PENENTUAN JURUSAN SISWA SEKOLAH MENENGAH ATAS DISESUAIKAN DENGAN MINAT SISWA MENGGUNAKAN ALGORITMA FUZZY C-MEANS

by Moch Arief Soeleman

Submission date: 06-May-2020 09:41PM (UTC+0700)
Submission ID: 1317542986
File name: 9-Article_Text-23-1-10-20171123.pdf (162.25K)
Word count: 5232
Character count: 32614

PENENTUAN JURUSAN SISWA SEKOLAH MENENGAH ATAS DISESUAIKAN DENGAN MINAT SISWA MENGGUNAKAN ALGORITMA *FUZZY C-MEANS*

Altanova Reza[1], Abdul Syukur[2], M. Arief Soeleman[3]
[123]Pascasarjana Teknik Informatika Universitas Dian Nuswantoro

ABSTRACT

Majors are held no valid high school in Indonesia is conducted when students are still in grade X. This includes the areas of interest Majors Natural Sciences, Social Sciences, and science of language. Majors will depend on the capability of student achievement in the areas of interest / courses available and in accordance with the conditions at the school. If it is not possible then the only department of particular interest are provided in school. The results of tests that tested students' interest through psychological tests aimed to help the school and the students themselves so that later, the lessons will be given to students become more focused as it has in accordance with the capability in the field of interest. Fuzzy C-Means algorithm is an algorithm that is easy and is often used in the technique of grouping the data as it takes an estimate efficient and does not require a lot of parameters. Several studies have concluded that the Fuzzy C-Means algorithm can be used to classify data based on certain attributes in this study will be used Fuzzy C-Means algorithm to classify the student data High School (SMA) based on the value of the core subjects for the majors that are appropriated to the interests test results. The study also examined the level of accuracy of Fuzzy C-Means algorithm in determining the majors in high school. Application of Fuzzy C-Means algorithm in determining the majors in the 278 high school students were tested in this study, indicating that the FCM algorithm has a good degree of accuracy (in an average of 82.01%) by including interest test scores compared with the manual method based on the selection of individual students only 63.67%.
Keywords: clustering, Majors Students, Accuracy Fuzzy C-Means

1. PENDAHULUAN

Sesuai kurikulum yang berlaku di seluruh Indonesia, maka siswa kelas X SMA yang naik ke kelas XI akan mengalami pemilihan jurusan / penjurusan. Penjurusan yang tersedia di SMA meliputi Ilmu Alam (IPA), Ilmu Sosial (IPS), dan Ilmu Bahasa. Penjurusan akan disesuaikan dengan minat dan kemampuan siswa. Tujuannya agar kelak di kemudian hari, pelajaran yang akan diberikan kepada siswa menjadi lebih terarah karena telah sesuai dengan minatnya. Sebelum waktu penjurusan, guru BK/BP telah melakukan psikotes sehingga potensi siswa secara psikologis lebih dapat lebih tergali dan penjurusan yang akan dilakukan tidak salah arah[1].

Minat merupakan suatu kegiatan yang dilakukan oleh siswa secara tetap dalam melakukan proses belajar. Sesuai dengan pendapat Slameto (2010: 57) minat adalah kecenderungan yang tetap untuk memperhatikan dan mengenang beberapa kegiatan. Kegiatan yang diminati siswa, diperhatikan terus-menerus yang disertai rasa senang dan diperoleh rasa kepuasan. Lebih lanjut dijelaskan minat adalah suatu rasa suka dan ketertarikan pada suatu hal atau aktivitas, tanpa ada yang menyuruh. Seseorang yang memiliki minat terhadap kegiatan tertentu cenderung memberikan perhatian yang besar terhadap kegiatan tersebut[3].

Kegiatan penjurusan di sekolah dilaksanakan untuk menyeleksi, memilih dan mengumpulkan siswa atau peserta didik yang memiliki kemampuan yang sama untuk mengikuti satu program pendidikan (jurusan) yang sama. Penentuan jurusan berdasarkan minat siswa sangat perlu dilaksanakan agar diperoleh hasil belajar siswa yang baik, sesuai dengan kurikulum yang diselenggarakan khususnya di Sekolah Menengah Atas.

Penulis bermaksud untuk meneruskan penelitian yang telah dilakukan oleh peneliti sebelumnya (Bahar, 2011, Penentuan Jurusan Sekolah Menengah Atas dengan Algoritma *Fuzzy C-Means*, Program

Pasca Magister TI UDINUS Semarang). Bahar menyarankan agar peneliti berikutnya menambahkan variabel minat yang didapat dari test psikologi untuk dijadikan variabel komputasi dengan *Fuzzy C-Means*[4]. Karena di beberapa sekolah sudah ada yang menambahkan variabel minat untuk menentukan jurusan dan juga di dukung oleh pendapat peneliti lain bahwa varibel minat akan berpengaruh pada kesungguhan belajar peserta didik di sekolah.[3] Pada penelitian ini penulis menggunakan data yang berbeda dengan peneliti sebelumnya, dan juga melakukan perubahan terhadap atribut-atribut data yang digunakan karena menyesuaikan dengan kondisi sekolah tempat penulis mengambil data. Akan tetapi, metode pengujian algoritma *Fuzzy C-Means* yag digunakan tetap mengacu pada peneliti sebelumnya.

Banyak metode (algortima) yang dapat digunakan untuk penentuan konsentrasi jurusan, salah satunya adalah dengan menggunakan konsep *Fuzzy Cluster Means* (FCM) yakni dengan cara mengelompokkan data nilai siswa menurut kemiripan yang dimilikinya. Metode ini dipilih karena kemampuannya untuk melakukan pengelompokan suatu objek/data yang belum memiliki klasifikasi, ke dalam kelas tertentu menurut kesamaan yang dimilikinya berdasarkan derajat keanggotaan dengan cara minimalisasi nilai fungsi objektifnya[4].

Dengan demikian tujuan penelitian ini menggunakan algoritma *Fuzzy C-Means* diharapkan dapat membantu untuk menentukan penjurusan siswa Sekolah Menengah Atas menjadi lebih akurat dan dapat diterapkan di Indonesia, terutama di bumi Kalimantan Selatan.

