

ANALISA PERBANDINGAN ALGORITMA K-MEANS, DECISION TREE, DAN NAÏVE BAYES UNTUK SISTEM PENGELOMPOKKAN SISWA OTOMATIS

Oktariani Nurul Pratiwi

Teknik Informatika, Fakultas Teknik, Universitas Widyatama
Bandung
oktariani.pratiwi@widyatama.ac.id

Abstrak

Pembelajaran kooperatif adalah proses pembelajaran dengan mengikuti beberapa instruksi yang melibatkan siswa bekerja dalam tim untuk mencapai tujuan bersama dalam kondisi yang mencakup unsur-unsur yang telah ditentukan. Pembelajaran kooperatif mampu membuat siswa lebih percaya diri dan mampu mengungkapkan pendapat sehingga, dapat meningkatkan kemampuan siswa lebih efektif. Salah satu faktor penting pembelajaran kooperatif adalah komposisi siswa dalam kelompok. Pada teorinya, penentuan kelompok belajar di sekolah dilakukan berdasarkan pengamatan guru terhadap siswa dengan mempertimbangkan banyak parameter. Namun, hal ini tidak mudah karena membutuhkan waktu lama, dan membuat beban kerja guru bertambah. Sehingga, cara termudah penentuan kelompok belajar adalah ditentukan secara acak. Risikonya, proses pembelajaran kooperatif tidak berjalan dengan efektif. Dibutuhkan sebuah sistem yang mampu menentukan komposisi anggota belajar siswa secara otomatis. Dalam paper ini dipaparkan hasil analisis perbandingan algoritma K-means, Decision Tree, dan Naive Bayes yang digunakan untuk mengelompokkan siswa. Dari hasil pengelompokkan menggunakan algoritma tersebut, kemudian dibandingkan dengan pengelompokkan yang dilakukan oleh guru. Hasilnya, Naive Bayes mampu mengelompokkan siswa lebih baik dengan nilai akurasi 70,37%.

Kata kunci :

Pembelajaran Kooperatif, Pengelompokkan Siswa Otomatis, K-Means, Decision Tree, Naive Bayes.

Abstract

Cooperative learning is a process of learning that following some instructions involving students work in team to achieve some goals. Cooperative learning

makes students more confident to tell opinions, and improve the ability of students more effectively. Formation of students in a group is important. Determination of group learning in school is based on the observation of teacher use many parameters. But, it is not easy. It takes a long time, and increase teacher task. The easiest way of determining groups of students is determined randomly. Hence, cooperative learning become inefficient. This process need a system that capable to determine the formation of the student's learning automatically. In this paper presented the results of a comparative analysis of K-means algorithm, Decision Tree, Naive Bayes. Then, the results of automatic students formation compared to the formation of teacher's. The best result of the experiment is used Naive Bayes 70,37% accurately.

Keywords :

Cooperative Learning, Automatic Grouping of Student, K-means, Decision Tree, Naive Bayes.

I. PENDAHULUAN

Proses pembelajaran yang biasa terjadi baik melalui media *online learning* maupun pembelajaran di kelas adalah proses pembelajaran satu arah. Dimana guru menyampaikan materi dan siswa menerima materi yang diberikan. Hal ini menimbulkan beberapa permasalahan dalam proses pembelajaran, diantaranya siswa menjadi pasif (Shen, Wang, Gao, Novak, & Tang, 2009) dan tidak percaya diri dalam mengungkapkan pendapat (Hoic-bozic, Mornar, Boticki, & Member, 2009). Oleh karena itu, metode pembelajaran dirancang agar mampu mendorong siswa untuk dapat belajar aktif.

Salah satu metode pembelajaran aktif adalah pembelajaran secara kooperatif. Pembelajaran secara kooperatif merupakan pendekatan proses pembelajaran secara bersama dalam kelompok kecil dengan menerima perbedaan anggota kelompok dan menyelesaikan permasalahan bersama (Anthony S. Dallmann-Jones, 1994). Berdasarkan penelitian-penelitian yang telah dilakukan, metode ini dinilai berhasil meningkatkan proses pembelajaran menjadi lebih efektif.

Salah satunya adalah penelitian yang dilakukan dengan membandingkan mahasiswa teknik elektro yang belajar tanpa metode kooperatif dengan mahasiswa yang belajar menggunakan metode kooperatif. Didapat hasil bahwa mahasiswa yang belajar dengan metode kooperatif dapat mengalami peningkatan pemahaman hingga 84%, sedangkan mahasiswa yang belajar tanpa metode kooperatif hanya mengalami peningkatan sebesar 25% saja (Tahir & Othman, 2010).

Selain itu, penelitian lainnya dilakukan dengan mengujicobakan proses pembelajaran yang melakukan diskusi melalui sistem *computer supported collaborative learning*. Dimana 75% dari siswa yang melakukan proses pembelajaran melalui sistem tersebut menjadi lebih paham dalam menerima materi (Elmahadi & Osman, 2012).

