

# DATA MINING DENGAN METODE K-MEANS UNTUK PENGELOMPOKAN MAHASISWA YANG MENGUNJUNGI PERPUSTAKAAN BERDASARKAN DATA KUNJUNGAN DAN IPK

M. Syafrizal Zain<sup>1</sup>, Sarjon Defit<sup>2</sup>, Sumijan<sup>3</sup>

<sup>1</sup> Mahasiswa Magister Ilmu Komputer UPI YPTK Padang

<sup>2</sup> Dosen Magister Ilmu Komputer UPI YPTK Padang

<sup>3</sup> Dosen Magister Ilmu Komputer UPI YPTK Padang

msyafrizalzain@gmail.com

## Abstrak

Data mining merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu database. Salah satu teknik data mining yaitu clustering. Clustering adalah mengelompokkan sejumlah objek ke dalam cluster dimana cluster yang baik adalah cluster yang memiliki tingkat kesamaan yang tinggi antar objek di dalam suatu cluster dan tingkat ketidaksamaan yang tinggi dengan objek cluster yang lainnya. Clustering disini menggunakan salah satu metode algoritma yaitu k-means. Algoritma k-means merupakan algoritma yang akan menghitung jarak dari masing-masing data ke pusat cluster kemudian menghitung rata-rata dari jumlah kelompok sehingga disebut k-means, atau k rata-rata. Dengan menggunakan teknik data mining tersebut maka dapat digunakan untuk mengelompokkan mahasiswa yang mengunjungi perpustakaan. Apakah kunjungannya termasuk rendah, sedang, atau tinggi. Sehingga diharapkan hasil akhir dari penggunaan data mining ini dapat menghasilkan kelompok siswa yang dapat membantu pihak instansi dalam mengambil keputusan.

**.Kata kunci:** Data Mining, Clustering, Algoritma k-means, Perpustakaan

## 1. Pendahuluan

### 1.1. Latar Belakang

Pada saat sekarang, kebutuhan akan informasi sudah mudah didapatkan, baik itu dari media cetak dan media non cetak seperti internet. Kemudahan ini membuat semua orang dapat membaca dimanapun dan kapanpun tanpa harus menggunakan fasilitas yang telah disediakan seperti perpustakaan. Perpustakaan pada saat sekarang telah menjelma menjadi fasilitas digital yang dapat dilihat tanpa harus mengunjungi perpustakaan fisik yang ada. Terlebih dilingkungan kampus, para mahasiswa dapat dengan mudah mencari sekian banyak jurnal dan buku di internet tanpa harus lagi datang ke perpustakaan. Walaupun demikian, fasilitas perpustakaan tetap digunakan karena beberapa aspek kelebihan yang ada. Beberapa kelemahan perpustakaan dalam bentuk digital akan membuat orang tetap memilih perpustakaan biasa, adapun kelemahan tersebut adalah Pertama, tidak semua pengarang mengizinkan karyanya didigitalkan. Pastinya, pengarang akan berpikirkir tentang royalti yang akan diterima bila karyanya didigitalkan. Kedua, masih banyak masyarakat Indonesia yang buta akan teknologi. Apalagi, bila perpustakaan digital ini dikembangkan dalam perpustakaan di pedesaan. Ketiga, masih sedikit pustakawan yang

belum mengerti tentang tata cara mendigitalkan koleksi perpustakaan. Itu artinya butuh sosialisasi dan penyuluhan tentang perpustakaan digital. (Gatot Subrata, 2009)

Aktivitas mahasiswa di perpustakaan memiliki peranan penting dalam menunjang keberhasilan belajar selama mereka duduk di bangku kuliah. Proses pembelajaran pada mahasiswa tidak hanya terjadi pada waktu mereka mendapatkan kuliah saja, melainkan mahasiswa bisa belajar diluar jam kuliah. Perpustakaan dimanfaatkan sebagai salah satu sumber belajar yang mereka butuhkan, baik yang berhubungan dengan perkuliahan atau diluar materi kuliah yang ada. Keberhasilan mahasiswa dalam pendidikannya dapat diukur dari prestasi akademik yang didapatkan, atau bisa disebut dengan index prestasi kumulatif (IPK). Tinggi rendahnya IPK mahasiswa dapat dipengaruhi oleh beberapa faktor. Dalam beberapa penelitian faktor yang mempengaruhi prestasi mahasiswa adalah nilai Unas, jenis kelamin dan kepuasan terhadap fasilitas jurusan atau kampus (Suparto, 2016) serta penelitian lain juga menyebutkan sebagian besar faktor sarana dan fasilitas yang mempengaruhi prestasi belajar mahasiswa (Suprapti, 2015). Fasilitas disini termasuk pada perpustakaan, karena perpustakaan adalah tempat membaca, dan membaca adalah langkah awal untuk menambah wawasan serta pengetahuan, sehingga mahasiswa dapat meraih prestasi terbaiknya.

