

***Data Clustering* Menggunakan K-Mean sebagai Pendukung Pengembangan Sistem Personalisasi E-Learning**

Purwono Hendradi¹⁾, Harry Budi Santoso²⁾, Ahmad Arief Prasetyo³⁾

Teknik Informatika Universitas Muhammadiyah Magelang^{1&3)}, Ilmu Komputer, Universitas Indonesia²⁾
Jl. Mayjend Bambang Soegeng KM 5 Magelang, 0293 326945^{1&3)}.,Depok Jawa Barat 16424²⁾
p_hendra@ummgl.ac.id¹⁾, harrybs@cs.ui.ac.id²⁾, insaan.khairan@gmail.com³⁾

Abstrak

Personalisasi web merupakan bentuk dari perubahan budaya dalam Teknologi Informasi dan Komunikasi (TIK). Saling berbagi informasi antar penyedia jasa layanan web membuat tampilan halaman web semakin dinamis yaitu menyesuaikan kebutuhan dan keinginan dari pengguna. Dalam pemanfaatan TIK untuk peningkatan pendidikan, personalisasi web diterapkan pada *learning management system*. Pengelompokan siswa dengan mengacu pada gaya belajar, kemampuan belajar dan lainnya, akan menjadi pertimbangan dalam personalisasi *e-learning* ini. *Data Clustering* dari siswa dalam berinteraksi dengan sistem adalah kombinasi data nilai saat Sekolah Lanjutan Tingkat Atas (SLTA) dan data prestasi matakuliah pendukung pada semester sebelumnya, sehingga menghasilkan pengelompokan siswa yang akan dijadikan rekomendasi untuk pengembangan personalisasi *e-learning*. Sistem pembelajaran *e-learning* dapat memiliki kemampuan yang menyerupai seorang pengajar dalam mengelola kelas konvensional. Harapannya fungsi *e-learning* sebagai alternatif media pembelajaran, juga menjadi suplemen atau pendukung pembelajaran.

Kata kunci: Personalisasi *e-learning*, Data clustering, suplemen

1. Pendahuluan

Sistem pembelajaran *e-learning*, saat ini sudah mulai lumrah dikembangkan dan diimplementasikan di lingkungan institusi pendidikan, mulai dari tingkatan sekolah dasar hingga perguruan tinggi. Sistem ini juga digunakan oleh pihak non sekolah, seperti pihak penerbit sertifikasi kompetensi, yang menggunakan sistem ini sebagai sarana memasarkan produk layanannya : mulai dari pendaftaran, pembukaan kelas, interaksi dan uji sertifikasi. Sebagai contoh *CISCO Network Academy (CISCO Netacad)*. Perubahan dari model pembelajaran konvensional menjadi *e-learning* bukan berarti *full* pembelajaran berbasis internet, akan tetapi masih ada pembelajaran tatap mukanya. Hal seperti ini sering diistilahkan sebagai *blended learning*, yaitu dalam satu periode skenario pembelajaran disiapkan skenario pembelajaran via internet dan pembelajaran tatap muka, karena ada hal-hal yang tidak bisa tergantikan yang harus dilakukan dengan pembelajaran tatap muka.

Kehadiran sistem pembelajaran *online (e-Learning)* yang diharapkan mampu meningkatkan intensitas belajar mandiri belum bias menunjukkan peran pentingnya karena di Indonesia masih banyak pembelajar yang belum memiliki budaya belajar mandiri. Oleh karena itu, akan dibangun sistem *e-Learning* yang mampu mendorong pembelajar untuk berperan aktif, menitik beratkan pada *personalization* yang meliputi kemampuan menyesuaikan diri (adaptif) terhadap tingkat kemampuan pembelajar, serta mengoleksi sumberdaya pengetahuan yang saling mendukung [1].