Salah satu metode pembelajaran kooperatif yang populer diimplementasikan adalah metode pembelajaran kooperatif Jigsaw. Hasil implementasi metode Jigsaw pada mahasiswa teknik elektro memberikan hasil yang lebih baik dibandingkan mahasiswa yang belajar secara tradisional (Tahir, Othman, & Yahaya, 2011). Metode Jigsaw ini telah mampu meningkatkan keaktifan siswa selama proses belajar sehingga 91.3% siswanya mengalami peningkatan nilai. Peningkatan nilai dan kesuksesan proses belajar ini salah satunya dipengaruhi oleh komposisi siswa dalam kelompok (Liu, Joy, & Griffiths, 2009).

Pengelompokan siswa pada kelompok yang tepat mampu mengimprovisasi hasil pembelajaran menjadi lebih baik (Henry, 2013). Pada pembelajaran kooperatif Jigsaw formasi yang dibentuk adalah formasi heterogen dan homogen. Heterogen berdasarkan perbedaan kemampuan siswa dan materi diskusi, serta homogen berdasarkan materi diskusi. Formasi dari kelompok tersebut dapat ditentukan

secara acak, ditentukan sendiri oleh siswa atau secara manual oleh guru (Felder & Brent, 2007).

Namun, untuk mendapatkan formasi siswa pada suatu kelompok dengan tepat, tidak dapat dilakukan secara acak (Anthony S. Dallmann-Jones, 1994), (Johnson & Johnson, 1985). Peneliti pada bidang pembelajaran secara kooperatif pun menegaskan bahwa kegagalan pembelajaran kooperatif adalah dikarenakan formasi kelompok yang kurang tepat (Slavin, 1987), (Han & Kamber, 2006). Dibutuhkan formasi yang tepat sesuai kebutuhan. Dapat berdasarkan karakteristik siswa tersebut yang diantaranya: jenis kelamin, sikap, ketertarikan pada materi, motivasi diri, kepercayaan diri, dan performansi belajar (Graf & Bekele, 2006) maupun hal lainnya.

Proses pengelompokan siswa yang dilakukan secara manual dapat memakan waktu banyak terutama pada kelas yang memiliki jumlah siswa banyak (Graf & Bekele, 2006) selain itu formasi kelompok pun belum tentu tepat. Sehingga dibutuhkan sistem pembelajaran kooperatif yang dapat membantu proses pengelompokan siswa dengan tepat secara otomatis dan cepat. Tujuan dari penelitian ini adalah untuk mengetahui algoritma yang paling baik pada kasus pengelompokan siswa.

II. PENELITIAN TERKAIT

Rahel (Henry, 2013) melakukan penelitian dengan melakukan pendekatan untuk pengelompokan siswa berdasarkan atribut-atribut berupa data akademik siswa dan non-akademik. Data tersebut diantaranya adalah kemampuan menggunakan Bahasa Inggris, tingkat motivasi dalam belajar, level motivasi diri, ketertarikan diri pada materi pelajaran dan hubungan dengan teman-teman. Faktor non-akademik diantaranya perbedaan jenis kelamin, umur, keadaan ekonomi orang tua, latar belakang pendidikan orang tua, gaya orang tua dalam mendidik, dan penjelasan mengenai keunikan diri siswa, ditambah dengan atribut nilai siswa. Metode pengelompokan siswa secara otomatis menggunakan *Ant Colony Optimization (ACO)*. Hasil penelitian ini menunjukkan bahwa ACO lebih baik daripada Bayesian.

Penelitian lainnya adalah yang dilakukan Shuangyan Liu, dkk (Liu et al., 2009) yang membangun iGLS (*Intelligent Grouping based on*

Learning Styles). iGLS merupakan sistem yang dibangun untuk mengelompokkan siswa secara otomatis sebagai penunjang pembelajaran online. iGLS mengelompokkan berdasarkan *learning styles* siswa. Dimana siswa dikelompokkan dengan siswa lain yang memiliki *learning styles* yang berbeda. Penelitian ini belum diujicoba secara langsung pada siswa.

III. KAJIAN PUSTAKA

Pada dasarnya, proses pengelompokkan siswa dapat dilakukan menggunakan metode klasifikasi dan clustering. Pada penelitian ini algoritma yang digunakan adalah K-Means, Decision Tree dan Naïve Bayes. Sebelum dilakukan pengelompokkan, tahapan preproses terlebih dahulu dilakukan.