IAIN Bukittinggi merupakan salah satu institusi yang memiliki perpustakaan sebagai fasilitas pendukung bagi mahasiswanya. Kunjungan mahasiswa ke perpustakaan termasuk tinggi dilihat dari data yang ada pada setiap buku isian masuk ke perpustakaan. Penelitian ini akan mencoba menganalisa kelompok mahasiswa yang berkunjung ke perpustakaan dengan nilai prestasi akademik yang didaparkannya. Apakah terdapat korelasi atau hubungan antara prestasi akademik serta seringnya mahasiswa tersebut ke perpustakaan.

## **1.2. Perumusan Masalah**

Dari latar belakang diatas, maka dapat dirumusan masalah sebagai berikut :

1. Bagaimana penerapan algoritma K-Means untuk mengelompokan data IPK mahasiswa terhadap data kunjungannya ke perpustakaan?
2. Bagaimana kelompok mahasiswa yang akan terbentuk dalam hasil penelitian ini dapat memberikan rekomendasi kepada pihak kampus?
3. Bagaimana pengaruh seringnya mahasiswa mengujungi perpustakaan dengan prestasi akademik yang diperolehnya?

## **1.2. Batasan Masalah**

Dengan luasnya cakupan yang dapat terkait dengan tesis ini, maka terdapat batasan-batasan yang perlu diberlakukan. Batasan-batasan tersebut adalah:

1. Penelitian ini akan mengelompokan mahasiswa berdasarkan data kunjungannya ke perpustakaan dan data IPK dari masing-masing mahasiswa tersebut, sehingga data yang akan digunakan hanya data IPK dan data kunjungan.
2. Untuk menghasilkan informasi yang diperlukan, maka aata yang digunakan merupakan data transkrip nilai mahasiswa yang wisuda April 2016, Oktober 2016 dan April 2017, yang diambil dari bagian Akademik IAIN Bukittinggi, serta data kunjungan dari perpustakaan IAIN Bukittinggi.
3. Metode yang digunakan adalah metode K-Means, dan untuk uji coba serta penampilan hasil dari penelitian ini, maka penulis akan menggunakan Rapid Miner.

#### **1.4. Tujuan Penelitian**

Tujuan dari penelitian ini adalah:

1. Untuk mengetahui apakah ada hubungan antara prestasi mahasiswa dengan seringnya mahasiswa tersebut datang ke perpustakaan, sehingga terlihat perbedaan dan kelompoknya berdasarkan nilai IPK yang didapatkan.
2. Menerapkan metode K-Means untuk melakukan pengelompokan terhadap data mahasiswa.
3. Mendapatkan hasil akhir pengelompokan mahasiswa yang dapat digunakan oleh pihak institusi sebagai acuan pengambilan keputusan.

#### **1.5. Manfaat Penelitian**

Manfaat yang akan diperoleh dalam penelitian ini adalah :

1. Penelitian ini diharapkan dapat membantu IAIN Bukitinggi khususnya bagian perpustakaan untuk mengambil keputusan dari hasil pengelompokan mahasiswa, sehingga kedepannya ada peningkatan yang diperoleh.
2. Memperoleh informasi tentang hubungan prestasi seorang mahasiswa dengan intensitasnya mengunjungi perpustakaan.
3. Dapat dijadikan referensi untuk penelitian selanjutnya.

### **2. Tinjauan Literatur**

#### **2.1 KDD (*Knowledge Discovery in Database*)**

*Knowledge Discovery in Database* (KDD) atau *data mining* adalah area disiplin ilmu untuk mengambil informasi dari data mentah atau metode untuk mencari pola data yang belum diketahui polanya sebelumnya (Nisha Rani, 2016). Tujuan dari data mining adalah untuk menarik pengetahuan abstrak dari sebuah database yang besar. Analisa data diambil dari pola abstrak tersebut dapat melakukan proses pengambilan keputusan dengan sangat mudah (Hari Ram, et al, 2013).