Dalam tatap muka antara pendidik (Guru atau Dosen) dengan peserta didik (siswa) seringkali menimbulkan suatu persepsi dari seorang pendidik terhadap peserta didiknya. Seperti pendidik menyimpulkan kemampuan peserta didik, pendidik meramalkan tingkat keberhasilan peserta didik dan sebagainya. Sebenarnya hal ini dilakukan karena pendidik memiliki intuisi untuk menganalisis dan menyimpulkan berdasarkan data dan interaksi dengan peserta didik di kelas. Data siswa sebelum mengikuti pelajaran dan data siswa saat mengikuti pelajaran dijadikan bahan untuk menganalisis. Kemampuan pendidik seperti ini yang akan digunakan untuk mengembangkan sistem *e-learning*, yaitu sistem *e-learning* yang dapat menganalisa kemampuan siswa berdasarkan data dan interaksi dengan sistem. Seperti perkembangan web saat ini, interaksi pengguna dengan web media sosial, membuat konten dari halaman web media sosial berbeda-beda pada masing-masing pengguna.

Untuk menghasilkan sistem *e-learning* seperti diatas, digunakan proses *data mining* dengan menggunakan metode *clustering* yang *unsupervised* yaitu K-Mean, karena metode ini diharapkan dapat menghasilkan *cluster* yang lebih alamiah. Yang menjadi permasalahan, sejauh mana K-Mean dapat menghasilkan *clustering* siswa yang dapat digunakan sebagai pendukung keputusan untuk mengembangkan *e-learning* yang mengadung unsur seperti *web personalize*. Dalam hal ini digunakan

studi kasus mata matakuliah *Network and Communication* (Netcom) pada Prodi Teknik Informatika Universitas Muhammadiyah Magelang.

2. Tinjauan Pustaka

Educational Data Mining (EDM) adalah mengaplikasikan teknik *Data Mining* dalam dunia pendidikan. Proses EDM menjadikan guru dan peneliti pendidikan mendapat informasi dengan mengolah data mentah menjadi informasi dipahami [2]. Dengan alat dalam EDM, pengelompokan siswa dapat lebih efektif menggunakan algoritma *clustering*. Tidak ada algoritma *clustering* yang dapat memberikan *cluster* sempurna pada semua set data. Pengguna dapat menggunakan alat ini lebih baik, hanya ketika mereka lebih memahami hal itu. Metode *clustering* dibagi menjadi dua macam, yaitu *supervised* dan *unsupervised*. Contoh untuk *unsupervised* adalah *clustering K-Means* dan *K-Medoid* sedangkan *supervised* adalah *Fuzzy C-Means*.

Penggunaan EDM dalam suatu siklus pembelajaran adalah saat sistem pembelajaran berlangsung. EDM dijalankan dengan data hasil interaksi siswa dengan pengajar dan juga data-data lainnya dari seorang siswa [3]. Output dari EDM ada dua arah, untuk pengajar dan pengelola system dihasilkan adalah informasi dan pengetahuan yang digali dari siswa. Adapun pada sisi siswa adalah rekomendasi untuk kelanjutan pembelajarannya.

Selain itu keuntungan besar dari penggunaan metode EDM adalah dapat dimasukkan sebagai bagian dari pengembangan sistem e-learning. [4]. Dengan EDM dapat ditemukan pengetahuan baru dari pengalihan data siswa, sehingga berpotensi meningkatkan beberapa aspek kualitas pendidikan dan untuk meletakkan dasar untuk proses belajar lebih efektif [5].