III.1. Cooperative Learning

Cooperative learning atau pembelajaran kooperatif merupakan pembelajaran kolaboratif yang telah dispesifikasikan lebih detail. Pembelajaran kooperatif adalah proses pembelajaran dengan mengikuti beberapa instruksi yang melibatkan siswa bekerja dalam tim untuk mencapai tujuan bersama, di bawah kondisi yang mencakup unsur-unsur yang telah ditentukan (Faust & Paulson, 1998). Terdapat bukti persuasif bahwa tim yang belajar secara kooperatif akan mencapai tingkat pemikiran yang lebih tinggi dibandingkan siswa yang belajar diam-diam secara individu (Johnson & Johnson, 1985). Belajar secara bersama-sama memberikan siswa kesempatan untuk terlibat dalam diskusi, mengambil tanggung jawab untuk pembelajaran mereka dan dengan demikian membentuk pola pikir siswa yang kritis (Totten, 1991). Salah satu metode *cooperative learning* adalah metode Jigsaw. Jigsaw ini dikembangkan oleh Profesor Aronson dan pertama kali digunakan di kelas pada tahun 1971.

III. 2. Learning Style

Salah satu atribut yang digunakan pada *student model* digunakan sebagai *student information* adalah *learning style*. Pemodelan *learning style*/gaya belajar merupakan penilaian yang dapat menjadi masukan bagi seseorang untuk menentukan proses belajar yang tepat untuk digunakan. Pada penelitian ini, *learning style* yang digunakan adalah VAK *learning style*. Dimana *learning style* terdiri dari *auditory*, *visual*, dan *kinaesthetic*. Pada penelitian ini kode *learning style* yang digunakan adalah 2 untuk *visual*, 1 untuk

auditory, 3 untuk *kinaesthetic*. Adapun penjelasan mengenai ketiga gaya belajar tersebut adalah:

1. Seseorang yang termasuk kelompok **auditory** adalah seseorang yang cocok belajar dengan mendengarkan kata yang disebutkan oleh diri sendiri atau orang lain, suara atau musik. Orang-orang pada kelompok ini akan menggunakan frasa ‘ceritakan padaku’, ‘ayo kita berdiskusi’ dan akan memberikan performansi belajar terbaik setelah mendengarkan instruksi dari ahli. Orang-orang ini senang jika diberikan instruksi melalui telepon, dan dapat mengingat semua kata dari lagu yang mereka dengarkan.
2. Seseorang yang termasuk gaya belajar **visual** adalah seseorang yang sebaiknya belajar dengan melihat atau mengobservasi sesuatu, termasuk gambar, diagram, demonstrasi, tampilan, film, atau lainnya. Orang-orang yang termasuk pada kelompok ini akan menggunakan frasa seperti ‘coba perlihatkan’, ‘ayo kita lihat’ dan akan mampu mengeluarkan kemampuan terbaiknya setelah membaca instruksi atau melihat seseorang memperagakan sesuatu. Mereka akan bekerja berdasarkan instruksi dan perintah tertulis.
3. Seseorang yang termasuk kelompok **kinaesthetic** adalah seseorang yang sebaiknya belajar dengan pengalaman fisik seperti menyentuh, merasakan, memegang, mengerjakan, dan berlatih dengan mencoba. Orang-orang pada kelompok ini akan menggunakan frasa ‘biarkan aku mencoba’, ‘bagaimana perasaanmu?’ dan akan memberikan kemampuan terbaiknya dalam mengerjakan sesuatu dengan mencoba langsung, belajar berdasarkan yang mereka coba. Mereka adalah orang yang senang bereksperimen, mencoba, dan tidak melihat pada instruksi dulu.

III.3. Computer Supported Cooperative Learning

Computer supported collaborative learning (CSCL) adalah studi yang berpusat pada makna dan praktik berdasarkan proses kegiatan bersama yang tersimpan (Stahl, Koschmann, & Suthers, 2006). CSCL diawali dengan perkembangan *computer assisted instruction*, *intelligent tutoring system*, *Logo*

as Latin, hingga kemudian berkembang menjadi CSCL pada pertengahan 1990 (Stahl et al., 2006).

Pada penelitian ini, CSCL dilengkapi dengan *intelligent tutoring* dimana sistem memiliki kemampuan cerdas dalam mengetahui kebutuhan siswa. *Intelligent tutoring* merupakan sistem komputer yang mampu menyediakan proses pembelajaran yang adaptif sehingga sesuai dengan siswa secara personal dan efektif (Nouh, P, & Nadarajan, 2007). *Intelligent tutoring* berdiri berdasarkan tiga bidang penting yang mendukungnya, yaitu (Woolf, 2009):

1. *Computer science* terutama pada bidang *artificial intelligence* (AI) berfokus pada bagaimana otomatisasi dari pembelajaran dapat dibangun. Bagaimana sistem dapat mengenali performansi seseorang dalam belajar, memprediksi kegiatan berdasarkan data kegiatan yang sudah lampau, merencanakan aksi pembelajaran personal, dan memberikan timbal balik serta rekomendasi kegiatan pembelajaran.
2. Sisi psikologi khususnya *cognitive science* yang memetakan bagaimana manusia belajar dan mengajar. *Cognitive science* menginvestasikan bagaimana kecerdasan manusia dan komputer berinteraksi dengan lingkungan untuk memperoleh pengetahuan, menyimpan dan menggunakan pengetahuan untuk menentukan penyelesaian masalah.
3. Dasar pendidikan sebagai acuan penting proses pembelajaran terutama yang dilakukan di lingkungan sekolah. Dengan fokus pada dampak yang manusia dapatkan berdasarkan proses mengajar dan belajar, komunikasi, kurikulum, ujian dan motivasi.