Hasil penelitian ini diharapkan dapat dapat bermanfaat dalam memberikan gambaran dan pemahaman penerapan algoritma *Fuzzy C-Means* pada studi kasus penentuan jurusan siswa Sekolah Menengah Atas yang disesuaikan dengan minat siswa.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terkait

2.1.1 Penentuan Jurusan SMA dengan *Fuzzy C-Means*

Algoritma *Fuzzy C-Means* merupakan satu algoritma yang mudah dan sering digunakan di dalam teknik pengelompokan data karena membuat suatu perkiraan yang efsien dan tidak memerlukan banyak parameter. Beberapa penelitian telah menghasilkan kesimpulan bahwa algoritma *Fuzzy C-Means* dapat dipergunakan untuk mengelompokkan data berdasarkan atribut-atribut tertentu. Pada penelitian ini akan digunakan algoritma *Fuzzy C-Means* untuk mengelompokkan data siswa Sekolah Menengah Atas berdasarkan Nilai mata pelajaran inti untuk proses penjurusan. Penelitian ini juga menguji tingkat akurasi algoritma *Fuzzy C-Means* dalam penentuan jurusan pada Sekolah Menengah Atas.

Penerapan algoritma *Fuzzy C-Means* dalam penentuan jurusan di Sekolah Menengah Atas pada 81 sampel data siswa yang diuji dalam penelitian ini menunjukkan bahwa Algoritma *Fuzzy C-Means* memiliki tingkat akurasi yang lebih tinggi (rata-rata 78,39%), jika dibandingkan dengan metode penentuan jurusan secara manual yang selama ini dilakukan (hanya memiliki tingkat akurasi rata-rata 56,17 %) [4]

2.1.2 Penggunaan Sistem Inferensi *Fuzzy* untuk Penentuan Jurusan di SMA Negeri 1 Bireuen

Dalam kegiatan akademik siswa SMA, penentuan jurusan sangat diperlukan untuk membantu siswa lebih fokus terhadap kemampuan yang telah dimiliki. Keputusan penentuan jurusan dibuat oleh pihak yang berkompeten di sekolah. Salah satu aplikasi logika *fuzzy* adalah pendukung keputusan dengan Sistem Inferensi *Fuzzy* Mamdani. Dalam Sistem Inferensi *Fuzzy* Mamdani untuk memperoleh output diperlukan empat tahap, yaitu pembentukan himpunan *fuzzy*, pembentukan *rule*s, aplikasi fungsi implikasi dan inferensi aturan serta defuzzifikasi.

Penelitian ini membangun Sistem Inferensi *Fuzzy* Mamdani dalam penentuan jurusan di SMA N 1 Bireuen. Variabel inputnya adalah NIPA, NIPS, IQ, Minat dan kapasitas kelas. Variabel outputnya adalah IPA dan IPS. Dari pengujian data output, diperoleh nilai output IPA dan IPS untuk Sistem Inferensi *Fuzzy*. Dari percobaan yang dilakukan terhadap data siswa kelas X [5].

2.1.3 *Clustering* Data Nilai Mahasiswa untuk Pengelompokan Konsentrasi Jurusan Menggunakan *Fuzzy Cluster Means*

Pada penelitian ini dibahas mengenai implementasi metode *Fuzzy Cluster-Means* terhadap data-data bobot nilai mahasiswa pada matakuliah tertentu untuk memperoleh pengelompokan konsentrasi jurusan

menurut perolehan nilai akademik berdasarkan kemiripan yang dimilikinya. Analisa data dilakukan dengan bantuan aplikasi MATLAB untuk pembentukan *cluster* data yang sesuai dengan yang diharapkan. Hasil penelitian ini berupa tiga buah data *cluster* yang bisa digunakan untuk pendukung keputusan terhadap penentuan konsentrasi dan dari 126 data mahasiswa, diperoleh sebanyak 28 mahasiswa masuk ke dalam *cluster* 1 (multimedia), 70 mahasiswa pada *cluster* 2 (web) dan 28 mahasiswa pada *cluster* 3 (pemrograman). (Tb. Ai Munandar, Wahyu Oktri Widyarto, Harsiti, 2013.) [6]

2.2. Landasan Teori

2.2.1 Konsep *Clustering* dalam *Data mining*

Konsep dasar *data mining* adalah menemukan informasi tersembunyi dalam sebuah basis data dan merupakan bagian dari *Knowledge Discovery in Databased* (KDD) untuk menemukan informasi dan pola yang berguna dalam data [7]. *Data mining* mencari informasi baru, berharga dan berguna dalam sekumpulan data dengan melibatkan komputer dan manusia serta bersifat iteratif baik melalui proses ing otomatis ataupun manual. Secara umum sifat *data mining* adalah:

a. *Predictive*: menghasilkan model berdasarkan sekumpulan data yang dapat digunakan untuk memperkirakan nilai data yang lain. Metode yang termasuk dalam prediktif *data mining* adalah:
   1) Klasifikasi: pembagian data ke dalam beberapa kelompok yang telah ditentukan sebelumnya
   2) Regresi: memetakan data ke suatu *prediction variable*
   3) Time Series Analisys: pengamatan perubahan nilai atribut dari waktu ke waktu

b. *Descriptive*: mengidentifikasi pola atau hubungan dalam data untuk menghasilkan informasi baru. Metode yang termasuk dalam *Descriptive Data mining* adalah:
   1) *Clustering* : identifikasi kategori untuk mendeskrIPSikan data
   2) *Association Rules*: Identifikasi hubungan antar data yang satu dengan yang lainnya.
   3) *Summarization*: pemetaan data ke dalam subset dengan deskrpsi sederhana.
   4) *Sequence Discovery*: identifikasi pola sekuensial dalam data.

Clustering membagi data menjadi kelompok-kelompok atau *cluster* berdasarkan suatu kemiripan atribut-atribut diantara data tersebut [7]. Karakteristik tiap *cluster* tidak ditentukan sebelumnya, melainkan tercermin dari kemiripan data yang terkelompok di dalamnya. Oleh sebab itu hasil *clustering* seringkali perlu diinterprestasikan oleh pihak-pihak yang benar-benar mengerti mengenai karakter domain data tersebut. Selain digunakan sebagai metode yang independen dalam *data mining*, *clustering* juga digunakan dalam pra-pemrosesan data sebelum data diolah dengan metode *data mining* yang lain untuk meningkatkan pemahaman terhadap domain data.

Karakteristik terpenting dari hasil *clustering* yang baik adalah suatu *instance* data dalam suatu *cluster* lebih ”mirip” dengan *instance* lain di dalam *cluster* tersebut daripada dengan *instance* di luar dari *cluster* itu [8]. Ukuran kemiripan (similarity measure) tersebut bisa bermacam-macam dan mempengaruhi perhitungan dalam menentukan anggota suatu *cluster*. Jadi tipe data yang akan di-*cluster* (kuantitatif atau kualitatif) juga menentukan ukuran apa yang tepat digunakan dalam suatu algoritma. Selain kemiripan antar data dalam suatu *cluster*, *clustering* juga dapat dilakukan berdasarkan jarak antar data atau *cluster* yang satu dengan yang lainnya. Ukuran jarak (*distance* atau *dissimilarity measure*) yang merupakan kebalikan dari ukuran kemiripan ini juga banyak ragamnya dan penggunaannya juga tergantung pada tipe data yang akan di-*cluster*. Kedua ukuran ini bersifat simetris, dimana jika A dikatakan mirip dengan B maka dapat disimpulkan bahwa B mirip dengan A.