#### **2.2. Data Mining**

Data mining adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam database, data warehouse, atau penyimpanan informasi lainnya (Gunawan Abdillah, et al, 2016). Data mining berkaitan dengan bidang ilmu – ilmu lain, seperti database system, data warehousing, statistik, machine learning, information retrieval, dan komputasi tingkat tinggi. Data mining didefinisikan sebagai proses menemukan pola-pola dalam data. Proses ini otomatis atau seringnya semiotomatis.

#### **2.3. Clustering**

Clustering adalah metode yang mengelompokkan data kedalam cluster, dimana objek dengan kesamaan tinggi berada pada cluster yang sama, tetapi objek yang tidak sama berada pada cluster yang berbeda. Jadi clustering adalah metode pengelompokan objek data kedalam kelompok yang berbeda, seperti objek data yang sama masuk ke cluster yang sama, dan objek data yang berbeda masuk ke cluster berbeda (Naina Pal, et al, 2014). Terdapat banyak algoritma clustering yang dalam penggunaannya tergantung pada tipe data yang akan dikelompokkan dan apa tujuan dari pembuatan aplikasinya. Algoritma tersebut dapat digunakan untuk mengelompokkan objek ke dalam cluster-cluster, kemudian dari hasil clustering akan dideteksi keberadaan outlier dalam data tersebut. Sedangkan data yang digunakan bertipe data numeric.

## 2.4. K-Means Clustering

K-Means merupakan salah satu metode data clustering non hierarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lainnya (Benri MM dan Herlina LS, 2015).

Metode K-Means pertama kali diperkenalkan oleh MacQueen JB pada tahun 1976. Metode ini adalah salah satu metode non hierarki yang umum digunakan. Metode K-Means sangat terkenal karena kemudahannya dan kemampuannya untuk mengelompokkan data besar dan outlier dengan sangat cepat. Dalam metode K-Means setiap data harus termasuk ke cluster tertentu pada suatu tahapan proses, pada tahapan proses berikutnya dapat berpindah ke cluster yang lain. Hasil cluster dengan metode K-Means sangat bergantung pada nilai pusat kelompok awal yang diberikan. Pemberian nilai awal yang berbeda bisa menghasilkan kelompok yang berbeda.

Algoritma K-Means pada awalnya mengambil sebagian dari banyaknya komponen dari populasi untuk dijadikan pusat cluster awal. Pada step ini pusat cluster dipilih secara acak dari sekumpulan populasi data. Berikutnya K-Means menguji masing-masing komponen didalam populasi data dan menandai komponen tersebut ke salah satu pusat cluster yang telah di definisikan tergantung dari jarak minimum antar komponen dengan tiap-tiap pusat cluster. Posisi pusat cluster akan dihitung kembali sampai semua komponen data digolongkan kedalam tiap-tiap cluster dan terakhir akan terbentuk posisi cluster baru. Adapun tujuan dari data clustering ini adalah untuk meminimalisasikan objective function yang diset dalam proses clustering, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu cluster dan memaksimalkan variasi antar cluster (Handi K S dan Sushermanto, 2011).

Tahapan Algoritma K-Means adalah (Nurul Rohmawati W, et al, 2015) :

1. Menentukan k sebagai jumlah cluster yang ingin dibentuk.
2. Membangkitkan nilai random untuk pusat cluster awal (centroid) sebanyak k
3. Menghitung jarak setiap data input terhadap masing-masing centroid menggunakan rumus jarak Euclidian (Euclidian Distance) hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Berikut adalah persamaan Euclidian Distance:

$$d(x_i, \mu_j) = \sqrt{(\sum [(x_i - \mu_j)]^2)} \dots \dots \dots (2.1)$$

Dimana :

- $x_i$  = data kriteria
- $\mu_j$  = centroid pada cluster ke-j

4. Mengklasifikasikan setiap data berdasarkan kedekatannya dengan centroid (jarak terkecil).
5. Memperbaharui nilai centroid. Nilai centroid baru diperoleh dari rata-rata cluster yang bersangkutan dengan menggunakan rumus:

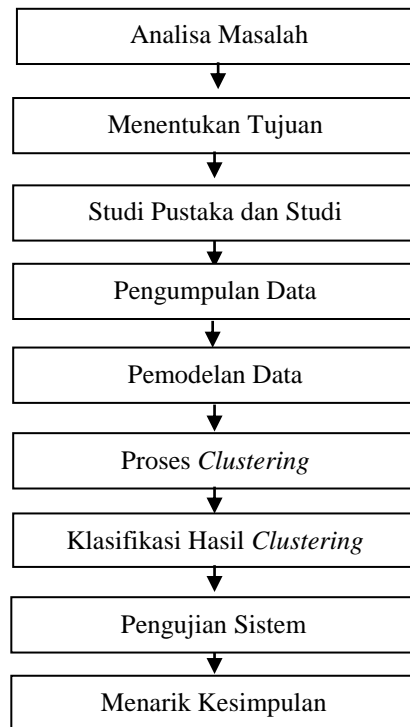
$$\mu_j(t+1) = 1/N_{sj} \sum_{j \in S_j} x_j \dots \dots \dots (2.2)$$

Dimana :

- $\mu_j(t+1)$  = centroid baru pada iterasi ke (t+1)
- $N_{sj}$  = banyak data pada cluster  $S_j$

6. Melakukan perulangan dari langkah 2 hingga 5 hingga anggota tiap cluster tidak ada yang berubah.
7. Jika langkah 6 telah terpenuhi, maka nilai pusat cluster ( $\mu_j$ ) pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data

### 3. Metodologi



Gambar 1. Kerangka Penelitian

#### 3.1. Analisa Masalah

Langkah analisa masalah bertujuan untuk memahami masalah yang ada. Dengan menganalisa masalah, diharapkan masalah dapat dipahami, sehingga kebutuhan sistem dapat diketahui dan ditentukan. Dari menganalisa masalah ini diharapkan hasil akhir berupa laporan yang nantinya dapat membantu mengambil keputusan.

#### 3.2. Menentukan Tujuan

Berdasarkan analisa terhadap masalah, maka ditentukan tujuan yang akan dicapai, terutama yang dapat mengatasi masalah-masalah yang ada.

#### 3.3. Studi Pustaka dan Studi Lapangan

Studi pustaka merupakan pemahaman terhadap berbagai landasan teori yang terkait dengan pelaksanaan penelitian yang akan dilakukan. Penelaahan terhadap literature yang terkait bertujuan antara lain untuk mengetahui prinsi apa itu data mining, prinsip-prinsip dan cara kerja metode algortima K-Means dan juga pembahasan mengenai penelitian terdahulu, yang bersumber baik itu melalui buku-buku, jurnal, dan situs internet yang ada. Sehingga diperoleh suatu pemahaman terhadap tahapan-tahapan dalam penyelesaian permasalahan peneitian.

Setelah melakukan studi pustaka, maka selanjutnya melakukan studi lapangan yaitu melakukan observasi ke tempat penelitian, serta pengumpulan data yang diperlukan untuk menyelesaikan masalah penelitian.

### **3.4. Pengumpulan Data**

Pengumpulan data digunakan untuk mengumpulkan data-data dan informasi-informasi yang diperlukan dalam penyelesaian masalah penelitian. Pengumpulan data pada penelitian ini menggunakan metode pengumpulan data study literature dan telaah dokumen.

Study Literature dilakukan dengan cara mencari bahan materi yang berhubungan dengan permasalahan, perancangan, metode K-Means Clustering, guna mempermudah proses implementasi sistem. Pencarian materi dilakukan melalui pencarian pada buku dan internet.

Telaah dokumen adalah pengumpulan data dengan cara mengumpulkan dan mempelajari dokumen-dokumen yang didapatkan dari pihak perpustakaan IAIN Bukittinggi, dan dari metode ini diperoleh hasil pengumpulan data sebanyak 105 data mahasiswa.

### **3.5. Pemodelan Data**

Pada tahapan ini akan dilakukan pemodelan terhadap data dengan menentukan variabel yang digunakan, kemudian dilanjutkan dengan proses clustering terhadap data. Proses clustering menggunakan data kriteria IPK dan data kunjungan perpustakaan.

### **3.6. Proses Clustering**

Tahap ini akan diterapkan metode K-Means untuk mengelompokkan data. Hasil pengelompokkan ini kemudian akan digunakan untuk melihat kelompok mahasiswa yang mengunjungi perpustakaan.