Ketiga bidang tersebut dibentuk dalam komponen yang membangun *intelligent tutoring*. Komponen tersebut adalah (Bruyere, 2014):

1. *Expert model*, merepresentasikan komponen pengetahuan ahli dalam menentukan pembelajaran dan penentuan rangking berdasarkan kebiasaan ahli. Dalam penelitian ini,

expert model disesuaikan dengan kurikulum yang berlaku pada pendidikan tingkat pertama.

2. *Learner model* atau *student model*, merepresentasikan pengetahuan siswa atau keadaan pengetahuan siswa saat ini. *Student model* dapat dideskripsikan terdiri dari:
 - a. *Personal information* (nama siswa, jenis kelamin, umur, *learning style*, dsb)
 - b. *Real-time information* (waktu belajar, jenis materi yang digunakan, kecepatan dalam mengerjakan, dsb)
 - c. *Status information* (kemampuan siswa, status, hal yang disukai)
 - d. *Learning-history information* (log belajar sebelumnya)
3. *Pedagogy model*, merepresentasikan komponen aturan penentu proses pembelajaran dan penilaian kemampuan siswa. Dalam penelitian ini, *pedagogy model* yang digunakan adalah pembelajaran kooperatif dengan pendekatan metode Jigsaw.
4. *Domain knowledge*, berupa format metadata yang mengindikasikan atribut dari element-elemen pembelajaran, seperti level kesulitan soal, materi pembelajaran, atau atribut lainnya.

III.4. Preproses

Tahapan-tahapan pada preproses data tersebut ditujukan untuk dapat menyediakan data yang lebih berkualitas dan menyediakan kebutuhan atribut yang tepat (Romero, Ventura, Pechenizkiy, & Baker, 2011). Preproses data penting untuk dilakukan karena *real data* yang akan diproses seringkali tidak lengkap. Preproses data digunakan untuk menyelesaikan permasalahan data yang hilang dan *outliers*. Pada data pendidikan, data yang hilang sering terjadi akibat siswa tidak melengkapi kuesioner atau tidak menjawab pertanyaan. Sedangkan *outliers* diakibatkan karena data tidak sesuai dengan model secara keseluruhan. Preproses pada penelitian ini adalah

1. Feature Extraction dengan *discretization*, nilai dari data numerik diubah menjadi data interval yang akan menjadi data baru.

- Contohnya adalah dengan mengubah nilai 82.5 menjadi 'lulus'. Hal ini dibutuhkan untuk mengurangi besarnya range data dari setiap poin nilai yang memungkinkan dan sebenarnya tidak terlalu berpengaruh.
2. *Data cleaning* berguna untuk mendeteksi dan membuang data yang tidak lengkap, *outliers*, dan tidak konsisten guna mengimprovisasi kualitas data yang digunakan (Han & Kamber, 2006).
 3. Upaya yang juga digunakan untuk mengurangi besar dimensi data adalah dengan *feature selection*. *Feature selection* merupakan proses analisa untuk mengetahui data yang memiliki korelasi paling tinggi dan paling berpengaruh dalam menentukan hasil (Romero et al., 2011). Metode yang banyak digunakan diantaranya adalah *information gain*. *Information gain* mampu memberikan informasi seberapa penting atribut pada sebuah keputusan dengan mencari nilai terbesar dari nilai *information gain* sebuah atribut.
 4. *Data transformation*. *Data transformation* dilakukan untuk mengubah data agar sesuai untuk dilakukan proses *mining*. Proses *data transformation* yang dilakukan adalah:
 1. *Aggregation*, proses menyimpulkan data. Sebagai contoh, menarik kesimpulan dari nilai harian menjadi nilai semester.
 2. *Generalization*, proses menggeneralisasikan data pada hirarki yang lebih tinggi. Sebagai contoh, nilai afektif, nilai kognitif dan nilai psikomotor pada mata pelajaran matematika digeneralisasikan secara keseluruhan menjadi nilai mata pelajaran matematika.

III.5. K-Means

K-Means merupakan salah satu algoritma clustering yang termasuk dalam algoritma *unsupervised learning* (Wu, 2009). K-means akan mengelompokkan data menjadi k kelompok yang diinginkan. Dimana pada setiap kelompok atau *cluster* akan terdapat titik pusat (centroid).