Ada beberapa macam rumus perhitungan jarak antar*cluster*. Untuk Tipe data numerik, sebuah data set X beranggotakan $x1 \in X$, $i = 1, ..., n$, tiap item direpresentasikan sebagai vektor X1= {Xi1, Xi2, Xim} dengan m sebagai jumlah dimensi dari item. Rumus-rumus yang biasa digunakan sebagai ukuran jarak antara Xi dan Xj untuk data numerik ini antara lain:

a. *Euclidean Distance*

$$\left(\sum_{i=0}^{n}\left(x_{ik}-x_{ij}\right)^{2}\right)^{\frac{1}{2}} \quad \text{.....(1)}$$

Ukuran ini sering digunakan dalam *clustering* karena sederhana. Ukuran ini memiliki masalah jika skala nilai atribut yang satu sangat besar dibandingkan nilai atribut lainnya. Oleh sebab itu, nilai-nilai atribut sering dinormalisasi sehingga berada dalam kisaran 0 dan 1.

b. *City Block Distance* atau *Manhattan Distance*

$$\sum_{i=0}^{n}\left|x_{ik}-x_{jk}\right| \quad \text{.....(2)}$$

Jika tiap item digambarkan sebagai sebuah titik dalam grid, ukuran jarak ini merupakan banyak sisi yang harus dilewati suatu titik untuk mencapai titik yang lain seperti halnya dalam sebuah peta jalan.

c. *Minkwoski Metric*

$$\left(\sum_{i=0}^{n}\left(x_{ik}-x_{ij}\right)^{p}\right)^{\frac{1}{p}} \quad \text{.....(3)}$$

Ukuran ini merupakan bentuk umum dari *Euclidean Distance* dan *Manhattan Distance*. *Euclidean Distance* adalah kasus dengan nilai p=2 sedangkan *Manhattan Distance* merupakan bentuk *Minkwoski* dengan p=1. Dengan demikian, lebih banyak nilai numerik yang dapat ditempatkan pada jarak terjauh di antara 2 vektor. Seperti pada *Euclidean Distance* dan juga *Manhattan Distance*, ukuran ini memiliki masalah jika salah satu atribut dalam vektor memiliki rentang yang lebih besar dibandingkan atribut-atribut lainnya.

d. *Cosine – Corelation* (ukuran kemiripan dari model *Euclidean* n-dimensi)

$$\frac{\sum_{k=1}^{m}\left(x_{ik}.x_{jk}\right)}{\sqrt{\sum_{k=0}^{m}x_{ik}^{2}.x_{jk}^{2}}} \quad \text{.....(4)}$$

Ukuran ini bagus digunakan pada data dengan tingkat kemiripan tinggi walaupun sering pula digunakan bersama pendekatan lain untuk membatasi dimensi dari permasalahan.

Dalam mendefinisikan ukuran jarak antar-*cluster* yang digunakan beberapa algoritma untuk menentukan *cluster* yang terdekat, perlu dijelaskan mengenai atribut-atribut yang menjadi referensi dari suatu *cluster* [8]. Untuk suatu *cluster* Km berisi N item {Xm1, Xm2, ..., Xmn}.

e. *Centroid*: suatu besaran yang dihitung dari rata-rata nilai dari setiap item dari suatu *cluster* menurut rumus:

$$C_m = \frac{\sum_{i=1}^{n}|x_{mi}|}{N} \quad \text{.....(5)}$$

f. *Medoid*: item yang letaknya paling tengah

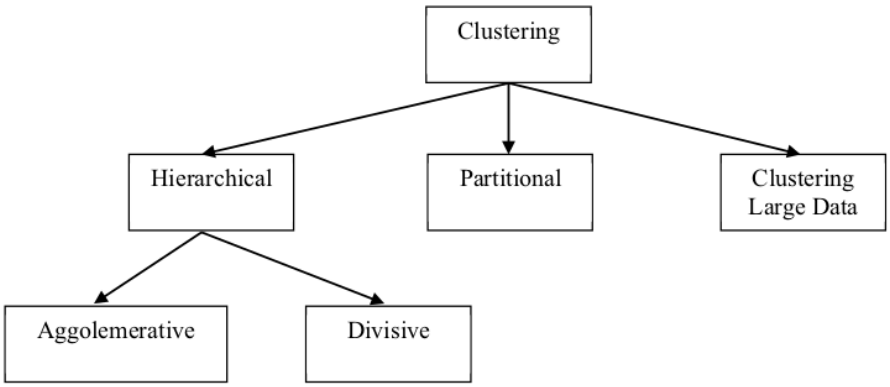
Metode-metode untuk mencari jarak antar-*cluster*:

a. *Single Link*: jarak terkecil antar suatu elemen dalam suatu *cluster* dengan elemen lain di *cluster* yang berbeda.
b. *Complete Link*: jarak terbesar antar satu elemen dalam suatu *cluster* dengan elemen lain di *cluster* yang berbeda.

c. *Average*: jarak rata-rata antar satu elemen dalam suatu *cluster* dengan elemen lain di *cluster* yang berbeda.
d. *Centoid*: jarak antar *centroid* dari tiap *cluster* dengan *centroid cluster* lainnya.
e. *Medoid*: jarak antar *medoid* dari tiap *cluster* dengan *medoid cluster* lainnya.

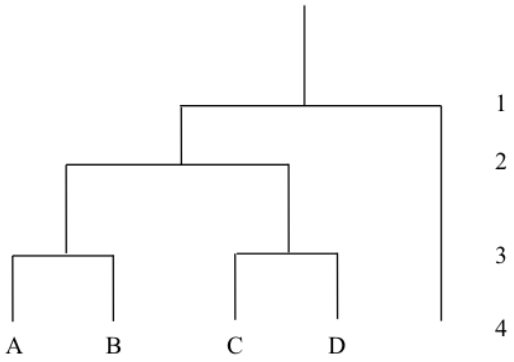
2.2.2 Algoritma *Clustering*

Secara umum pembagian algoritma *clustering* dapat digambarkan sebagai berikut:



Gambar 1. Kategori Algoritma *Clustering*

Hierarchical clustering menentukan sendiri jumlah *cluster* yang dihasilkan. Hasil dari metode ini adalah suatu struktur data berbentuk pohon yang disebut *dendogram*. Data dikelompokkan secara bertingkat dari yang paling bawah, tiap *instance* data merupakan satu *cluster* sendiri, hingga tingkat paling atas secara keseluruhan data membentuk satu *cluster* besar berisi *cluster-cluster* seperti gambar 2.