Dalam perkuliahan, prestasi mahasiswa dipengaruhi oleh asal program studi, asal kota dan asal SMA [6]. Data ini diolah dan dipadu dengan data hasil tes perkuliahan dengan menggunakan EDM. Hasil pengolahannya menghasilkan informasi yang berguna untuk pengembangan dan pengambilan keputusan untuk proses pendidikan. Variasi dari latar belakang SLTA dapat menjadikan segmen yang membagi tiap profile siswa. Siswa dari SLTA yang umum (SMA) lebih sukses dalam materi Sistem Informasi Manajemen (SIM) dibandingkan siswa dari sekolah vokasi (SMK) pada pengetahuan dasar manajemen atau keterampilan komputer. Hasil analisis ini juga digunakan dalam membentuk berbagai strategi tingkat makro dan mikro di departemen MIS. [7]

Untuk melaksanakannya diperlukan kegiatan yang mengkombinasikan penilaian, pengukuran dan evaluasi. Penilaian dapat diartikan sebagai kegiatan menafsirkan data hasil pengukuran dan evaluasi digunakan sebagai penyediaan informasi yang dapat dijadikan sebagai bahan pertimbangan dalam mengambil keputusan [8]. Dalam penelitian ini menggunakan penilaian dan evaluasi, penilaian digunakan untuk menafsirkan data hasil pengukuran untuk menentukan tingkat kompetensi. Sedangkan evaluasi digunakan untuk menyediakan data dalam mengambil keputusan atas scenario/rencana pembelajaran.

Metode EDM yang dibutuhkan dalam RPP (Rencana Pelaksanaan Pembelajaran) adalah yang dapat memberikan informasi mengenai pengelompokan siswa, yaitu pengelompokan siswa untuk membagi siswa dalam melaksanakan pembelajaran dan pengelompokan siswa untuk mengukur tingkat perkembangan siswa. Untuk keperluan pertama diatas digunakan metode *clustering Fuzzy C-mean*, sedangkan keperluan yang kedua digunakan metode *K-Mean*. [9].

Dengan tujuan menciptakan sebuah personalisasi e-learning untuk lingkungan sekolah menengah kejuruan, dikembangkanlah personalisasi e-learning yang mengadopsi dimensi visual/verbal [10]. Sistem memiliki sebuah algoritma yang dapat menentukan tipe materi pembelajaran (visual/verbal) yang dapat yang sesuai dengan gaya belajar siswa [11]. Personalisasi pembelajaran dengan memanfaatkan e-learning memberikan keleluasan dalam memfasilitasi perbedaan karakteristik, kebutuhan, dan preferensi setiap pembelajar.

2.1. Algoritma K-Mean

Algoritma untuk melakukan *K-Means clustering* adalah sebagai berikut

Pilih K buah titik centroid secara acak

Kelompokkan data sehingga terbentuk K buah *cluster* dengan titik centroid dari setiap *cluster* merupakan titik centroid yang telah dipilih sebelumnya.

Proses pengelompokan data ke dalam suatu *cluster* dapat dilakukan dengan cara menghitung jarak terdekat dari suatu data ke sebuah titik centroid. Perhitungan jarak Euclidean dapat digunakan untuk menghitung jarak antar 2 buah data. Rumus perhitungan jarak Euclidean adalah sebagai berikut :

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (2.1)$$

Di mana:

- Dij = Jarak Kuadrat Euclidean antar objek ke-i dengan objek ke-j
- P = Jumlah variabel *cluster*
- Xik = Nilai atau data dari objek ke-i pada variabel ke-k
- Xjk = Nilai atau data dari objek ke-j pada variabel ke-k (Everitt, 1993).

Perbaharui nilai titik centroid.

Untuk memperbarui titik centroid terdapat beberapa langkah :

Untuk setiap catatan, tentukan pusat kelompok terdekatnya dan tetapkan catatan tersebut sebagai kelompok anggota dari kelompok yang terdekat pusat kelompoknya.

Hitung BCV (*BetweenClusterVariation*) = jarak antar cluster

$$BCV = d(m_1, m_2) + d(m_2, m_3) + d(m_1, m_3) \tag{2.2}$$

Hitung WCV (*WithinClusterVariation*) = jarak antar anggota dalam cluster.

$$WCV = \sum_{i=1}^k \sum_{p=ci} d(p, mi)^2 \tag{2.3}$$

$$\text{Hitung rasio} = \frac{BCV}{WCV} \tag{2.4}$$

Bandingkan rasio tersebut dengan rasio sebelumnya jika sudah ada. Jika rasio tersebut nilainya semakin besar, maka lanjutkan ke langkah ke-4. Namun jika tidak, hentikan prosesnya.

