

# SISTEM PEMBAGIAN KELAS KULIAH MAHASISWA DENGAN METODE K-MEANS DAN K-NEAREST NEIGHBORS UNTUK MENINGKATKAN KUALITAS PEMBELAJARAN

Gede Aditra Pradnyana<sup>1)</sup>, Agus Aan Jiwa Permana<sup>2)</sup>

<sup>1)</sup>Jurusan Pendidikan Teknik Informatika, Universitas Pendidikan Ganesha  
Jalan Udayana no. 11 Singaraja, Bali

<sup>2)</sup>Jurusan Manajemen Informatika, Universitas Pendidikan Ganesha  
Jalan Udayana no. 11 Singaraja, Bali

e-mail: [gede.aditra@undiksha.ac.id](mailto:gede.aditra@undiksha.ac.id)<sup>1)</sup>, [agus.aan@undiksha.ac.id](mailto:agus.aan@undiksha.ac.id)<sup>2)</sup>

## ABSTRAK

Permasalahan yang terjadi saat pembentukan atau pembagian kelas mahasiswa adalah perbedaan kemampuan yang dimiliki oleh mahasiswa di setiap kelasnya yang dapat berdampak pada tidak efektifnya proses pembelajaran yang berlangsung. Pengelompokan mahasiswa dengan kemampuan yang sama merupakan hal yang sangat penting dalam rangka meningkatkan kualitas proses belajar mengajar yang dilakukan. Dengan pengelompokan mahasiswa yang tepat, mereka akan dapat saling membantu dalam proses pembelajaran. Selain itu, membagi kelas mahasiswa sesuai dengan kemampuannya dapat mempermudah tenaga pendidik dalam menentukan metode atau strategi pembelajaran yang sesuai. Penggunaan metode dan strategi pembelajaran yang tepat akan meningkatkan efektifitas proses belajar mengajar. Pada penelitian ini dirancang sebuah metode baru untuk pembagian kelas kuliah mahasiswa dengan mengkombinasikan metode K-means dan K-Nearest Neighbors (KNN). Metode K-means digunakan untuk pembagian kelas kuliah mahasiswa berdasarkan komponen penilaian dari mata kuliah prasyaratnya. Adapun fitur yang digunakan dalam pengelompokan adalah nilai tugas, nilai ujian tengah semester, nilai ujian akhir semester, dan indeks prestasi kumulatif (IPK). Metode KNN digunakan untuk memprediksi kelulusan seorang mahasiswa di sebuah matakuliah berdasarkan data sebelumnya. Hasil prediksi ini akan digunakan sebagai fitur tambahan yang digunakan dalam pembentukan kelas mahasiswa menggunakan metode K-means. Pendekatan yang digunakan dalam penelitian ini adalah Software Development Live Cycle (SDLC) dengan model waterfall. Berdasarkan hasil pengujian yang dilakukan diperoleh kesimpulan bahwa jumlah cluster atau kelas dan jumlah data yang digunakan mempengaruhi dari kualitas cluster yang dibentuk oleh metode K-means dan KNN yang digunakan. Nilai Silhouette Indeks tertinggi diperoleh saat menggunakan 100 data dengan jumlah cluster 10 sebesar 0,534 yang tergolong kelas dengan kualitas medium structure.

**Kata Kunci:** K-means, K-Nearest Neighbors, pengelompokan mahasiswa.

## ABSTRACT

Problems that occur when the division of student classes is the difference in the ability of students in each class that can impact on the ineffectiveness of the learning process in the class. Grouping students with the same ability is very important in order to improve the quality of teaching and learning process. By grouping the right students, they will be able to help each other in the learning process. Dividing the class of students according to their ability can facilitate the educator in determining the appropriate method or learning strategy, this will improve the effectiveness of the teaching and learning process. This research developed a new method for class division of student lecture by combining K-means and K-Nearest Neighbors (KNN) method. The K-means method is used for class division of student lectures based on the assessment component of the pre-requisite course. The features used in the grouping is the value of the task, the value of the midterm exam, the value of the final exam of the semester, and the cumulative achievement index (GPA). The KNN method is used to predict students' graduation in a course based on previous data. This prediction result will be used as an additional feature used in the formation of student classes using K-means method. The approach used in this research is Software Development Live Cycle (SDLC) with waterfall model. From the test results can be concluded that the number of clusters or classes and the amount of data used affects the quality of clusters formed by the K-means and KNN methods used. The highest value of Silhouette index is obtained when using 100 data with cluster 10 amount of 0,534 (medium structure).

**Keywords:** K-means, K-Nearest Neighbors, student clustering.

## I. PENDAHULUAN

PROSES belajar mengajar yang efektif tentu akan sangat menentukan mutu serta kualitas pendidikan. Berbagai cara dapat dilakukan untuk membuat proses belajar mengajar dalam sebuah institusi pendidikan menjadi lebih efektif, seperti peningkatan fasilitas belajar mengajar, peningkatan kualitas tenaga pengajar, dan pembagian kelas yang terjadwal dengan baik. Permasalahan yang umum terjadi saat pembentukan atau pembagian kelas mahasiswa adalah perbedaan kemampuan yang dimiliki oleh mahasiswa di setiap kelasnya. Pengelompokan mahasiswa dengan kemampuan yang sama merupakan hal yang sangat penting dalam rangka meningkatkan kualitas proses belajar mengajar [1-3]. Dengan pengelompokan mahasiswa di kelas yang sesuai,

mereka akan dapat saling membantu dalam proses pembelajaran [3]. Selain itu, membagi kelas mahasiswa sesuai dengan kemampuannya dapat mempermudah tenaga pendidik dalam menentukan metode atau strategi pembelajaran yang sesuai. Metode dan strategi pembelajaran yang tepat akan meningkatkan efektifitas proses belajar mengajar [2]. Oleh karena itu, pembagian kelas kuliah mahasiswa yang tepat dan sesuai merupakan suatu hal yang sangat penting untuk meningkatkan kualitas pembelajaran.