Algoritma K-means adalah sebagai berikut (Han & Kamber, 2006):

1. Pilih K buah titik *centroid* secara acak
2. Kelompokkan data sehingga terbentuk K buah *cluster* dengan titik *centroid* dari setiap *cluster* merupakan titik *centroid* yang telah dipilih sebelumnya
3. Perbaharui nilai titik *centroid*
4. Ulangi langkah 2 dan 3 sampai nilai dari titik *centroid* tidak lagi berubah

Data yang menjadi anggota tiap *cluster* dihitung berdasarkan jarak terdekat data tersebut pada titik *centroid*. Rumus untuk menghitung jarak tersebut adalah (Maimon & Rokach, 2010):

$$d(x_i, x_j) = (|x_{i1} - x_{j1}| + \dots + |x_{ip} - x_{jp}|^g)^{1/g}$$

Dimana,

- g = 1, untuk menghitung jarak Manhattan
- g = 2, untuk menghitung jarak Euclidean
- g = ∞, untuk menghitung jarak Chebychev

x_i, x_j adalah dua buah data yang akan dihitung jaraknya

p = dimensi dari sebuah data

Pembaruan suatu titik centroid dapat dilakukan dengan rumus berikut:

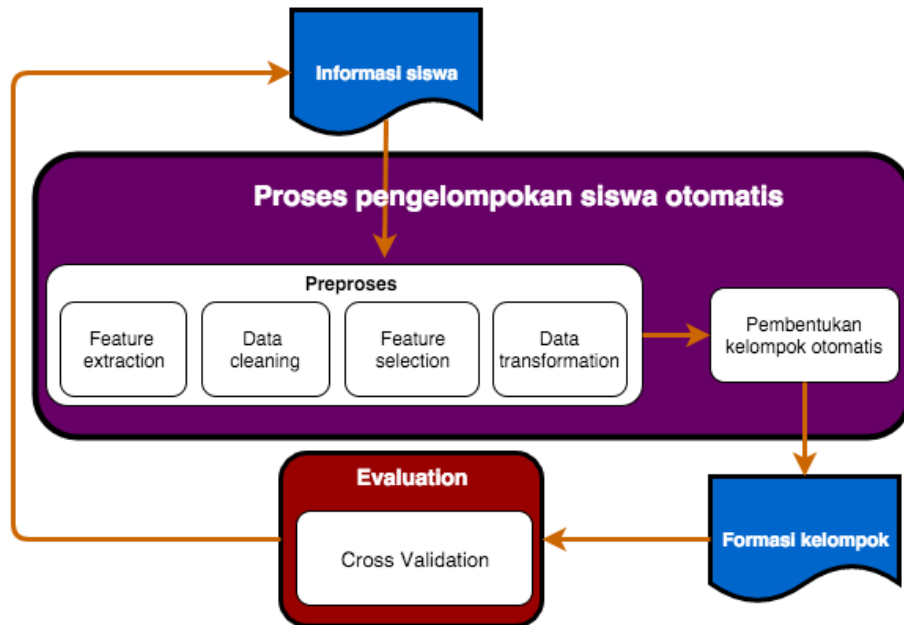
$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q$$

Dimana,

- μ_k = titik centroid dari cluster ke-K
- N_k = banyaknya data pada cluster ke-K
- x_q = data ke-q pada cluster ke-K

III.6. Decision Tree

Decision Tree merupakan algoritma yang termasuk dalam *supervised learning*. Dikatakan *supervised learning* karena pada proses pengelompokan algoritma ini menggunakan data awal untuk membentuk aturan keputusan. Sehingga, data yang telah dikumpulkan harus sudah memiliki label kelompok.



Gambar 1. Proses pengelompokan siswa otomatis

Algoritma decision tree adalah sebagai berikut (Han & Kamber, 2006):

1. Pilih atribut sebagai akar
2. Buat cabang untuk masing-masing nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sesuai

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Rumus untuk menghitung gain adalah sebagai berikut:

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \text{Entropy}(S_i)$$

Dimana,

S = Himpunan semesta data

A = Atribut

N = Jumlah partisi atribut A

$|S_i|$ = Jumlah kasus pada partisi ke-i

$|S|$ = Jumlah kasus dalam S

III.7. Naïve Bayes

Seperti halnya Decision Tree, algoritma Naïve Bayes termasuk dalam algoritma *supervised learning*.

Dimana membutuhkan data latih yang sudah memiliki label untuk dijadikan data awal pembentukan model.