Ulangi langkah 2 dan 3 sampai nilai dari titik centroid tidak lagi berubah.

Pembaharuan suatu titik centroid dapat dilakukan dengan rumus berikut:

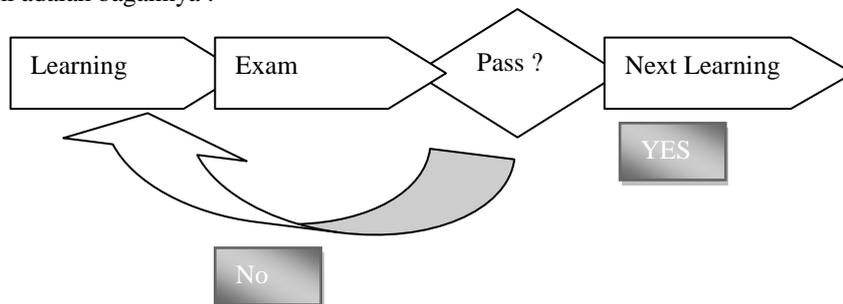
$$m_i = \frac{\sum c_i}{n} \tag{2.5}$$

Di mana:

- mi = titik centroid dari cluster ke-i
- c = cluster
- n = banyaknya data pada cluster

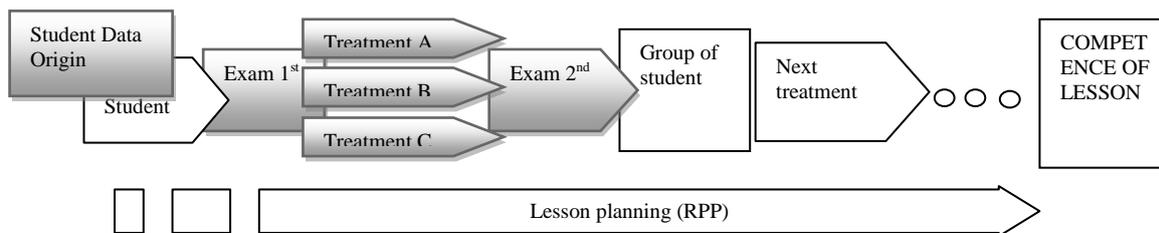
3. Metode Penelitian

Dalam pembelajaran konvensional, peserta didik mengikuti proses yang disebut kegiatan belajar mengajar (*learning*). Setiap target capaian pembelajaran (CP) akan diadakan ujian (*Exam*). Dari setiap ujian ini ada dua kondisi, yaitu lulus (*pass*) atau tidak. Bila lulus maka peserta didik akan menuju ke materi belajar berikutnya (*next learning*) dan bila tidak lulus harus mengulangi ke proses sebelumnya. Berikut ini adalah bagannya :



Gambar 1. Proses pembelajaran yang ada

Selanjutnya dikembangkan suatu pengalian data (*data mining*) dari data hasil ujian dengan peserta didik dalam bentuk cluster. Hasil cluster ini berupa kelompok yang selain menggambarkan prestasi akademik juga menggambarkan potensi akademik, sehingga sistem pembelajaran yang dikembangkan akan menyediakan treatment lanjutan sebelum menuju materi belajar berikutnya. Berikut ini adalah bagannya :