Jurusan Pendidikan Teknik Informatika merupakan salah satu jurusan yang berada di bawah Fakultas Teknik dan Kejuruan Universitas Pendidikan Ganesha (Undiksha). Dari tahun ke tahun jumlah mahasiswa yang diterima terus mengalami peningkatan. Oleh karena itu, dalam proses perkuliahan mahasiswa pada satu angkatan harus dibagi ke dalam beberapa kelas. Berdasarkan hasil observasi dan wawancara dengan Ketua Jurusan Pendidikan Teknik Informatika, Bapak I Made Gede Sunarya, S.Kom., M.Cs., proses pembentukan kelas yang dilakukan di Jurusan Pendidikan Teknik Informatika masih dilakukan secara acak (*random*). Mahasiswa pada masing-masing semester dikelompokkan secara acak menjadi beberapa kelas, dengan jumlah maksimal 30 orang mahasiswa di setiap kelasnya. Untuk matakuliah yang memiliki sebuah atau lebih matakuliah prasyarat, penguasaan mahasiswa terhadap materi matakuliah prasyarat yang berbeda-beda dalam sebuah kelas kerap menyulitkan pengajar dalam mengelola proses pembelajaran.

*Educational Data Mining* (EDM) merupakan bidang penelitian yang menggunakan teknik-teknik *data mining* untuk lebih memahami proses belajar siswa. EDM akan mengekstrak informasi-informasi tersembunyi dari sekumpulan data pendidikan. Salah satu metode dalam EDM yang kerap digunakan untuk pengelompokan data siswa adalah clustering [4]. Pengelompokan data siswa dengan metode clustering merupakan hal penting dalam bidang EDM dan pembuatan perangkat pembelajaran cerdas (*intelligent learning tools*). Metode *clustering* dapat digunakan untuk mengelompokkan siswa ke dalam kelompok yang homogen (*intra-cluster*) atau heterogen (*inter-cluster*) [5-7].

Penelitian-penelitian yang dilakukan umumnya berfokus pada bagaimana mendapatkan kelas yang sesuai bagi siswa maupun tenaga pendidik. Salah satunya adalah penelitian [2] yang berjudul “*Grouping Students Academic Performance using One-Way Clustering*”. Penelitian yang dilakukan pada sekolah dasar di Selangor Malaysia ini bertujuan untuk membandingkan kinerja dari metode *hierarchical clustering* dan *K-means clustering* dalam pembentukan kelas siswa serta melihat karakteristik siswa dari kelas yang berhasil dibentuk. Dari hasil penelitian yang dilakukan diperoleh kesimpulan bahwa metode *K-means* lebih tepat digunakan untuk mengelompokkan siswa dan menentukan karakter kelompok siswa tersebut. Penelitian serupa sebelumnya telah dilakukan oleh [8] yang berjudul “Pembagian Kelas Kuliah Mahasiswa Menggunakan Pengklasteran Fuzzy”. Penelitian yang dilakukan menggunakan studi kasus Jurusan Teknik Elektro Universitas Diponegoro. Penelitian yang dilakukan menunjukkan bahwa algoritma fuzzy *c-means* yang digunakan dapat membagi kelas kuliah mahasiswa lebih baik dibandingkan dengan algoritma subtraktif. Algoritma subtraktif menghasilkan *cluster* yang sulit diukur tingkatan kemampuan siswa di dalamnya.

Penelitian lain juga dilakukan [1] yang berjudul “*Clustering Student for Group Based Learning in Foreign Language Learning*”. Penelitian yang dilakukan berfokus pada mekanisme pembentukan kelompok siswa otomatis dengan memanfaatkan metode *clustering* untuk pembelajaran bahasa asing. Pada penelitian ini digunakan dua jenis strategi pengelompokan siswa, yaitu *similarity clustering* dan *complementation clustering*. Hasil penelitian yang dilakukan menunjukkan bahwa metode yang dihasilkan lebih fleksibel dan komprehensif dibandingkan dengan pengelompokan secara manual dalam pembelajaran bahasa asing.

Pada penelitian ini akan dikembangkan sebuah sistem baru dalam pembentuk dan pembagian kelas mahasiswa di Jurusan Pendidikan Teknik Informatika Undiksha dengan menggunakan metode *K-means clustering* dan metode *K-nearest neighbors* (KNN). Sistem pembagian kelas kuliah mahasiswa ini digunakan untuk setiap matakuliah yang memiliki matakuliah prasyarat untuk dapat diikuti oleh mahasiswa. Metode *K-means* tepat digunakan dalam pembentukan kelas mahasiswa dengan jumlah kelas yang ditentukan karena jumlah kelompok yang akan terbentuk dengan metode *K-means* akan ditentukan diawal. Pada penelitian ini metode *K-means* digunakan untuk mengelompokkan mahasiswa berdasarkan fitur berupa komponen penilaian dari mata kuliah prasyaratnya. Adapun fitur yang digunakan dalam pengelompokan dengan metode *K-means* adalah nilai tugas, nilai ujian tengah semester, nilai ujian akhir semester, indeks prestasi kumulatif (IPK) serta fitur baru berupa prediksi kelulusan mahasiswa tersebut. Metode KNN digunakan untuk memprediksi kelulusan seorang mahasiswa di sebuah matakuliah berdasarkan data sebelumnya. Hasil prediksi ini akan digunakan sebagai fitur tambahan yang digunakan dalam pembentukan kelas mahasiswa menggunakan metode *K-means*.

## II. METODE PENELITIAN

Pendekatan yang akan digunakan dalam penelitian mengacu pada *System Development Live Cycle* (SDLC) dengan model waterfall. Model *waterfall* merupakan proses pengembangan perangkat lunak secara sekuensial dengan daftar tahapan yang mengalir ke bawah [9]. Tahapan pertama pada penelitian ini adalah melakukan pendefinisian masalah yang ingin diselesaikan. Setelah mendefinisikan masalah yang ingin diselesaikan langkah selanjutnya adalah melakukan pengumpulan data untuk mendukung penyelesaian permasalahan yang dihadapi. Setelah data yang diperlukan terkumpul, data dianalisis sebagai dasar dalam pembuatan sistem. Dalam tahap ini juga didefinisikan kebutuhan-kebutuhan dalam pengembangan sistem. Perancangan dan pengembangan sistem dilakukan setelah kebutuhan-kebutuhan sistem dikumpulkan. Hasil penelitian ini diperoleh dari proses uji coba produk dengan mengacu pada aspek penilaian produk dan penerapannya di Jurusan Pendidikan Teknik Informatika Universitas Pendidikan Ganesha.

Lokasi penelitian adalah Jurusan Pendidikan Teknik Informatika, Universitas Pendidikan Ganesha yang menyediakan objek penelitian seperti data mata kuliah, data ruang kelas, data mahasiswa, dan data nilai mahasiswa. Untuk data mahasiswa dan data nilai mahasiswa diperoleh dari Sistem Informasi Akademik (SIAK) Universitas Pendidikan Ganesha yang terdiri dari data yang akan mengambil mata kuliah serta data yang sudah mengambil mata kuliah untuk digunakan sebagai data *training*.