### **3.7. Klasifikasi Hasil Clustering**

Setelah proses clustering tahap selanjutnya adalah proses klasifikasi. Disini akan terlihat hasil dari clustering yaitu kelompok mahasiswa yang datang ke pustaka apakah berbanding lurus dengan IPK yang diterima.

### **3.8. Pengujian Sistem**

Pengujian sistem sangat penting di lakukan karena kekurangan dan kelebihan dari sistem yang sudah dirancang dapat terlihat disini. Pengujian di lakukan dengan cara melihat apakah semua fungsi dan tahapan yang dilalui dapat diselesaikan oleh sistem yang dibangun, serta sistem dapat menampilkan hasil sesuai rancangan yang telah ditentukan.

### **3.9. Menarik Kesimpulan**

Tahap menarik kesimpulan merupakan tahapan terakhir dalam penelitian. Pada tahap ini hasil akhir dari penelitian sudah terbentuk berupa informasi yang dapat digunakan, dari informasi yang didapatkan ini, maka ditarik kesimpulan yang dapat membantu peneliti untuk mengambil keputusan.

#### 4. Hasil dan Diskusi

**Tabel 4.1 Data awal**

<b>Mahasiswa Ke</b>	<b>Kunjungan</b>	<b>IPK</b>
1	0	2,87
2	2	2,99
3	5	3,24
4	0	2,84
5	9	3,42
6	0	2,78
7	10	3,55
8	1	3,05
9	8	3,28
10	3	3,04
11	7	3,21
12	6	3,00
13	4	2,98
14	0	3,32
15	0	3,14
16	12	3,50
17	7	2,95
18	0	2,98
19	9	3,50
20	6	3,52
21	6	3,11
22	10	3,52
23	0	2,90
24	10	3,67
25	0	3,01
26	9	3,63
27	5	3,02
28	13	3,84
29	0	3,01
30	4	3,09
31	9	3,57
32	12	3,50
33	13	3,62
34	9	3,42
35	10	3,26
36	6	3,28
37	3	3,32
38	5	3,42
39	1	3,02
40	6	3,27
41	8	3,46
42	10	3,44
43	2	3,12
44	13	3,63
45	8	3,38
46	0	2,98
47	2	2,84

48	1	3,03
49	3	3,03
50	11	3,59

Data awal ini akan diproses dengan algoritma *K-Means* adapun langkah-langkahnya adalah sebagai berikut :

1. Menentukan nilai k sebagai jumlah kluster

Cluster yang digunakan berjumlah 3 cluster, yaitu :

- a.  $C_1$  adalah cluster untuk mahasiswa dengan IPK rendah dan kunjungan ke perpustakaan rendah
- b.  $C_2$  adalah cluster untuk mahasiswa dengan IPK sedang dan kunjungan ke perpustakaan sedang
- c.  $C_3$  adalah cluster untuk mahasiswa dengan IPK tinggi dan kunjungan ke perpustakaan tinggi.

2. Menentukan nilai awal centroid

Langkah kedua adalah menentukan pusat cluster secara acak pada data awal yang ada pada tabel diatas.

- a.  $C_1 (2,78;0)$ , diambil dari data mahasiswa ke- 6 sebagai nilai centroid awal untuk cluster 1, yang berarti mahasiswa dengan IPK 2,78 dengan kunjungan ke perpustakaan 0 kali.
- b.  $C_2 (3,27;6)$ , diambil dari data mahasiswa ke- 40 sebagai nilai centroid awal untuk cluster 2, yang berarti mahasiswa dengan IPK 3,27 dengan kunjungan perpustakaan 6 kali.
- c.  $C_3 (3,84;13)$ , diambil dari data mahasiswa ke- 28 sebagai nilai centroid awal untuk cluster 3, yang berarti mahasiswa dengan IPK 3,84 dengan kunjungan ke perpustakaan 13 kali.