Gambar 2. Sistem yang akan dikembangkan

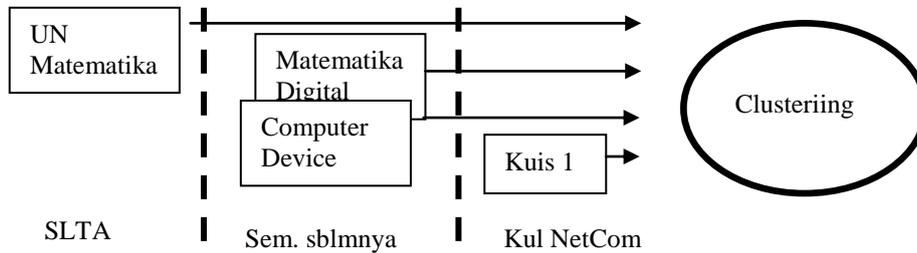
■ L-2

Sebagai studi kasus adalah matakuliah Netcom (*Network And Communication*) pada program studi Teknik Informatika S1 Universitas Muhammadiyah Magelang. Yaitu data pada saat perkuliahan awal, yang melibatkan ujian pertama (*exam 1*) yang akan mempengaruhi penanganan berdasarkan clustering untuk menuju ujian berikutnya (*exam 2*).

4. Hasil dan Pembahasan

Data sampel yang digunakan dalam penelitian ini diperoleh dari Dosen Mata Kuliah *Network and Communication* Universitas Muhammadiyah Magelang. Adapun data yang dijadikan sampel dalam penelitian ini adalah data nilai ujian nasional mata pelajaran matematika mahasiswa dan data nilai kuis pertama mata kuliah *Network and Communication* program studi Teknik Informatika S1 angkatan 2014. Dengan total jumlah data sampel sebanyak 48 data dengan rincian sebagai berikut :

Variabel yang digunakan untuk proses pengclustoran adalah nilai Ujian Nasional mata pelajaran Matematika, diambil dari nilai ijazah mahasiswa, nilai *computer device*, nilai matematika *digital* dan nilai kuis yang diambil dari daftar nilai kuis milik Dosen pengampu mata kuliah *Network and Communication*. Nilai *computer device* dan matematika *digital* disini merupakan nilai matakuliah pada semester sebelumnya yang merupakan prasyarat dari matakuliah *Network and Communication*. Sedangkan nilai kuis merupakan nilai yg didapat dari nilai saat dosen mengadakan obeservasi pada *perkuliahan Network and Communication*.



Gambar 3. Data yang digunakan untuk clustering

Dalam melaksanakan clustering dengan K-mean, centroid ditentukan dulu sebanyak 3, untuk mewakili klasifikasi siswa yang akan dikelola dalam *e-learning*. Ketiga centroid ini dimaksudkan untuk mewakili kualitas, kurang, cukup dan baik. Berikut ini adalah perhitungannya :

Iterasi 1 : Berdasarkan data yang ada dapat di ambil satu contoh untuk proses perhitungan cluster (C), di mana :

$$\begin{aligned}
 K &= 3 \\
 m_1 &= (50, 65, 56, 70) \\
 m_2 &= (76, 78, 77, 75) \\
 m_3 &= (89, 90, 90, 100) \\
 \text{maka :}
 \end{aligned}$$

Pencarian cluster untuk data Pertama.

$$\begin{aligned}
 C_1 &= \sqrt{(Q^1 - 50)^2 + (Q^2 - 65)^2 + (Q^3 - 56)^2 + (Q^4 - 70)^2} \\
 &= \sqrt{(76 - 50)^2 + (78 - 65)^2 + (77 - 56)^2 + (80 - 70)^2} \\
 &= \sqrt{(-26)^2 + (13)^2 + (21)^2 + (10)^2} \\
 &= \sqrt{676 + 169 + 441 + 100} \\
 &= \sqrt{1386} \\
 &= 37,2290
 \end{aligned}$$

Lakukan perhitungan yang sama terhadap Cluster Kedua (C_2) hasil yang didapatkan = 5 dan Cluster Ketiga (C_3) hasil yang didapatkan = 29,6984

Cara di atas dilakukan untuk masing-masing data hingga data terakhir (*End Of File , EOF*)