Teorema Bayes memiliki bentuk umum sebagai berikut (Han & Kamber, 2006):

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Dimana,

X = data dengan kelas yang belum diketahui

H = hipotesis data X merupakan suatu kelas spesifik

$P(H|X)$ = probabilitas hipotesis H berdasarkan kondisi x (posteriori probabilistik)

$P(H)$ = probabilitas hipotesis H (prior probabilistik)

$P(X|H)$ = probabilitas X berdasarkan kondisi tersebut

$P(X)$ = probabilitas dari X

IV. METODE PENELITIAN

Metode penelitian yang digunakan pada penelitian ini adalah eksperimen. Dimana data penelitian yang digunakan adalah data siswa SMPN 2 Bandung kelas VIII-I, VIII-J, VIII-K tahun ajaran 2014/2015. Total siswa adalah 103 orang dengan

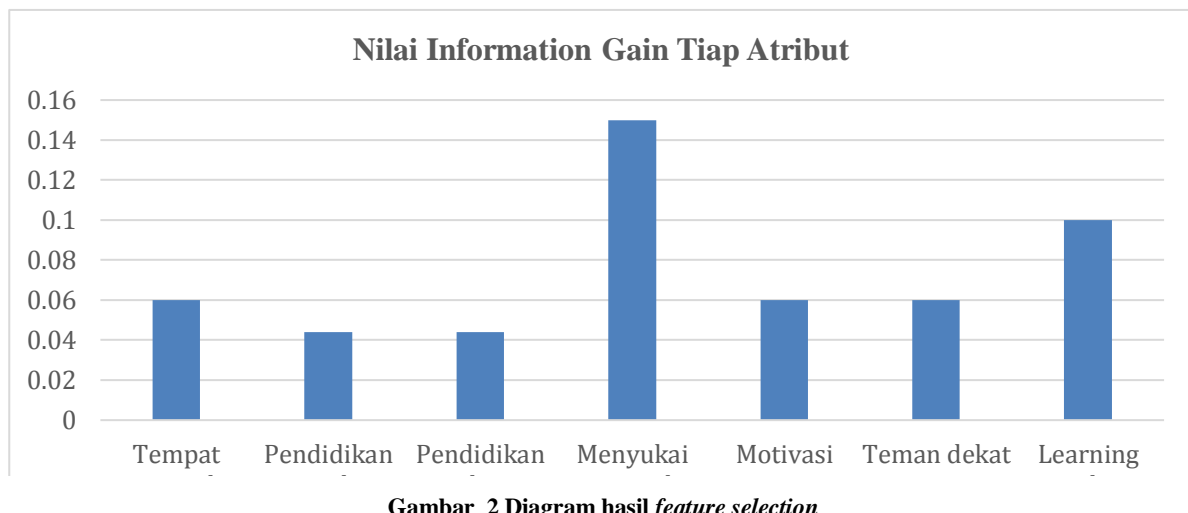
mengambil data selama semester genap pada mata pelajaran IPA.

Data siswa yang digunakan pada penelitian ini berupa faktor akademik dan faktor non-akademik. Faktor akademik berupa nilai kognitif siswa, *learning style* dan tingkat kesukaan siswa pada mata pelajaran tersebut. Faktor non-akademik berupa jenis kelamin, umur, latar belakang pendidikan orang tua, dan keadaan ekonomi orang tua. Seperti dilihat pada Gambar 1, metode penelitian dilakukan dengan mengumpulkan data siswa, data siswa dipreproses, kemudian pengelompokkan siswa secara otomatis dan *evaluation* data menggunakan *cross validation* dengan membandingkan hasil pengelompokkan siswa yang ditentukan oleh guru dengan hasil pengelompokkan yang ditentukan oleh hasil algoritma.

Cara pengambilan data dilakukan dengan siswa mengisi form yang terdapat pada sistem CSCL. Dari data yang dikumpulkan terdapat ketidaklengkapan data. Hal tersebut terjadi dikarenakan siswa tidak mengisi form pada sistem CSCL dengan lengkap. Untuk menangani hal tersebut, data siswa yang tidak lengkap diisi dengan nilai 0 jika berupa data numeric dan null jika data berupa string.

3. *Feature selection*

Feature selection dilakukan dengan metode *information gain*. Implementasi penghitungan *information gain* diterapkan pada beberapa atribut yang diambil dari siswa sebagai sampel awal. Adapun atribut data siswa yang digunakan adalah: *tempat tinggal, pendidikan orang tua, tingkat kesukaan pada mata pelajaran, motivasi, teman dekat, learning style (visual, audio,*



Gambar 2 Diagram hasil *feature selection*

Pada preproses yang dilakukan adalah sebagai berikut:

1. *Feature extraction*

Pada penelitian ini, *feature extraction* dilakukan dengan metode *discretization*. Dimana nilai numerik diubah menjadi data baru berupa interval. Data yang dikenakan *discretization* adalah data nilai siswa. Data nilai siswa diubah menjadi keterangan Baik dan Kurang. Siswa diberi keterangan Baik (B) jika nilainya antara nilainya lebih besar dari 70 dan diberikan keterangan Kurang (K) jika nilainya dibawah 70.

2. *Data cleaning*

kinestetik). Didapat hasil seperti terlihat pada Gambar 2. Disimpulkan bahwa atribut yang memiliki pengaruh pada kesuksesan *cooperative learning* adalah kelompok terdiri dari perbedaan penilaian suka dan tidak sukanya siswa terhadap materi dan jenis *learning style* siswa.

