). *Sistem Informasi Geografis Berbasis Web untuk Pemetaan Sebaran Alumni Menggunakan Metode K-Means*. Semarang: Universitas Diponegoro.
- Triwulandari, E. F. (2013). An Analysis of English Final Test for the First Semester of the Eleventh Grade Students of SMA Negeri Jombang 2012/2013. *E-Journal Unesa*, 2-6.
- Witten, I. H. (2011). *Data Mining Practical Machine Learning Tools and Techniques Third Edition*. Amsterdam: Morgan Kauffman Press.