Gambar 2. *Dendogram*

Divisive hierarchical clustering mengelompokkan data dari kelompok yang terbesar hingga ke kelompok yang terkecil, yaitu masing-masing *instance* dari kelompok data tersebut. Sebaliknya, *agglomerative hierarchical clustering* mulai mengelompokkan data dari kelompok yang terkecil hingga kelompok yang terbesar [8]. Beberapa algoritma yang menggunakan metode ini adalah: *Robust Clustering Using LinKs* (ROCK), *Chameleon*, *Cobweb, Shared Nearest Neighbor* (SNN).

Partitional clustering yang mengelompokkan data ke dalam k *cluster* dimana k adalah banyaknya *cluster* dari *input user*. Kategori ini biasanya memerlukan pengetahuan yang cukup mendalam tentang data dan proses bisnis yang memanfaatkannya untuk mendapatkan kisaran nilai input yang sesuai. Beberapa algoritma yang masuk dalam kategori ini antara lain: *K-Means, Fuzzy C-Means, Clustering Large Aplications* (CLARA), *Expectation Maximation* (EM), *Bond Energy Algorithm* (BEA), algoritma Genetika, Jaringan Saraf Tiruan.

Clustering Large Data, dibutuhkan untuk melakukan *clustering* pada data yang volumenya sangat besar sehingga tidak cukup ditampung dalam memori komputer pasa suatu waktu. Biasanya untuk mengatasi masalah besarnya volume data, dicari teknik-teknik untuk meminimalkan berapa kali algoritma harus membaca seluruh data. Beberapa algoritma yang masuk dalam kategori ini antara lain: *Balanced Iteratif Reducing and clustering using hierarchies* (BIRCH), Density *Based Spatial Clustering of Application With Noise* (DCSCAN), *Clustering Categorical Data Using Summaries* (CACTUS).

2.2.3 Algoritma *Fuzzy Clustering C-Means* (FCM)

Clustering dengan metode *Fuzzy C-Means* (FCM) didasarkan pada teori logika *fuzzy*. Teori ini pertama kali diperkenalkan oleh Lotfi Zadech (1965) dengan nama himpunan *fuzzy (fuzzy set).* Dalam teori *fuzzy* keanggotaan sebuah data tidak diberikan nilai secara tegas dengan nilai 1 (menjadi anggota) dan 0 (tidak menjadi anggota), melainkan dengan suatu nilai derajat keanggotaan yang jangkauan nilainya 0 sampai 1. Nilai keanggotaan suatu data dalam sebuah himpunan menjadi 0 ketika sama sekali tidak menjadi anggota, dan menjadi 1 ketika menjadi anggota secara penuh dalam suatu himpunan. Umumnya nilai keanggotaannya antara 0 dan 1. Semakin tinggi nilai keanggotannya maka semakin tinggi derajat keanggotaannya, dan semakin kecil maka semakin rendah derajat keanggotaannya. Kaitannya dengan *K-Means*, sebenarnya FCM merupakan versi *fuzzy* dari *K-Means* dengan beberapa modifikasi yang membedakannya dengan *K-Means* (Prasetyo, 2013) [9]

Diasumsikan ada sejumlah data dalam set data X yang berisi n data yang dinotasikan $X = \{x_1, x_2, \dots, x_n\}$, di mana setiap data memiliki fitur r dimensi: $x_{i1}, x_{i2}, \dots, x_{ir}$, dinotasikan $x_i = \{x_{i1}, x_{i2}, \dots, x_{ir}\}$. Ada sejumlah *cluster C* dengan *centroid*: $c_1, c_2, \dots, c_k$, k adalah jumlah *cluster*. Setiap data mempunyai derajat keanggotaan pada setiap *cluster*, dinyatakan dengan nilai di antara 0 dan 1, i menyatakan data x_i dan j menyatakan *cluster* c_j. jumlah nilai derajat keanggotaan setiap data x_i selalu sama dengan 1, yang diformulasikan pada persamaan berikut: [9]

$$\sum_{j=1}^{k} u_{ij} = 1 \quad \dots\dots (6)$$

Untuk *cluster* c_j, setiap *cluster* berisi paling sedikit satu data dengan nilai keanggotaan tidak nol, namun tidak berisi derajat satu pada semua data. *Cluster* c_j dapat diformulasikan sebagai berikut:

$$0 < \sum_{i=1}^{n} u_{ij} < n \quad \dots\dots (7)$$

Seperti halnya teori himpunan *fuzzy* bahwa suatu data bisa menjadi anggota di beberapa himpunan yang dinyatakan dengan nilai derajat keanggotaan pada setiap himpunan, maka dalam FCM setiap data juga menjadi anggota pada setiap *cluster* dengan derajat keanggotaan u_{ij}.

Nilai derajat keanggotaan data x_i pada *cluster* c_j, diformulasikan pada persamaan berikut :

$$u_{ij} = \frac{D(x_i, c_j)^{\frac{-2}{w-1}}}{\sum_{l=1}^{k} D(x_i, c_l)^{\frac{-2}{w-1}}} \quad \dots\dots (8)$$

Parameter c_j adalah *centroid cluster* ke-*j*, D( ) adalah jarak antara data dengan *centroid*, sedangkan w adalah parameter bobot pangkat (*weighting exponent*) yang diperkenalkan dalam FCM. w tidak memiliki nilai ketetapan, biasanya $w > 1$ dan umumnya diberi nilai 2.

Nilai keanggotaan tersebut disimpan dalam matriks *fuzzy pseudo-partition* berukuran Nxk, di mana baris merupakan data, sedangkan kolom adalah nilai keanggotaan pada setiap *cluster*. Bentuknya seperti di bawah ini:

$$U = \begin{bmatrix} u_{11}(x_1) & u_{12}(x_1) & \dots & u_{1k}(x_1) \\ u_{21}(x_1) & u_{22}(x_2) & \dots & u_{2k}(x_2) \\ \dots & \dots & \dots & \dots \\ u_{n1}(x_n) & u_{n1}(x_n) & \dots & u_{nk}(x_n) \end{bmatrix} \quad \dots\dots (9)$$

Untuk menghitung *centroid* pada *cluster* c_l pada fitur *j*, digunakan persamaan berikut :

$$c_{ij} = \frac{\sum_{i=1}^{N}(u_{il})^w x_{ij}}{\sum_{i=1}^{N}(u_{il})^w} \quad \ldots\ldots (10)$$

Parameter N adalah jumlah data, *w* adalah bobot pangkat, dan u_{il} adalah nilai derajat keanggotaan data x_i ke *cluster* c_l.