4. *Data transformation*

Pada data siswa yang dikumpulkan terdapat data nilai tugas dan nilai ulangan harian siswa selama satu semester. Untuk dapat diproses, data-data nilai tersebut ditransformasi secara *aggregation*.

Sehingga, data nilai tugas dan nilai ulangan harian siswa selama satu semester tersebut diubah menjadi nilai semester siswa. Selain itu, data pendidikan ayah dan pendidikan ibu ditransformasi menjadi pendidikan tinggi atau pendidikan dasar. Serta, pekerjaan ayah dan ibu yang ditransformasi menjadi ya atau tidak.

Setelah preproses, dilakukan pengelompokkan secara tradisional oleh guru dan secara otomatis menggunakan K-means, Decision Tree dan Naïve Bayes. Simulasi pengelompokkan ini menggunakan alat bantu WEKA.

V. HASIL ANALISIS

Siswa dikelompokkan secara tradisional oleh guru berdasarkan hasil pengamatan guru terhadap siswa. Masing-masing kelas dibagi menjadi 7 kelompok, dimana per kelompok terdiri dari 4-5 siswa. Hasil pengelompokkan guru kemudian dijadikan label kelas untuk setiap siswa. Setelah preproses, data kemudian disesuaikan dengan format .arff agar dapat diproses dengan alat bantu WEKA.

Penelitian ini menggunakan data yang sama dengan 3 macam algoritma, yaitu Naïve Bayes, K-means, dan Decision Tree. Dari tabel 1 dapat dilihat nilai akurasi algoritma Naïve bayes adalah 70,37%, Decision Tree 59,26% dan K-means 26%.

Tingkat kesalahan pengelompokkan yang paling besar terjadi pada K-means. *Cluster* yang terbentuk ternyata hanya 2 *cluster*. Dapat diambil keputusan bahwa K-means tidak sesuai untuk kasus pengelompokkan siswa. *Cluster* yang terbentuk oleh K-means membutuhkan tahapan selanjutnya. Karena, *cluster* tersebut terdiri dari siswa-siswa yang memiliki kesamaan. Sehingga setelah didapat hasil *cluster* K-means, sistem harus kembali mengambil siswa secara adil dari masing-masing *cluster* menjadi n kelompok hasil bagi seluruh siswa dengan banyak siswa per kelompok. Hasil akurasi yang lebih baik dari K-means adalah Decision Tree sebesar 59,26%. Berdasarkan hasil *tree* yang dihasilkan atribut nilai menjadi atribut yang paling berpengaruh terhadap proses pengelompokkan siswa. Meskipun nilai akurasi Decision Tree lebih baik, namun waktu yang dibutuhkan untuk membuat *tree* lebih lama dibandingkan K-means.

Tabel 1 Hasil perbandingan algoritma

	Naïve bayes	K-means	Decision Tree
Correctly classified	70,37%	26%	59,26%
Incorrectly classified	29,63%	74%	40,74%
Time build model	0 detik	0 detik	0,08 detik

Hasil akurasi tertinggi pengelompokkan siswa adalah menggunakan Naïve Bayes sebesar 70,37%. Diantara ketiga algoritma yang dibandingkan, Naïve Bayes menjadi algoritma terbaik dalam tingkat akurasi dan waktu yang dibutuhkan dalam proses pengelompokkan. Naïve Bayes selain mudah diimplementasi, juga mampu memberikan tingkat akurasi yang cukup baik.

VI. KESIMPULAN DAN SARAN

Pengelompokkan siswa merupakan hal penting dalam pembelajaran kooperatif. Mengelompokkan siswa menjadi masalah karena ada banyak atribut siswa yang harus dipertimbangkan dalam menentukan kelompok yang tepat. Pada penelitian ini, kelompok yang tepat merupakan kelompok yang memiliki keragaman atribut siswa dalam kelompok. Pertimbangan guru dalam mengelompokkan siswa menjadi acuan akurasi dalam mengelompokkan siswa secara otomatis. Uji coba pengelompokkan secara otomatis menggunakan algoritma learning, diantaranya Naïve Bayes, K-means, dan Decision Tree. Pembelajaran kooperatif dengan menggunakan JIGSAW mengharuskan kelompok yang dibangun terdiri dari siswa yang memiliki keberagaman (heterogen) dan antar kelompoknya memiliki kesamaan (homogen).

Hasil uji coba terhadap tiga algoritma menunjukkan bahwa Naïve Bayes mampu mengelompokkan siswa dengan tingkat akurasi 70,37% seperti guru mengelompokkan dan waktu yang dibutuhkan untuk mengelompokkan tercepat diantara K-means dan Decision Tree. Jika menggunakan metode Naïve Bayes, meskipun metode ini memberikan nilai akurasi paling tinggi, namun metode ini tidak dapat mengukur tingkat

keheterogenan dalam kelompok dan tingkat homogen antar kelompok.