Perancangan sistem dilakukan setelah semua kebutuhan sistem didapatkan melalui tahap analisis sistem. Pada tahapan ini dilakukan perancangan alur logika sistem yaitu dengan membuat *flow chart* sistem. *Flow chart* sistem bertujuan untuk menggambarkan tahapan logis pada sistem pembentukan dan pembagian kelas kuliah mahasiswa, mulai dari pemrosesan input hingga menjadi output yang diharapkan. Untuk menggambarkan proses-proses yang terjadi di dalam sistem saat melakukan proses pembagian kelas kuliah mahasiswa dengan metode K-means dan KNN dapat digambarkan dalam suatu *flow chart* sistem.

## III. KOMBINASI METODE K-MEANS DAN KNN DALAM PEMBAGIAN KELAS KULIAH MAHASISWA

### A. K-means

K-means merupakan salah satu metode *clustering* non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster* atau kelompok. Metode ini mempartisi ke dalam *cluster* sehingga data yang memiliki karakteristik yang sama (*high intra class similarity*) dikelompokkan ke dalam satu *cluster* yang sama dan yang memiliki karakteristik yang berbeda (*low inter class similarity*) dikelompokkan pada kelompok yang lain [10]. Proses *clustering* dimulai dengan mengidentifikasi data yang akan di-*cluster*,  $X_{ij} (i = 1, \dots, n; j = 1, \dots, m)$  dengan  $n$  adalah jumlah data yang akan dikelompokkan dan  $m$  adalah jumlah variabel. Pada awal iterasi, pusat setiap *cluster* yaitu *centroid*, ditetapkan secara bebas (sembarang),  $C_{kj} (k = 1, \dots, k; j = 1, \dots, m)$ . Kemudian dihitung jarak antara setiap data dengan setiap pusat *cluster*. Untuk melakukan penghitungan jarak data ke- $i$  ( $x_i$ ) pada pusat *cluster* ke- $k$  ( $c_k$ ), diberi nama ( $d_{ik}$ ), dapat digunakan formula Euclidean seperti pada (1)

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2} \quad (1)$$

Suatu data akan menjadi anggota dari *cluster* ke- $k$  apabila jarak dengan pusat *cluster* tersebut merupakan jarak terpendek dibandingkan dengan pusat *cluster* lain menggunakan (2). Selanjutnya, kelompokkan data-data yang menjadi anggota pada setiap *cluster*.

$$\min \sum_{k=1}^k d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2} \quad (2)$$

Nilai pusat *cluster* yang baru dapat dihitung dengan cara mencari nilai rata-rata dari data-data yang menjadi anggota pada *cluster* tersebut, dengan menggunakan rumus pada (3):

$$c_{kj} = \frac{\sum_{i=1}^p x_{ij}}{p} \quad (3)$$

Dimana  $x_{ij} \in$  *cluster* ke- $k$  dan  $p$  adalah banyaknya anggota cluster ke- $k$ .

Algoritma dasar dalam K-means adalah

- Tentukan jumlah *cluster* ( $k$ ), tetapkan pusat *cluster* (*centroid*) sembarang.
- Hitung jarak setiap data ke pusat *cluster* menggunakan (1).
- Kelompokkan data ke dalam *cluster* yang dengan jarak yang paling pendek ke pusat

- cluster* menggunakan (2).
- d. Hitung pusat *cluster* yang baru menggunakan (3) Ulangi langkah b sampai dengan d hingga sudah tidak ada lagi data yang berpindah ke *cluster* yang lain.

**B. K-Nearest Neighbors (KNN)**

Algoritma KNN merupakan metode yang menggunakan algoritma *supervised*. Algoritma *supervised learning* bertujuan untuk mendapatkan pola baru sedangkan *unsupervised learning* tujuannya untuk mendapatkan pola dalam sebuah data. Ide KNN didasarkan pada asumsi lokalitas di ruang data [10]. Prinsip kerja KNN adalah mencari jarak terdekat antara data yang akan dievaluasi dengan *K* tetangga (*neighbor*) terdekatnya dalam data pelatihan. Data *training* diproyeksikan ke ruang berdimensi banyak, yang mana masing-masing dimensi menjelaskan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi data training. Sebuah titik pada ruang ini ditandai kelas *c*, jika kelas *c* merupakan klasifikasi yang paling banyak ditemui pada *K* buah tetangga terdekat titik tersebut [11].

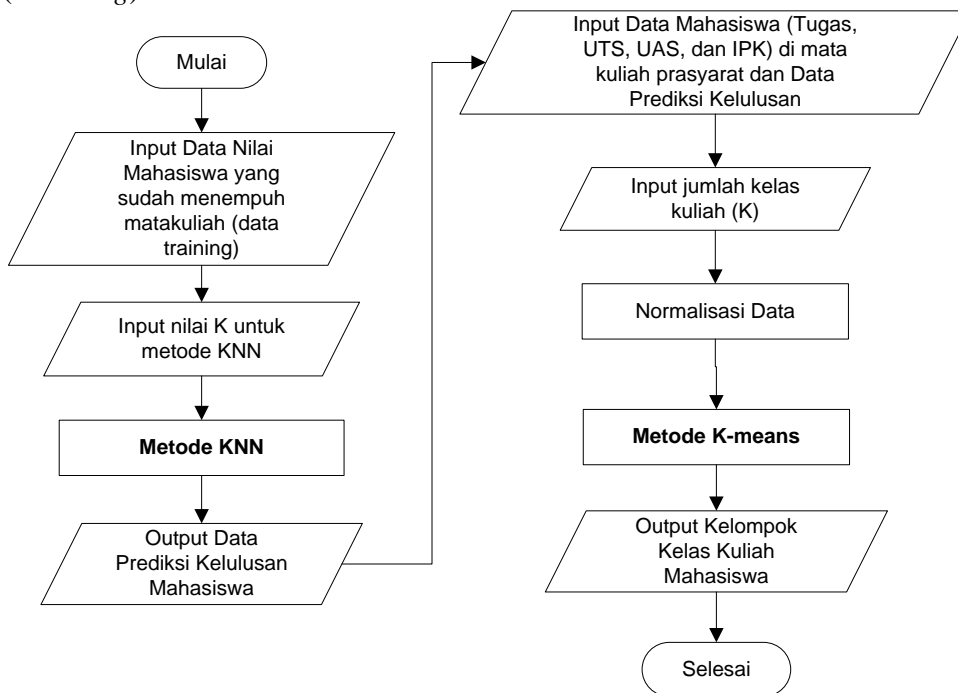
Dalam algoritma ini, nilai *K* yang terbaik itu tergantung pada jumlah data. Ukuran nilai *K* yang besar belum tentu menjadi nilai *K* yang terbaik begitupun juga sebaliknya.