Sementara fungsi objektif menggunakan persamaan berikut:

$$J = \sum_{i=l}^{N}\sum_{i=l}^{k}\left(u_{ij}\right)^w D(x_i, c_l)^2 \quad \ldots\ldots (11)$$

a. Menentukan data yang akan di *cluster* X, berupa matriks berukuran n x m (n=jumlah sampel data, m = atribut setiap data). Xij = data sampel ke-i (i=1,2,...,n), atribut ke-j (j=1,2,...,m).

b. Menentukan Nilai Parameter Awal:
- Jumlah *cluster* = c
- Pangkat = w
- Maksimum interasi = MaxIter
- Error terkecil yang diharapkan = ξ
- Fungsi objektif awal = $P_o = 0$
- Interasi awal = t =1

c. Membangkitkan bilangan random μik, i=1,2,3 ..., n; k=1,2,3 ...c; sebagai elemen-elemen matriks partisi awal U.

$$\mu_{ik}\varepsilon[0,1]; \quad 1 \le i \le n; \quad 1 \le k \le c \quad \ldots\ldots(12)$$

d. Menghitung jumlah setiap kolom:

$$Q_1 = \sum_{k=1}^{c} \mu_{ik} . \quad \ldots\ldots(13)$$

dengan j=1,2,...n.

e. menghitung pusat *cluster* ke-k: V_{kj}, dengan k=1,2,...c; dan j=1,2,...m

$$V_{kj} = \frac{\sum_{i=1}^{n}((\mu_{ik})^w X_{ij})}{\sum_{i=1}^{n}(\mu_{ik})^w} \quad \ldots\ldots (14)$$

f. menghitung fungsi objektif pada interasi ke-t :

$$P_t = \sum_{i=1}^{n}\sum_{k=1}^{C}\left(\left[\sum_{j=1}^{m}\left(X_{ij} - V_{kj}\right)^2\right](\mu_{ik})^w\right) \quad \ldots\ldots(15)$$

g. menghitung perubahan matriks partisi :

$$\mu_{ik} = \frac{\left[\sum_{j=1}^{m}\left(X_{ij}-V_{kj}\right)^2\right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{c}\left[\sum_{j=1}^{m}\left(X_{ij}-V_{kj}\right)^2\right]^{\frac{-1}{w-1}}} \quad \ldots\ldots (16)$$

dengan: i=1,2,...n; dan k=1,2,...c.

h. Memeriksa kondisi berhenti:
- Jika: ($|P_t - P_{t-1}| < \xi$) atau (t > MaxIter) maka berhenti
- Jika tidak: t=t+1, mengulang langkah ke-4.

2.3. Sistem Penjurusan dengan Pendekatan Klusterisasi *Fuzzy Clustering Means* (FCM)

Sistem Penjurusan merupakan proses penyeleksian siswa dalam menentukan jurusan. Dalam penjurusan ini, siswa diberi kesempatan memilih jurusan yang paling cocok dengan karakteristik dirinya. Sistem penjurusan ini dibutuhkan beberapa kriteria penilaian, seperti nilai raport, hasil tes bakat, dan minat siswa. Sistem penjurusan SMA dilakukan akhir semester 2 pada kelas X. Proses penjurusan ini biasanya dilakukan oleh pihak guru atau wali kelas dan guru BK (Bimbingan Konselimg). Setiap wali kelas akan memberikan nilai raport atau nilai akhir dari semua mata pelajaran, kemudian diserahkan kepada guru BK untuk dianalisa dan dilakukan perangkingan serta berdasarkan minat siswa terhadap jurusan [1]. Nilai minat di dapat dari test psikologi yang diselenggarakan oleh sekolah yang telah diprogramkan oleh guru BK kepada siswa di kelas X.

Sesuai kurikulum yang berlaku di seluruh Indonesia, maka siswa kelas X SMA yang naik ke kelas XI akan mengalami pemilihan jurusan/penjurusan. Penjurusan yang tersedia di SMA meliputi Ilmu Alam (IPA), Ilmu Sosial (IPS), dan Ilmu Bahasa. Penjurusan akan disesuaikan dengan minat dan kemampuan siswa. Tujuannya agar kelak di kemudian hari, pelajaran yang akan diberikan kepada siswa menjadi lebih terarah karena telah sesuai dengan minatnya. Dari keseluruhan mata pelajaran di SMA,

tidak seluruhnya dijadikan dasar untuk proses penjurusan, melainkan hanya mata pelajaran inti dari tiap jurusan tersebut. Mata pelajaran inti untuk jurusan IPA terdiri atas: Biologi, Fisika, Matematika IPA, Kimia. Mata pelajaran inti untuk IPS adalah: Sosiologi, Geografi, Sejarah dan Ekonomi. Sedangkan mata kuliah inti untuk jurusan Bahasa adalah: Bahasa Inggris dan Bahasa Indonesia [2].

Tujuan dilaksanakannya penjurusan adalah [2]:

a. Mengelompokkan siswa sesuai kecakapan, kemampuan, dan bakat, yang relatif sama.
b. Membantu mempersiapkan siswa melanjutkan studi dan memilih dunia kerja.
c. Membantu memperkokoh keberhasilan dan kecocokan atas prestasi yang akan dicapai di waktu mendatang (kelanjutan studi dan dunia kerja).

Waktu penentuan penjurusan bagi peserta didik untuk program IPA, IPS dan Bahasa dilakukan mulai akhir semester 2 (dua) kelas X. Pelaksanaan penjurusan program dimulai pada semester 1 (satu) kelas XI. Kriteria penjurusan program dilakanakan berdasarkan nilai akademik. Peserta didik yang naik kelas XI dan akan mengambil program tertentu yaitu: Ilmu Pengetahuan Alam (IPA) atau Ilmu Pengetahuan Sosial (IPS) atau Bahasa: boleh memiliki nilai yang tidak tuntas paling banyak 3 (tiga) mata pelajaran pada mata pelajaran-mata pelajaran yang bukan menjadi ciri khas program tersebut. Peserta didik yang naik ke kelas XI, dan yang bersangkutan mendapat nilai tidak tuntas 3 (tiga) mata pelajaran, maka nilai tersebut harus dijadikan dasar untuk menentukan program yang dapat diikuti oleh peserta didik, misalnya :