Oleh karena itu, penelitian lanjutan yang dapat dilakukan adalah pertama, perbaikan penggunaan metode Naïve Bayes yang digunakan untuk mengelompokkan siswa secara otomatis agar dapat mengukur tingkat ragam siswa dalam kelompok (heterogen) dan tingkat pemerataan siswa antar kelompok (homogen). Kedua, pengukuran kepuasan siswa terhadap hasil pengelompokkan dan nilai setelah siswa melakukan proses pembelajaran secara berkelompok menggunakan pretest dan posttest. Ketiga, penelitian selanjutnya dapat melibatkan siswa lebih banyak dan perbedaan atribut yang digunakan pada setiap mata pelajaran untuk proses pengelompokkan.

Diharapkan dengan adanya perbaikan dalam metode yang digunakan dapat memberikan rekomendasi kelompok bagi siswa. Dimana kelompok yang dibangun terdiri dari keberagaman siswa dalam kelompok yang merata antar kelompoknya. Sehingga, siswa dapat menemukan kelompok belajar yang mampu meningkatkan kemampuannya.

REFERENSI

- Anthony S. Dallmann-Jones. (1994). *The Expert Educator: A Reference Manual of Teaching Strategies for Quality Education*. Three Blue Herons Pub.
- Bruyere, H. (2014). *Intelligent Tutoring System*. US.
- Elmahadi, I. O., & Osman, I. (2012). Perceptions towards Computer Supported Collaborative Learning: A Case Study of Sudanese Undergraduate Students (pp. 158–161).
- Faust, J. L., & Paulson, D. R. (1998). Active Learning in the College Classroom, 9, 3–24.
- Felder, R. M., & Brent, R. (2007). *Cooperative Learning* *.
- Graf, S., & Bekele, R. (2006). Forming Heterogeneous Groups for Intelligent Collaborative Learning Systems with Ant Colony Optimization. In *Intelligent Tutoring Systems* (Vol. 4053, pp. 217 – 226). Springer Berlin Heidelberg.
- Han, J., & Kamber, M. (2006). *Data Mining* (Second Edi). San Francisco: Morgan Kaufmann Publishers.
- Henry, T. R. (2013). Creating Effective Student Groups: An Introduction to groupformation . org. In *Proceeding of the 44th ACM technical symposium on Computer science education* (pp. 645–650). Denver, CO, USA.
- Hoic-bozic, N., Mornar, V., Boticki, I., & Member, S. (2009). A Blended Learning Approach to Course Design and Implementation, 52(1), 19–30.
- Johnson, D. W., & Johnson, R. T. (1985). *The Internal Dynamics of Cooperative Learning Groups*. USA: Springer US.
- Liu, S., Joy, M., & Griffiths, N. (2009). iGLS: Intelligent Grouping for Online Collaborative Learning. *2009 Ninth IEEE International Conference on Advanced Learning Technologies*, 364–368. <http://doi.org/10.1109/ICALT.2009.41>
- Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook* (Second Edi). London: Springer Berlin Heidelberg.
- Nouh, Y., P, K., & Nadarajan, R. (2007). Intelligent Tutoring System-Bayesian Student Model. *2006 1st International Conference on Digital Information Management*, 257–262. <http://doi.org/10.1109/ICDIM.2007.369362>
- Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. (Eds.). (2011). *Handbook of Educational Data Mining* (Second Ed). London: CRC Press.
- Shen, R., Wang, M., Gao, W., Novak, D., & Tang, L. (2009). Mobile Learning in a Large Blended Computer Science Classroom: System Function, Pedagogies, and Their Impact on Learning. *IEEE Transactions on Education*, 52(4), 538–546. <http://doi.org/10.1109/TE.2008.930794>
- Slavin, R. E. (1987). Development and Motivational Perspective on Cooperative Learning: A Reconciliation. *Special Issue on Schools and Development*, 58(5), 1161–1167.
- Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective, 409–426.

- Tahir, N. M., & Othman, K. A. (2010). The jigsaw cooperative method amongst electrical engineering students. In *2010 2nd International Congress on Engineering Education* (pp. 229–233). Ieee. <http://doi.org/10.1109/ICEED.2010.5940797>
- Tahir, N. M., Othman, K. A., & Yahaya, F. H. (2011). Case study of jigsaw cooperative learning effect within Electrical Engineering courses. In *2011 International Conference on Business, Engineering and Industrial Applications* (pp. 20–23). Ieee. <http://doi.org/10.1109/ICBEIA.2011.5994245>
- Woolf, B. P. (2009). *Building Intelligent Interactive Tutors for revolutionizing e-learning*. USA.
- Wu, Z. (2009). Identification and prioritization of variables that influence learning content development in e-learning platforms. *2009 2nd IEEE International Conference on Computer Science and Information Technology*, 444–449. <http://doi.org/10.1109/ICCSIT.2009.5234912>