Langkah-langkah untuk menghitung algoritma KNN:

1. Menentukan nilai *K*
2. Menghitung jarak *euclidian* (*query instance*) dengan (1) pada masing-masing objek terhadap *training* data yang diberikan.
3. Kemudian mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak *euclidian* terkecil.
4. Mengumpulkan label class *Y* (klasifikasi *Nearest Neighborhood*).
5. Dengan menggunakan kategori, *Nearest Neighborhood* yang paling mayoritas maka dapat diprediksikan nilai *query instance* yang telah dihitung.

**C. Metode Kombinasi K-means dan KNN untuk Pembagian Kelas Kuliah**

Tahapan-tahapan dari sistem pembagian atau pengelompokan kelas mahasiswa dengan metode K-means dan KNN yang dikembangkan dapat dilihat pada Gambar 1. Tahapan awal adalah memproses data nilai kuliah mahasiswa sebelumnya yang telah mengambil mata kuliah, untuk dijadikan data training dalam proses prediksi dengan metode KNN. Hasil dari proses ini adalah prediksi kelulusan masing-masing mahasiswa yang akan mengambil mata kuliah atau yang akan dikelompokkan ke dalam beberapa kelas. Nilai prediksi kelulusan mahasiswa ini akan digabungkan dengan nilai tugas, ujian tengah semester (UTS), dan ujian akhir semester (UAS) pada mata kuliah prasyarat serta IPK semester mahasiswa untuk dijadikan fitur dalam proses pengelompokan (*clustering*) oleh metode K-means.



Gambar 1. Flow Chart Sistem Pembagian Kelas Kuliah Mahasiswa dengan Metode K-means dan KNN

TABEL 1.  
DATA TRAINING

NIM	Nilai di MK Organisasi Komputer				Status di MK Arsitektur Komputer
	Tugas	UTS	UAS	IPK	
1415051001	90	80	80	3,8	LULUS
1415051002	90	90	85	3,4	LULUS
1415051003	60	75	80	3	LULUS
1415051004	80	90	95	3,8	LULUS
1415051005	90	75	80	3,5	LULUS
1415051006	75	80	75	3,1	TIDAK LULUS
1415051007	50	60	75	2,9	TIDAK LULUS
1415051008	75	60	70	3	TIDAK LULUS
1415051009	90	80	70	3,2	LULUS
1415051010	80	60	70	2,8	TIDAK LULUS

Berikut ini adalah contoh proses yang dilakukan dalam sistem pembagian kelas kuliah yang mahasiswa yang akan dikembangkan:

- a. Misalkan akan dilakukan pembagian kelas mahasiswa untuk mata kuliah Arsitektur Komputer. Pembagian kelas ini akan dilakukan berdasarkan nilai mata kuliah prasyaratnya, yaitu mata kuliah Organisasi Komputer.
- b. Pencarian fitur kelulusan dilakukan dengan memprediksi kelulusan mahasiswa di mata kuliah Arsitektur Komputer berdasarkan data kelulusan sebelumnya. Prediksi ini dilakukan dengan menggunakan metode KNN sebagai berikut:
  - 1) Data *training*, nilai mahasiswa sebelumnya di mata kuliah Organisasi Komputer (mata kuliah prasyarat) dan kelulusannya di mata kuliah Arsitektur Komputer, seperti terlihat pada Tabel 1.
  - 2) Selanjutnya dihitung jarak *euclidian* ( $d$ ) antara data mahasiswa baru yang akan diprediksi dengan setiap data pada data *training*.

NIM	Tugas	UTS	UAS	IPK	Prediksi Kelulusan
1515051033	85	80	80	3	?

$$d(1415051001,1515051033) = \sqrt{(90 - 85)^2 + (80 - 80)^2 + (80 - 80)^2 + (3,8 - 3)^2} = 5,063$$

Perhitungan diatas dilakukan untuk setiap data *training*, sehingga diperoleh nilai  $d$  pada masing-masing data *training*.

- 3) Urutkan nilai  $d$  dari kecil ke besar kemudian tentukan nilai  $k$  untuk memprediksi kelulusannya.

NIM	Nilai di MK Organisasi Komputer				Status di MK Arsitektur Komputer	d
	Tugas	UTS	UAS	IPK		
1415051001	90	80	80	3,8	LULUS	5,063596
1415051005	90	75	80	3,5	LULUS	7,088723
1415051006	75	80	75	3,1	TIDAK LULUS	11,18079
1415051009	90	80	70	3,2	LULUS	11,18213
1415051002	90	90	85	3,4	LULUS	12,25398
1415051004	80	90	95	3,8	LULUS	18,72538
1415051010	80	60	70	2,8	TIDAK LULUS	22,91375
1415051008	75	60	70	3	TIDAK LULUS	24,4949
1415051003	60	75	80	3	LULUS	25,4951
1415051007	50	60	75	2,9	TIDAK LULUS	40,62032

Misalkan  $k$  ditentukan 5 maka prediksi dari mahasiswa tersebut adalah LULUS, yang dilihat dari 5 nilai teratas (data dengan jarak  $d$  terdekat) dengan mayoritas bernilai LULUS.

NIM	Tugas	UTS	UAS	IPK	Prediksi Kelulusan
1515051033	85	80	80	3	LULUS

- 4) Data prediksi kelulusan ini akan menjadi fitur tambahan yang akan digunakan dalam pembagian kelas mahasiswa dengan metode K-means *clustering*.