a. Apabila mata pelajaran yang tidak tuntas adalah Fisika, Kimia dan Geografi (2 mata pelajaran ciri khas program IPA dan 1 ciri khas program IPS), maka siswa tersebut secara akademik dapat dimasukkan ke program Bahasa.
b. Apabila mata pelajaran yang tidak tuntas adalah Bahasa Indonesia, Bahasa Inggris, dan Fisika, (2 mata pelajaran ciri khas Bahasa dan 1 ciri khas IPA), maka siswa tersebut secara akademik dapat dimasukkan ke program IPS.
c. Apabila mata pelajaran yang tidak tuntas adalah Ekonomi, Sosilologi, dan Bahasa Inggris (2 mata pelajaran ciri khas program IPS dan 1 ciri khas program Bahasa), maka peserta didik tersebut secara akademik dapat dimasukkan ke program IPA.
d. Apabila mata pelajaran yang tidak tuntas adalah Fisika, Ekonomi, dan Bahasa Indonesia (mencakup semua mata pelajaran yang menjadi ciri khas ketiga program di SMA) maka perlu diperhatikan prestasi nilai mata pelajaran yang lebih unggul daripada program lainya (siswa tersebut dapat dijuruskan ke program yang nilai prestasi mata pelajaran yang lebih unggul tersebut), atau dengan mempertimbangkan minat peserta didik. Untuk mengetahui minat peserta didik dapat dilakukan melalui angket/kuesioner dan wawancara, atau cara lain yang dapat digunakan untuk mendeteksinya.

Skala penilaian penilaian yang dapat dijadikan acuan bagi sekolah-sekolah di Indonesia adalah sebagai berikut [2]:

a. Nilai ketuntasan belajar untuk aspek pengetahuan dan praktik dinyatakan dalam bentuk bilangan bulat, dengan rentang 0 -100.
b. Ketuntasan belajar setiap indikator yang telah ditetapkan dalam suatu kompetensi dasar berkisar antara 0 – 100 %. Kriteria ideal ketuntasan untuk masing-masing indikator 75 %.
c. Satuan pendidikan dapat menentukan kriteria ketuntasan minimal (KKM) dibawah nilai ketuntasan belajar ideal. Satuan pendidikan diharapkan meningkatkan kriteria ketuntasan belajar secara terus menerus untuk mencapai kriteria ketuntasan ideal.
d. KKM ditetapkan oleh forum guru pada awal tahun pelajaran.
e. KKM tersebut dicantumkan dalam LHB dan harus diinformasikan kepada seluruh warga sekolah dan orang tua siswa.

Penentuan konsentrasi jurusan dalam penelitian ini adalah dengan menggunakan konsep *Fuzzy Cluster* Means (FCM) yakni dengan cara mengelompokkan data nilai siswa menurut kemiripan yang dimilikinya.

3. METODE PENELITIAN

3.1. Metode Pengumpulan data

Penelitian ini menggunakan data nilai prestasi dan hasil tes psikologi minat siswa Sekolah Menengah Atas (SMA) yang didapatkan dari SMA Negeri 1 Barabai Kabupaten Hulu Sungai Tengah Provinsi Kalimantan Selatan.

Data yang dibutuhkan dalam penelitian ini adalah data sekunder yang berupa data nilai prestasi siswa Kelas X angkatan tahun pelajaran 2012/2013 pada SMA Negeri 1 Barabai Kabupaten Hulu Sungai Tengah Provinsi Kalimantan Selatan dan nilai minat dari tes psikologi yang di dapatkan dari kerjasama sekolah dengan pihak ketiga. Sampel data yang digunakan dalam penelitian ini adalah data nilai sebelum penjurusan dan data nilai setelah penjurusan siswa SMA Negeri 1 Barabai angkatan tahun 2012 sebanyak 278 siswa (seluruh siswa kelas X Tahun Pelajaran 2012/2013).

3.2. Metode Pengolahan Data Awal

Data yang didapatkan dari akademik SMA Negeri 1 Barabai masih berupa data yang terdiri dari data rekapitulasi nilai sebelum peminatan / penjurusan dan data nilai sesudah peminatan / penjurusan. Untuk data rekapitulasi nilai di bagi menjadi dua kelompok nilai yaitu IPA dan IPS, karena di SMA Negeri 1 Barabai untuk siswa angkatan tahun ajaran 2012-2013 hanya menyelenggarakan 2 (dua) program peminatan / penjurusan.

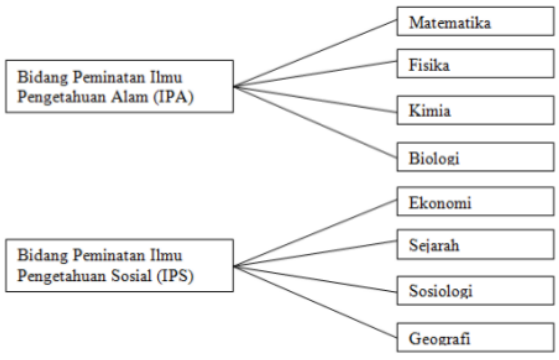
Pada rekapitulasi nilai sebelum peminatan / penjurusan harus dipilah dahulu kelompok nilai jurusan IPA (Matematika, Fisika, Kimia dan Biologi) dan kelompok nilai jurusan IPS (Sejarah, Geografi, Ekonomi dan Sosiologi). Data proses ini mempunyai atribut: Siswa, Nilai Rata-rata IPA, Nilai Rata-rata IPS, Nilai minat tes Psikologi.

Sedangkan rekapitulasi nilai sesudah peminatan/penjurusan mempunyai atribut: Siswa, Nilai Rata-rata Mata Pelajaran jurusan sesudah Peminatan Kls XI (Semester 1, Semester 2). Jumlah sampel data keseluruhan ada 278 data siswa angkatan tahun 2012.

3.3. Eksperimen dan Pengujian Model

Pada penelitian ini penulis menggunakan algoritma *Fuzzy C-Means* untuk membantu menentukan penjurusan siswa Sekolah Mengah Atas (SMA). Metode ini dipilih karena kemampuannya untuk melakukan pengelompokan suatu objek/data yang belum memiliki klasifikasi, ke dalam kelas tertentu menurut kesamaan yang dimilikinya berdasarkan derajat keanggotaan dengan cara minimalisasi nilai fungsi objektifnya[4].

Algoritma ini diterapkan pada data nilai siswa sebelum peminatan/ penjurusan tahun ajaran 2012-2013 ditambahkan dengan nilai minat tes psikologi dan data nilai siswa sesudah peminatan tahun ajaran 2013-2014. Pada tahap awal dilakukan pemetaan korelasi antara peminatan dengan mata pelajaran peminatan, hasilnya ditunjukan pada gambar berikut.



Gambar 3. Pemetaan antara Peminatan dengan Mata Pelajaran Penjurusan

Selanjutnya data nilai siswa diurutkan berdasarkan posisi daftar siswa dikelas sesudah penjurusan yang telah dijalani. Setelah itu ditentukan nilai bidang jurusan tertentu yang diperoleh dari hasil rata-rata mata pelajaran peminatan / penjurusan yang berada dalam kelompok bidang jurusan tersebut sebelum dilakukan penjurusan dan ditambahkan nilai minat dari hasil tes psikologi. Nilai minat IPA diberi label angka 1, dan nilai minat IPS diberi label angka 2. Kemudian Data ini digunakan sebagai data parameter ujicoba penjurusan menggunakan *Fuzzy C-Means* /data nilai sebelum.