Berikutnyag BCV (*BetweenClusterVariation*) atau jarak antar cluster menggunakan rumus (2.1).

$$\begin{aligned}
 &= \sqrt{(50 - 76)^2 + (65 - 78)^2 + (56 - 77)^2 + (70 - 75)^2} \\
 &\quad + \sqrt{(76 - 89)^2 + (78 - 90)^2 + (77 - 90)^2 + (75 - 100)^2} \\
 &\quad + \sqrt{(50 - 89)^2 + (65 - 90)^2 + (56 - 90)^2 + (70 - 100)^2} \\
 &= \sqrt{(-26)^2 + (-13)^2 + (-21)^2 + (-5)^2} + \sqrt{(-13)^2 + (-12)^2 + (-13)^2 + (-25)^2} \\
 &\quad + \sqrt{(-39)^2 + (-25)^2 + (-34)^2 + (-30)^2} \\
 &= \sqrt{676 + 169 + 441 + 25} + \sqrt{169 + 144 + 169 + 625} + \sqrt{1521 + 625 + 1156 + 900}
 \end{aligned}$$

$$\begin{aligned}
 &= \sqrt{1311} + \sqrt{1107} + \sqrt{4202} \\
 &= 36,2077 + 33,2176 + 64,8228 \\
 &= 134,3022
 \end{aligned}$$

Hitung WCV (*WithinClusterVariation*) atau jarak antar anggota dalam cluster menggunakan rumus (2.2).

Hasil kuadrat dari penjumlahan nilai terdekat dari masing-masing data, didapatkan hasil WCV = 13.547

Hitung rasio menggunakan rumus (2.3)

$$\text{Rasio} = \frac{134,3022}{13,547} = 0,009914$$

Kemudian tentukan titik centroidbaru menggunakan rumus (2.5)

Perhitungan dilakukan dengan menghitung rata-rata nilai dari masing-masing anggota cluster sebagai berikut:

Centroid 1

$$Q_1 = \frac{105}{2} = 52,50$$

$$Q_2 = \frac{135}{2} = 67,50$$

$$Q_3 = \frac{119}{2} = 59,50$$

$$Q_4 = \frac{115}{2} = 77,50$$

Centroid Baru 1 = (52.50, 67.50, 59.50, 77.50)

Centroid 2

$$Q_1 = \frac{2482}{32} = 77,56$$

$$Q_2 = \frac{2474}{32} = 77,31$$

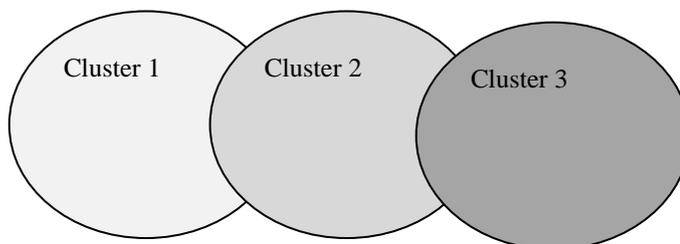
$$Q_3 = \frac{2491}{2} = 77,84$$

$$Q_4 = \frac{2670}{2} = 83,44$$

Centroid Baru 2 = (77.56, 77.31, 77.84, 83.44)

Kemudian ulangi langkah ke 7 sampai 9 untuk menghitung BCV, WCV dan rasio untuk melakukan iterasi ke 3. Bandingkan rasio iterasi ke 3 dengan rasio iterasi ke 2, jika rasio iterasi ke 3 nilainya semakin besar maka lanjutkan ke langkah berikutnya namun jika rasio semakin kecil maka proses berhenti atau selesai.