- 5) Sebelum memasuki proses *clustering*, data dinormalisasi terlebih dahulu agar tidak ada parameter atau fitur yang mendominasi dalam proses *clustering*. Normalisasi data dilakukan dengan persamaan berikut:

$$\text{Nilai Baru} = \frac{\text{Nilai Asal} - \text{Nilai Min}}{\text{Nilai Max} - \text{Nilai Min}}$$

- 6) Setelah melakukan normalisasi data, proses *clustering* diawali dengan menentukan jumlah kelas ( $k$ ) yang akan dibentuk.
- 7) Setelah menentukan jumlah kelas, selanjutnya menentukan  $k$  pusat *cluster* (*centroid*) secara acak.
- 8) Hitung jarak *euclidian* ( $d$ ) dari data yang akan dikelompokkan ke setiap *centroid*. Untuk melakukan penghitungan jarak data ke- $i$  ( $x_i$ ) pada pusat *cluster* (*centroid*) ke- $k$  ( $c_k$ ), diberi nama ( $d_{ik}$ ), dapat digunakan formula *Euclidean* seperti pada (1).
- 9) Kelompokkan data ke dalam *centroid* dengan jarak  $d$  terpendek atau terkecil. Suatu data akan menjadi anggota dari *cluster* ke- $k$  apabila jarak dengan pusat *cluster* tersebut merupakan jarak terpendek dibandingkan dengan pusat *cluster* lain menggunakan (2).
- 10) Hitung kembali *centroid* berdasarkan nilai rata-rata setiap fitur dari data dalam kelompok yang sama. Nilai pusat *cluster* yang baru dapat dihitung dengan cara mencari nilai rata-rata dari data yang menjadi anggota pada *cluster* tersebut, dengan menggunakan rumus pada (3), dimana  $x_{ij} \in$  *cluster* ke- $k$  dan  $p$  adalah banyaknya anggota *cluster* ke- $k$ .
- 11) Lakukan langkah 8-10 sampai tidak terjadi perubahan nilai *centroid* dan atau tidak ada data yang berpindah kelompok.

#### IV. HASIL DAN PEMBAHASAN

##### A. Pengembangan Sistem Pembagian Kelas Berbasis Web

Setelah melakukan perancangan, tahap selanjutnya adalah tahap implementasi atau pengembangan sistem. Pengembangan sistem mengacu pada hasil dari analisis kebutuhan dan proses perancangan. Sistem dikembangkan dengan bahasa pemrograman PHP dan basis data MySQL. Gambaran antar muka dari sistem yang dikembangkan dapat dilihat pada Gambar 2 sampai dengan Gambar 5.

Gambar 2 menunjukkan fitur manajemen data mahasiswa yang akan dilakukan pembagian kelasnya. Penambahan mahasiswa dapat dilakukan secara manual satu persatu atau dengan mengunggah file excel. Data mahasiswa terdiri dari nim, nama, angkatan, dan IPK. Selain menambahkan data mahasiswa, data mata kuliah juga ditambahkan dalam sistem. Data mata kuliah yang perlu disimpan adalah nama kode mata kuliah, nama mata kuliah, serta mata kuliah prasyarat. Setelah itu, data nilai tugas, uts, dan uas dari masing-masing mahasiswa sesuai dengan mata kuliah dimasukkan sesuai pada Gambar 3.

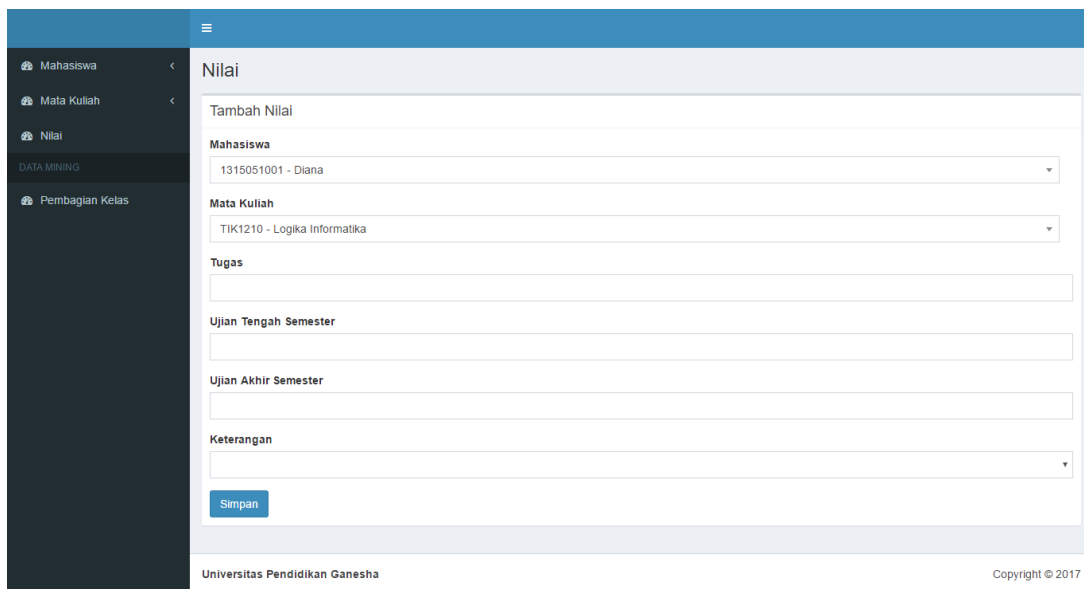
Proses pembagian kelas kuliah dimulai dengan menentukan beberapa parameter, yaitu nilai  $K$  untuk proses algoritma KNN dan nilai  $K$  untuk jumlah kelas yang akan dibentuk seperti pada Gambar 4. Pada proses pembagian kelas juga harus ditentukan data *training* yang digunakan. Proses pembagian kelas dan hasil berturut-turut dapat dilihat pada Gambar 5 dan Gambar 6.

##### B. Pengujian Pengaruh Jumlah Cluster

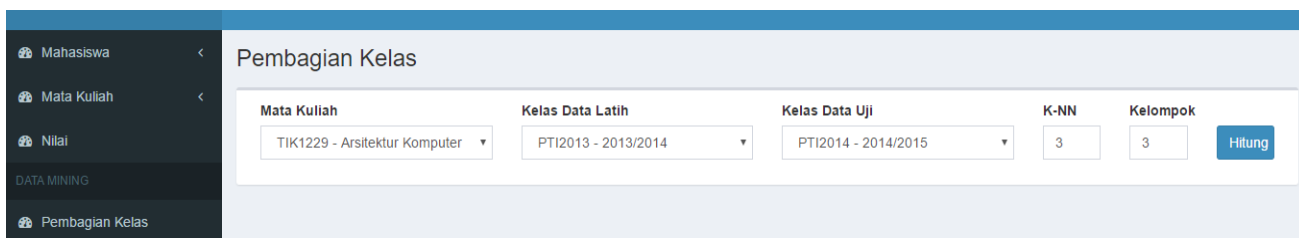
Setelah sistem diimplementasikan dengan baik, langkah selanjutnya adalah melakukan pengujian sistem. Pengujian sistem dilakukan dengan metode *black box testing*. Metode *black box testing* digunakan untuk meyakinkan bahwa fitur-fitur atau fungsi-fungsi yang terdapat dalam sistem yang dibangun sudah berjalan