Setelah parameter nilai rata-rata bidang minat diketahui, selanjutnya dilakukan pemetaan/klastering data mengikuti algoritma FCM:

a. Menetapkan matriks partisi awal U berupa matriks berukuran n x m (n adalah jumlah data, yaitu=278, dan m adalah parameter/atribut setiap data, yaitu=2). Xij = data sampel ke-i (i=1,2,...,n), atribut ke-j (j=1,2,...,m). Data yang digunakan

b. Menentukan Nilai Parameter Awal :
   1) - Jumlah *cluster* (c) = 2
   2) - Pangkat (w) = 2
   3) - Maksimum interasi (MaxIter) = 25
   4) - Error terkecil yang diharapkan (ξ) = 10^{-5}
   5) - Fungsi objektif awal (P_0) = 0
   6) - Interasi awal (t) = 1

c. Menguji data dengan bantuan aplikasi (tools), meliputi :
   1) Membangkitkan bilangan random μij, i=1,2,...,n; k=1,2,...c; sebagai elemen-elemen matriks partisi awal (U) dengan memberikan nilai sembarang dalam jangkauan [0,1] dengan syarat seperti pada persamaan (6).
   2) Menentukan Pusat Klaster (C)
   3) Pusat klaster dihitung menggunakan persamaan (10) .
   4) Menghitung Fungsi Objektif (P) berdasarkan persamaan (16).
   5) Menghitung Perubahan Matriks Partisi (U) berdasarkan persamaan (17).
   6) Mengecek kondisi berhenti
      Karena : $|P_t - P_0|$= 26,8753 - 0 lebih besar dari 0,00001 dan iterasi =1 maka dilanjutkan ke iterasi berikutnya sampai nilai $|Pt - Pt\text{-}1| < \xi$ yaitu 0,00001 atau t > MaxIter (iterasi bernilai 25). Jika tidak maka mengulangi dari langkah ke-4.
   7) Menyimpulkan hasil pengklasteran dari data yang di olah.

Setelah dilakukan penghitungan menggunakan algoritma *Fuzzy C-Means* sampai dengan iterasi 25 diperoleh nilai fungsi objektif = 26,94500 dan perubahan nilainya adalah : |Pt – Pt-1| = 26,94500 – 26,94485 = 0,00015. Karena nilai batas maksimum iterasi sudah tercapai maka proses dihentikan sampai di sini walaupun nilai |Pt – Pt-1| > 0,00001.

3.4. Metode Pengukuran

Akurasi penerapan algoritma *Fuzzy C-Means* (FCM) dalam peminatan / penjurusan di SMA diuji dengan cara : [4]

a. Mengelompokan data sampel siswa dan nilai rata-rata mata pelajaran peminatan / penjurusan sebelum peminatan dengan menggunakan algoritma *Fuzzy C-Means* untuk membagi siswa ke dalam bidang minat / jurusan tertentu (dalam kasus ini hanya kelompok IPA dan IPS) yang disesuaikan dengan kesamaan perolehan nilai rata-rata bidang peminatan / jurusan tersebut.

b. Membandingkan hasil peminatan / penjurusan *Fuzzy C-Means* dengan hasil peminatan / penjurusan yang telah dilaksanakan di tempat penelitian (terhadap data sampel nilai rata-rata mata pelajaran pada peminatan yang telah dijalani oleh siswa di sekolah).

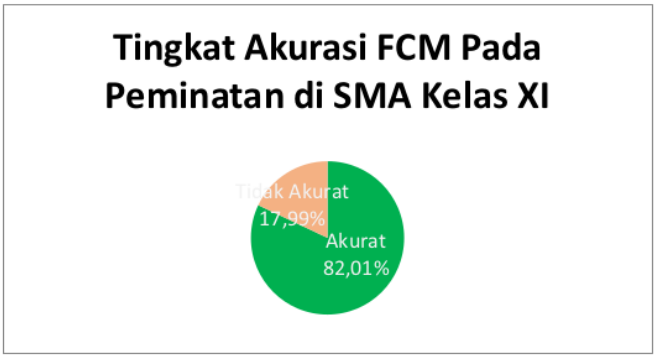
c. Jika minat yang dipilih oleh siswa sama dengan penjurusan oleh FCM dan nilai rata-rata mata pelajaran peminatan yang diperoleh setelah peminatan >= Kriteria Ketuntasan Minimal (KKM) yang ideal (>=75), maka FCM dinyatakan AKURAT.

d. Jika minat yang dipilih oleh siswa sama dengan penjurusan FCM dan nilai rata-rata mata pelajaran penjurusan yang diperoleh setelah penjurusan < Kriteria Ketuntasan Minimal (KKM) yang ideal, maka FCM dinyatakan TIDAK AKURAT.

e. Jika minat yang dipilih oleh siswa tidak sama dengan penjurusan FCM dan nilai rata-rata mata pelajaran penjurusan yang diperoleh setelah peminatan >= Kriteria Ketuntasan Minimal (KKM) yang ideal, maka FCM dinyatakan TIDAK AKURAT.
f. Jika minat yang dipilih oleh siswa tidak sama dengan penjurusan FCM dan nilai rata-rata mata pelajaran peminatan yang diperoleh setelah peminatan < Kriteria Ketuntasan Minimal (KKM) yang ideal, maka FCM dinyatakan AKURAT.
g. Kemudian dihitung persen tingkat akurasi FCM dengan:
h. % Akurasi = (Jumlah Data Akurat / Total Sampel) * 100

4. HASIL PENELITIAN

Dari matriks partisi U iterasi terakhir dapat diperoleh informasi mengenai kecenderungan siswa untuk masuk ke kelompok peminatan tertentu. Setiap peminatan memiliki derajat keanggotaan tertentu untuk menjadi anggota suatu kelompok. Derajat keanggotaan terbesar menunjukkan kecenderungan tertinggi seorang siswa untuk masuk menjadi anggota peminatan tertentu.



Gambar 4. Tingkat Akurasi FCM pada Peminatan di SMA Kelas XI

Hasil peminatan yang dilakukan oleh algoritma *Fuzzy C-Means* (FCM) dapat dijelaskan bahwa pada pelaksanaan penjurusan ke kelas XI, dari sebanyak 278 data 82,01% yang tepat dalam memilih peminatan. Sedangkan penjurusan dengan cara manual berdasarkan pilihan individu siswa hanya 63,67%, sehingga penjurusan dengan FCM lebih tinggi 18,34%. Akurasi peminatan algoritma FCM disajikan pada gambar di atas.