Dari hasil *clustering* di atas dapat dilihat bahwa cluster1 merupakan kelompok yang merepresentasikan grup bawah, cluster 2 mereprentasikan kelompok menengah dan *cluster* 3 merepresentasikan kelompok atas. Dilihat dari variable nilainya, terlihat bahwa ketiga *cluster* ini tidak sepenuhnya terpisah, setiap clusteringnya bila ditinjau dari rata-rata nilai yang digunakan dalam variabel ini memiliki irisan seperti gambar berikut :



Gambar 4. Algoritma pelaksanaan clustering

Hasil ini dapat dijadikan bahan pertimbangan dalam mengembangkan sistem personalisasi *e-learning*.

4. Simpulan

Untuk menjadi bahan pendukung keputusan dalam mengembangkan personalisasi sistem *e-learning*, metode clustering menggunakan K-mean dapat memisahkan siswa menjadi beberapa kelompok. Namun pemisahan ini bila ditinjau dari variable nilai dan dirata-rata, menghasilkan irisan, sehingga pemisahannya tidak terlihat jelas. Kedepan perlu dicoba metode clustering lain atau kombinasi lebih dari beberapa metode clustering untuk mendapatkan hasil yang maksimal.

Pengaruh matakuliah sebelumnya atau matakuliah prasyarat perlu dipertimbangkan stanndar sistem penilaiannya, karena bila berbeda standar, akan mempengaruhi validitas dari hasil clustering. Sehingga clustering ini haruslah dimulai dengan menstandarkan sistem penilaian dan dimulai dari matakuliah awal, yaitu pada semester satu.

■ **L-2**

Daftar Referensi

- [1] Agus Sasmito Aribowo, "E-Learning Cerdas Dengan Personalisasi Menggunakan Teknik Data Mining Dan Decision Support System," in *Seminar Nasional Informatika 2010 (semnasIF 2010)*, Yogyakarta, 2010, pp. 80-86.
- [2] Li Zheng Chen Xu, "Effectiveness Analysis of The Application of Clustering in Student," in *International Conference on Education Technology and Information System (ICETIS 2013)*, 2013, pp. 988-991.
- [3] S.Ventura C. Romero, "Educational data mining: A survey from 1995 to 2005," *Expert System With Application an Internasional Journal* , vol. 33, no. 1, pp. 135-146, 2007.
- [4] Suryo Guritno, Retantyo Wardoyo, Ahmad Ashari Bernard Renaldy Suteja, "Personalization Ssistem E-Learning Berbasis Ontology," *Makara Sains*, pp. 192-200, 2010.
- [5] A., Peralta, D., Benítez, J.M. , Herrera, F Fernandez, "E-learning and Educational Data Mining in Cloud Computing : An Overview," *Int. J. Learning Technology*, vol. 9, no. 1, pp. 25-52, 2014.
- [6] Narwati, "Pengelompokan Siswa Menggunakan Algoritma K-Means," *Dinamika Informatika*, pp. 12-16, 2010.
- [7] Bertan Y. Badur Osman N. Darcan, "Student Profiling on Academic on Academic Performance Using Cluster Analysis," *Journal of e-Learning & Higher Education*, vol. 2012, p. 8, 2012.
- [8] Hadi Susanto. (2013, May) Wong Kapetakan's Blog. [Online].
<https://bagawanabiyasa.wordpress.com/2013/05/29/penilaian-pengukuran-dan-evaluasi/>
- [9] Purwono Hendradi, "Implementasi Educational Data Mining Untuk Implementasi Kurikulum Berbasis Kompetensi (Kbk)," *Jurnal Sistem Informasi*, pp. 40-43, 2016.
- [10] L.K. Silverman R.M. Felder. (2002) Learning Styles. [Online].
<http://www4.ncsu.edu/unity/lockers/users/f/felder/public/Papers/LS-1988.pdf>
- [11] Suryani Mira, "Pengembangan Personalisasi Gaya Belajar pada E-learning dengan Menggunakan Felder Silverman Learning Style Model untuk Sekolah Menengah Kejuruan (SMK)," in *Conference: Seminar Nasional Pendidikan Teknik Informatik* , Bali, 2014.