#	NIM	Nama	Angkatan	IPK	Aksi
1.	1315051001	Diana	2013/2014	2.07	[Edit] [Delete]
2.	1315051002	Dakota	2013/2014	0.92	[Edit] [Delete]
3.	1315051003	Herman	2013/2014	2.06	[Edit] [Delete]
4.	1315051004	Madaline	2013/2014	3.34	[Edit] [Delete]
5.	1315051005	Dalton	2013/2014	0.19	[Edit] [Delete]
6.	1315051006	Maris	2013/2014	0.76	[Edit] [Delete]
7.	1315051007	Galwin	2013/2014	3.29	[Edit] [Delete]
8.	1315051008	Mara	2013/2014	2.61	[Edit] [Delete]
9.	1315051009	Mari	2013/2014	0.83	[Edit] [Delete]

Gambar 2. Manajemen Data Mahasiswa



Gambar 3. Manajemen Nilai Kuliah Mahasiswa



Gambar 4. Proses Penentuan Parameter Pembagian Kelas

**Hasil Pembagian Kelas**

K-Nearest Neighbors | Data Latih

#	NIM	Tugas	Nilai Mata Kuliah Prasyarat (ID/IS)	IPK	Status Kelulusan Mata Kuliah	
		UTS	UAS			
1.	1315051001	32.44	36.31	11.83	2.07	TIDAK LULUS
2.	1315051002	16.76	92.32	21.46	0.92	TIDAK LULUS
3.	1315051003	90.38	95.61	1.81	2.06	LULUS
4.	1315051004	20.77	83.84	59.08	3.34	TIDAK LULUS
5.	1315051005	77.24	34.15	47.28	0.19	TIDAK LULUS
6.	1315051006	24.76	83.83	86.15	0.76	TIDAK LULUS
7.	1315051007	32.04	67.27	75.37	3.29	TIDAK LULUS
8.	1315051008	64.82	23.39	64.34	2.61	TIDAK LULUS
9.	1315051009	4.95	73.17	38.11	0.83	TIDAK LULUS
10.	1315051010	7.00	95.80	44.58	1.47	TIDAK LULUS

K-Means | Data Latih

#	NIM	Tugas	Nilai Mata Kuliah Prasyarat	IPK	Status Kelulusan Mata Kuliah	
		UTS	UAS			
1.	1415051001	22.33	31.65	25.37	1.70	TIDAK LULUS
2.	1415051002	0.61	2.11	1.78	1.07	TIDAK LULUS
3.	1415051003	40.42	24.02	84.29	2.76	TIDAK LULUS
4.	1415051004	14.32	42.61	9.46	0.99	TIDAK LULUS
5.	1415051005	6.60	3.75	25.91	0.32	TIDAK LULUS
6.	1415051006	74.40	92.55	34.37	2.45	TIDAK LULUS
7.	1415051007	23.78	25.97	48.16	1.90	TIDAK LULUS
8.	1415051008	98.49	25.05	95.39	0.38	TIDAK LULUS
9.	1415051009	37.09	91.22	16.46	1.92	TIDAK LULUS
10.	1415051010	29.92	1.62	25.23	1.14	TIDAK LULUS

**K-Means | Hasil**

**Kelas 1**

#	NIM	Nama
1.	1415051001	Nicole
2.	1415051002	Cedric
3.	1415051004	Flavia
4.	1415051005	Brianna
5.	1415051010	Julie
6.	1415051019	Declan
7.	1415051007	Aline
8.	1415051013	Orlando

**Kelas 2**

#	NIM	Nama
1.	1415051003	Chrysette
2.	1415051008	Ira
3.	1415051012	Burke
4.	1415051018	Carlos
5.	1415051020	Byron

**Kelas 3**

#	NIM	Nama
1.	1415051006	Maxwell
2.	1415051009	Quinn
3.	1415051014	Jocelyn
4.	1415051015	Amaya
5.	1415051016	Cotton
6.	1415051017	Aika
7.	1415051011	Ishmael

Gambar 5. Hasil Proses Pembagian Kelas

dengan baik. Setelah tahap pengujian selesai, selanjutnya dilakukan proses evaluasi sistem. Evaluasi sistem yang akan dilakukan berfokus pada evaluasi kinerja dari metode K-means dan KNN dalam membagi kelas kuliah mahasiswa. Evaluasi dilakukan dengan mengamati waktu proses, membandingkan perhitungan manual dengan perhitungan oleh sistem, serta evaluasi kualitas kelas kuliah (*cluster*) yang dihasilkan melalui perhitungan *Silhouette Index* (SI). SI digunakan untuk memvalidasi sebuah data, *cluster* tunggal, atau bahkan keseluruhan *cluster*. Metode ini banyak digunakan untuk memvalidasi *cluster* yang menggabungkan nilai kohesi dan separasi.

*Silhouette Index* (SI) atau *silhouette coefficient* disebut juga digunakan untuk melihat kualitas dan kekuatan *cluster*, seberapa baik suatu objek ditempatkan dalam suatu *cluster*. Metode ini juga digunakan untuk memvalidasi sebuah data, *cluster* tunggal, atau bahkan keseluruhan *cluster*. Metode ini banyak digunakan untuk memvalidasi *cluster* yang menggabungkan nilai *cohesion* dan *separation* [12]. Rentang nilai SI adalah -1 hingga +1. Nilai SI mendekati 1 menunjukkan bahwa data tersebut tidak tepat berada pada *cluster* tersebut. SI bernilai 0

atau mendekati 0 maka posisi data berada pada perbatasan dua *cluster*.

Tahapan perhitungan dari *silhouette index* adalah sebagai berikut:

- a. Hitung rata-rata jarak dari suatu data misalkan *i* dengan semua data lain yang berada dalam satu *cluster*.

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j)$$

dengan *j* adalah data lain dalam suatu *cluster A* dan *d(i,j)* adalah jarak data *i* dengan *j*.