5. KESIMPULAN

Dari Hasil pengujian algoritma *Fuzzy C-Means* (FCM) dalam penentuan jurusan di Sekolah Menengah Atas pada 278 data siswa yang diuji dalam penelitian ini, menunjukkan bahwa Algoritma FCM memiliki tingkat akurasi yang baik (yaitu rata-rata 82,01%) dengan memasukkan nilai tes minat dibandingkan dengan cara manual berdasarkan pilihan individu siswa hanya 63,67%, sehingga penjurusan dengan FCM lebih tinggi 18,34%.

Proses klastering pada penelitian ini hanya dibagi ke dalam 2 (dua) klaster, karena mengikuti jumlah penjurusan yang di sekolah tempat pengambilan data (jurusan IPA dan IPS). Sehingga belum mengacu kepada penjurusan / peminatan Sekolah Menengah Atas pada umumnya yang terbagi ke dalam 3 (tiga) jurusan, yaitu jurusan IPA, IPS dan Bahasa.

UCAPAN TERIMAKASIH

Penelitian ini dapat terselesaikan karena bantuan berbagai pihak, oleh karena itu peneliti berterimakasih kepada pihak-pihak yang mendukung terlaksananya penelitian yaitu para pembimbing penelitian, penguji,, serta pihak-pihak lain yang mendukung terlaksananya penelitian ini.

PERNYATAAN ORIGINALITAS

"Saya menyatakan dan bertanggung jawab dengan sebenarnya bahwa Artikel ini adalah hasil karya sendiri kecuali cuplikan dan ringkasan yang masing-masing telah saya jelaskan sumbernya". [ALTANOVA REZA-P31.2012.01263]

DAFTAR PUSTAKA

[1] Departemen Pendidikan Nasional. (2004) *Panduan Penilaian Penjurusan Kenaikan Kelas dan Pindah Sekolah*, Jakarta: Direktorat Pendidikan Menengah Umum, Jakarta.

[2] Departemen Pendidikan Nasional (2006) *Panduan Penyusunan Laporan Hasil Belajar Peserta Didik Sekolah Menengah Atas (SMA)*, Direktorat Jenderal Manajemen Pendidikan Dasar Dan Menengah Direktorat Pembinaan SMA, Jakarta.

[3] Roida Eva Flora Siagian, 2012. *Pengaruh minat dan Kebiasaan Belajar Siswa terhadap Prestasi Belajar Matematika*. Jurnal formatif 2(2) : 122-131.

[4] Bahar, 2011, *Penentuan Jurusan SMA dengan Fuzzy C-Means*. Tesis, Pasca Sarjana Magister Teknik Informatika UDINUS Semarang.

[5] Zati Azmiana, Faigiziduhu Bu'ulolo, dan Partano Siagian, 2013, *Penggunaan Sistem Inferensi Fuzzy Untuk Penentuan Jurusan di SMA Negeri 1 Bireuen*, Saintia Matematika Vol. 1 No. 3, pp 233-247.

[6] Tb. Ai Munandar, Wahyu Oktri Widyarto, Harsiti, 2013, *Clustering* data nilai mahasiswa untuk pengelompokan Konsentrasi Jurusan menggunakan *Fuzzy Cluster* Means, Universitas Serang Raya Banten.

[7] Dunham, Margaret,H. (2003), *Data mining* Introuctory and Advanced Topics, New Jersey, Prentice Hall.

[8] Kantardzic, Mehmed, (2003), *Data mining* Concepts Models, Methods, and Algorithms, New Jersey, IEEE

[9] Eko Prasetyo,M.Kom, 2014, *Data mining* Mengolah Data Menjadi Informasi menggunakan Matlab,Andi Offset,Yogyakarta.

PENENTUAN JURUSAN SISWA SEKOLAH MENENGAH ATAS DISESUAIKAN DENGAN MINAT SISWA MENGGUNAKAN ALGORITMA FUZZY C-MEANS

ORIGINALITY REPORT

9%	6%	3%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	docplayer.info Internet Source	2%
2	M A Muchtar, I Jaya, Manguji Nababan, U Andayani, Lisa Noprianti Siregar, Erna B Nababan, Opim S Sitompul. "Separation of Basic Words in Angkola Batak Text Documents using Enhanced Confix Stripping Stemmer Case: Mandailing Ethnic", IOP Conference Series: Materials Science and Engineering, 2019 Publication	2%
3	ebookinga.com Internet Source	1%
4	eprints.dinus.ac.id Internet Source	1%
5	yunita113070288.wordpress.com Internet Source	1%

6	Nelpita Ulandari, Rahmi Putri, Febria Ningsih, Aan Putra. "Efektivitas Model Pembelajaran Inquiry terhadap Kemampuan Berpikir Kreatif Siswa pada Materi Teorema Pythagoras", Jurnal Cendekia : Jurnal Pendidikan Matematika, 2019 Publication	<1%
7	J.C. Ribes, G. Delaunay, E. Merle, M. Mouillet. "Smart Diagnosis on the AIRIX Induction Accelerator", International Journal of Smart Engineering System Design, 2003 Publication	<1%
8	António Manuel Lourenço Canelas, Jorge Manuel Correia Guilherme, Nuno Cavaco Gomes Horta. "Chapter 4 Monte Carlo-Based Yield Estimation: New Methodology", Springer Science and Business Media LLC, 2020 Publication	<1%
9	teknois.stikombinaniaga.ac.id Internet Source	<1%
10	doczz.net Internet Source	<1%
11	www.sman3bdg.sch.id Internet Source	<1%
12	jurnal.pnj.ac.id Internet Source	<1%

13	Agustian Noor. "Perangkat Lunak Pembelajaran Metode Kriptografi WAKE (Word Auto Key Encryption)", Jurnal Sains dan Informatika, 2017 Publication	<1%
14	membangunbersama-sama.blogspot.com Internet Source	<1%
15	docobook.com Internet Source	<1%
16	Lisna Lisna, Agnes Vincentia, Noferdiman Noferdiman, Jasmine Masyitha Amelia. "INVENTORY OF FISHING GEAR IN KECAMATAN TUNGKAL ILIR, TANJUNG JABUNG BARAT, JAMBI", AQUASAINS, 2018 Publication	<1%
17	R. Xu. "Survey of Clustering Algorithms", IEEE Transactions on Neural Networks, 5/2005 Publication	<1%
18	Sunil Bharitkar. "", IEEE Transactions on Audio Speech and Language Processing, 2/2007 Publication	<1%
19	IFMBE Proceedings, 2013. Publication	<1%

Exclude quotes	Off	Exclude matches	Off

Exclude bibliography On