- b. Hitung rata-rata jarak dari data ke-*i* tersebut dengan semua data di *cluster* lain, dan diambil nilai terkecilnya.

$$d(i, C) = \frac{1}{|A|} \sum_{j \in C} d(i, j)$$

dengan *d(i,C)* adalah jarak rata-rata data *i* dengan semua objek pada cluster lain *C* dimana  $A \neq C$

$$b(i) = \min_{C \neq A} d(i, C)$$

- c. Nilai SI dihitung dengan

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

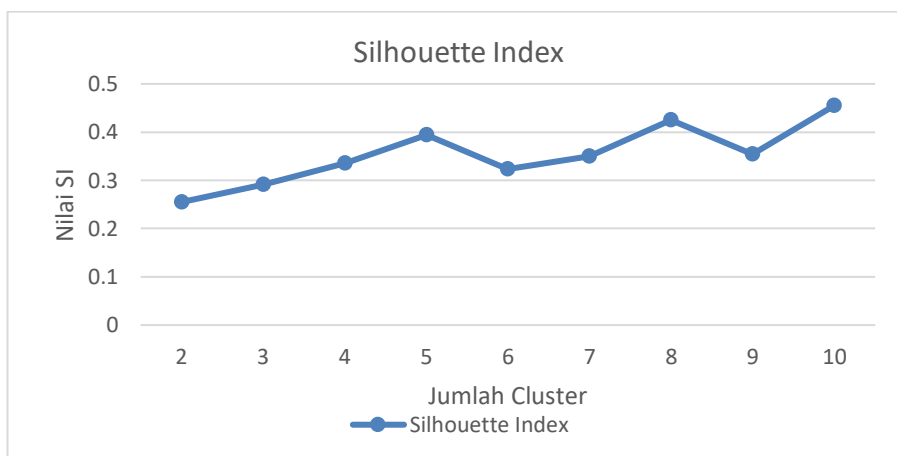
Adapun ukuran nilai SI menurut Kaufman dan Rousseeuw [13] yaitu:

- a.  $0.7 < SI \leq 1$  strong structure
- b.  $0.5 < SI \leq 0.7$  medium structure
- c.  $0.25 < SI \leq 0.5$  weak structure
- d.  $SI \leq 0.25$  no structure

Pengujian pengaruh jumlah *cluster* dilakukan dengan menggunakan 100 data uji untuk mengetahui jumlah *cluster* terbaik dengan melihat dari nilai *Silhouette Index* yang dihasilkan. Nilai *cluster* (*k*) yang akan diujikan adalah nilai *k* 2 sampai 5. Hasil pengujian dapat dilihat pada Tabel 2. Pada hasil proses pengujian, kualitas *cluster* terbaik diperoleh pada penggunaan nilai *k* = 10 dengan nilai rata-rata *Silhouette Index* tertinggi 0.455. Grafik pengaruh nilai *k* ditunjukkan pada Gambar 6.

TABEL 2.  
HASIL PENGUJIAN PENGARUH JUMLAH CLUSTER

Nilai Cluster ( <i>k</i> )	Nilai Silhouette Index ( <i>SI</i> )
2	0,225
3	0,291
4	0,336
5	0,394
6	0,323
7	0,350
8	0,425
9	0,354
10	0,455

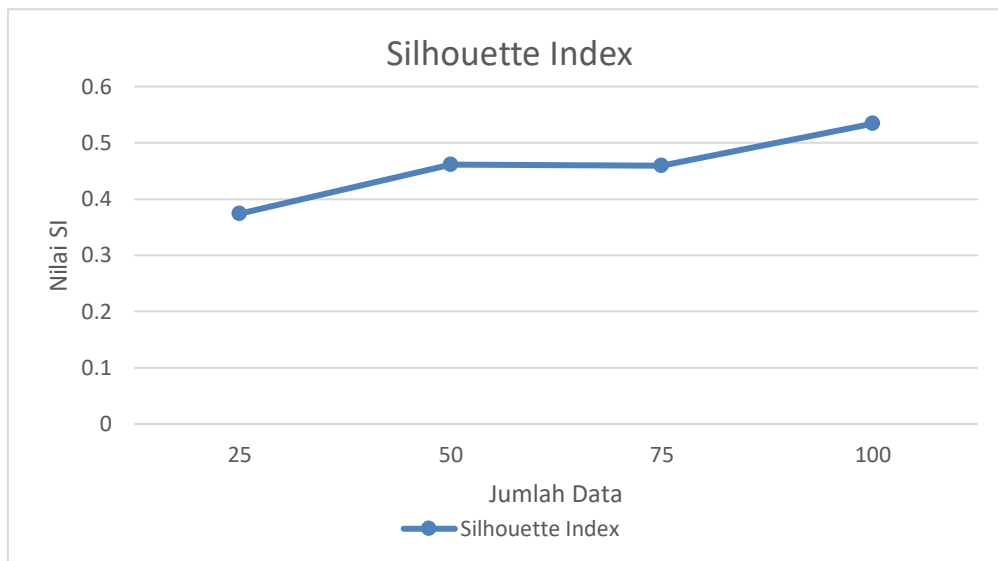


Gambar 6. Grafik Pengaruh Jumlah *Cluster* Terhadap *SI*



TABEL 3.  
PENGUJIAN PENGARUH JUMLAH DATA

<i>Jumlah Data</i>	<i>Nilai Silhouette Index (SI)</i>
25	0,374
50	0,461
75	0,459
100	0,534



Gambar 7. Grafik Pengaruh Jumlah Data Terhadap SI

Hasil pada grafik terlihat semakin banyak jumlah penggunaan nilai  $k$ , maka nilai *Silhouette Index* yang dihasilkan relatif semakin tinggi. Hal ini menunjukkan bahwa semakin banyak jumlah *cluster* atau kelas yang dibentuk akan mengakibatkan jarak *intra cluster* semakin kecil dan jarak *inter cluster* semakin besar yang menunjukkan bahwa kualitas *cluster* yang dibentuk semakin baik. Berdasarkan hasil pengamatan hal ini dapat disebabkan dari karakteristik data yang digunakan. Data mahasiswa antara satu dengan lainnya tidak terlalu berbeda, sehingga diperlukan jumlah *cluster* yang banyak untuk dapat memisahkan dengan baik.

### C. Pengujian Pengaruh Jumlah Data

Pengujian pengaruh jumlah data dilakukan untuk mengetahui pengaruh banyaknya data terhadap hasil dari pengujian sebelumnya. Jumlah data yang digunakan pada pengujian ini yaitu 25, 50, 75, dan 100 dari data sampel yang digunakan pada penelitian ini. Hasil pengujian ditunjukkan pada Tabel 3. Pada hasil proses pengujian, jumlah data terbaik diperoleh pada penggunaan 100 data dengan nilai rata-rata *Silhouette Index* tertinggi 0.534. Grafik pengaruh jumlah data ditunjukkan pada Gambar 7.

Hasil pada grafik menunjukkan semakin banyak data yang digunakan maka nilai *Silhouette Index* yang dihasilkan relatif semakin tinggi. Hal ini menunjukkan bahwa jumlah data berpengaruh terhadap kualitas dari *cluster*. Ketika jumlah data yang digunakan semakin sedikit, maka nilai *Silhouette Index* yang dihasilkan semakin rendah dikarenakan semakin besarnya jarak *inter cluster* dan semakin kecilnya jarak *intra cluster* dimana menggambarkan tingkat kemiripan data dalam satu *cluster* yang besar dan tingkat kemiripan data antar *cluster* yang kecil. Adapun ukuran nilai SI menurut Kaufman dan Rousseeuw [13] yaitu:

- 1)  $0.7 < SI \leq 1$  strong structure
- 2)  $0.5 < SI \leq 0.7$  medium structure
- 3)  $0.25 < SI \leq 0.5$  weak structure
- 4)  $SI \leq 0.25$  no structure

Berdasarkan pengujian diatas diperoleh hasil nilai *Silhouette Indeks* tertinggi adalah 0,534 yang berarti tergolong dalam *medium structure*.

## V. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dapat diperoleh beberapa kesimpulan sebagai berikut:

1. Pada penelitian ini telah berhasil dikembangkan sebuah metode baru dalam pembagian kelas kuliah mahasiswa dengan menggunakan metode K-means *clustering* dan KNN. Metode K-means digunakan sebagai metode pengelompokan siswa ke dalam beberapa kelas dalam suatu mata kuliah berdasarkan nilai tugas, ujian tengah semester, serta ujian akhir semester mahasiswa di mata kuliah prasyaratnya, dan juga IPK semester mahasiswa. Metode K-means tersebut akan dikombinasikan dengan Metode KNN

dalam memprediksi kelulusan mahasiswa di matakuliah terkait, untuk selanjutnya menjadi fitur atau parameter tambahan dalam proses *clustering*. Dari metode tersebut kemudian dikembangkan sebuah sistem berbasis web dengan menggunakan bahasa pemrograman PHP dengan *framework* Laravel.

2. Berdasarkan hasil pengujian yang dilakukan diperoleh kesimpulan bahwa jumlah *cluster* dan jumlah data yang digunakan mempengaruhi dari kualitas *cluster* yang dibentuk oleh metode K-means dan KNN yang digunakan. Kualitas *cluster* dinilai berdasarkan nilai *Silhouette Indeks*. Semakin banyak jumlah *cluster* dan data yang digunakan cenderung akan menghasilkan kualitas *cluster* atau pengelompokan yang semakin baik. Nilai *Silhouette Indeks* tertinggi diperoleh saat menggunakan 100 data dengan jumlah *cluster* 10 sebesar 0,534 (*medium structure*).

Berdasarkan proses penelitian dan temuan yang berhasil diperoleh adapun saran yang dapat diberikan untuk penelitian selanjutnya antara lain:

1. Dapat menggunakan metode lain dalam melakukan pengelompokan kelas mahasiswa misalkan dengan metode *k-medoids* atau menggunakan algoritma *clustering* yang tidak tergolong berbasis partisi seperti *hierarchical clustering*.
2. Dapat menggunakan metode yang menghasilkan kualitas *cluster* yang lebih baik saat membagi dalam kelompok yang tidak terlalu banyak. Hal ini karena pada kasus pembagian kelas mahasiswa sangat jarang dilakukan pembagian kelas lebih dari 5 kelas.

#### DAFTAR PUSTAKA

- [1] Li, L., Luo, X., & Chen, H. (2015). Clustering Students for Group-Based Learning in Foreign Language Learning. *International Journal of Cognitive Informatics and Natural Intelligence*, 55-72.
- [2] Nasir, N. A., Rasid, N. S., & Ahmad, N. (2014). Grouping Students Academic Performance Using One-Way Clustering. *International journal of Science Commerce and Humanities*, 131-138.
- [3] Wang, Y. (2007). On Laws of Work Organization in Human Cooperation. *International Journal of Cognitive Informatics and Natural Intelligence*, 1-15.
- [4] Rana, S., & Garg, R. (2016). Evaluation of Student's Performance of an Institute Using Clustering Algorithms. *International Journal of Applied Engineering Research*, 3605-3609.
- [5] Hamalainen, W., & Kumpulainen, V. (2014). Evaluation of clustering methods for adaptive learning system. *Artificial Intelligence*, 1-23.
- [6] Pradnyana, G. A., & Djunaidy, A. (2013). Metode Weighted Maximum Capturing untuk Klasterisasi Dokumen Berbasis Frequent Itemsets. *Jurnal Ilmu Komputer*, 6(2).
- [7] Pradnyana, G. A. (2012). Perancangan dan Implementasi Automated Document Integration fengan Menggunakan Algoritma Complete Linkage Agglomerative Hierarchical Clustering. *Jurnal Ilmu Komputer*, 5(2).
- [8] Hadi, H. Y., Isnanto, R., & Setiyono, B. (2011). *Pembagian Kelas Kuliah Mahasiswa Menggunakan Algoritma Pengklasteran Fuzzy*. Semarang: Undergraduate Thesis Jurusan Teknik Elektro Fakultas Teknik Undip.
- [9] Bassil, Youssef. (2012). A Simulation Model for the Waterfall Software Development Life Cycle. *International Journal of Engineering & Technology (iJET)*, 2(5).
- [10] Tan, P., Steinbach, & Kumar, V. (2006). *Introduction to Data Mining*. United States: Pearson Education.
- [11] Nanja, M., & Purwanto. (2015). Metode K-Nearest Neighbors berbasis Forward Selection untuk Prediksi Harga Kooditi Loba. *Jurnal Pseudocode*, 53-64.
- [12] Prasetyo, E. (2014). *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi Publisher.
- [13] Kaufman, L., & Rousseuw, P. J. (1990). *Finding Groups in Data*. New York: John Wiley&Sons.