3. Menghitung jarak masing-masing objek ke centroid awal

Untuk menghitung jarak objek ke centroid, digunakan rumus Euclidean Distance:

Perhitungan jarak mahasiswa pertama ke pusat cluster pertama adalah:

$$d_{11} = \sqrt{(2,87 - 2,78)^2 + (0 - 0)^2} = 0,09000$$

Perhitungan jarak mahasiswa pertama ke pusat cluster kedua adalah:

$$d_{12} = \sqrt{(2,87 - 3,27)^2 + (0 - 6)^2} = 6,01332$$

Perhitungan jarak mahasiswa pertama ke pusat cluster ketiga adalah:

$$d_{13} = \sqrt{(2,87 - 3,84)^2 + (0 - 13)^2} = 13,03614$$

Perhitungan dilanjutkan ke mahasiswa ke 2 sampai ke mahasiswa ke- 50 dan dihitung jaraknya masing-masing ke pusat cluster pertama seperti contoh mahasiswa ke-1 di atas dan didapatkan hasil seperti terlihat pada tabel 4.2.

4. Melakukan pengelompokan setiap objek kedalam cluster berdasarkan jarak minimumnya ke pusat cluster. Mahasiswa pertama, jarak terkecil dilihat dari nilai minimum dari ketiga cluster dan nilai minimum terletak pada cluster ke-1 dengan nilai 0,09000 sehingga mahasiswa pertama masuk ke cluster 1. Dan untuk mahasiswa ke-2 nilai minimum terletak pada cluster ke-1 juga dengan nilai 2,01099. Untuk cluster dari mahasiswa selanjutnya



dapat dilihat pada tabel 4.2 dan mahasiswa yang termasuk dalam clusternya ditandai dengan simbol ( V ).

**Tabel 4.2 Jarak antar data mahasiswa terhadap clusternya dan cluster dari masing-masing mahasiswa pada iterasi ke-1**

Mahasiswa Ke	C1	C2	C3	C1	C2	C3
1	0,09000	6,01332	13,03614	V		
2	2,01099	4,00979	11,03279	V		
3	5,02112	1,00045	8,02247		V	
4	0,06000	6,01539	13,03840	V		
5	9,02273	3,00375	4,02199		V	
6	0,00000	6,01998	13,04314	V		
7	10,02960	4,00979	3,01398			V
8	1,03581	5,00484	12,02598	V		
9	8,01561	2,00002	5,03126		V	
10	3,01125	3,00880	10,03195		V	
11	7,01319	1,00180	6,03298		V	
12	6,00403	0,27000	7,05022		V	
13	4,00500	2,02092	9,04100		V	
14	0,54000	6,00021	13,01040	V		
15	0,36000	6,00141	13,01883	V		
16	12,02158	6,00441	1,05622			V
17	7,00206	1,04995	6,06565		V	
18	0,20000	6,00700	13,02842	V		
19	9,02875	3,00880	4,01442		V	
20	6,04546	0,25000	7,00731		V	
21	6,00907	0,16000	7,03796		V	
22	10,02734	4,00780	3,01702			V
23	0,12000	6,01140	13,03394	V		
24	10,03953	4,01995	3,00481			V
25	0,23000	6,00563	13,02647	V		
26	9,04005	3,02152	4,00551		V	
27	5,00576	1,03078	8,04192		V	
28	13,04314	7,02317	0,00000			V
29	0,23000	6,00563	13,02647	V		
30	4,01199	2,00808	9,03120		V	
31	9,03461	3,01496	4,00910		V	
32	12,02158	6,00441	1,05622			V
33	13,02711	7,00874	0,22000			V
34	9,02273	3,00375	4,02199		V	
35	10,01151	4,00001	3,05555			V
36	6,02080	0,01000	7,02236		V	
37	3,04821	3,00042	10,01351		V	
38	5,04079	1,01119	8,01102		V	
39	1,02840	5,00625	12,02798	V		
40	6,01998	0,00000	7,02317		V	
41	8,02885	2,00900	5,01442		V	
42	10,02176	4,00361	3,02655			V
43	2,02869	4,00281	11,02354	V		
44	13,02776	7,00925	0,21000			V
45	8,02247	2,00302	5,02112		V	

46	0,20000	6,00700	13,02842	V		
47	2,00090	4,02305	11,04536	V		
48	1,03078	5,00576	12,02731	V		
49	3,01040	3,00958	10,03275		V	
50	11,02978	5,01023	2,01556			V
Total				16	23	11

##### 5. Menentukan centroid baru

Pusat cluster baru ditentukan dari rata-rata masing-masing anggota cluster.

- a. Untuk parameter IPK cluster 1 memiliki 16 anggota, yaitu mahasiswa ke 1, 2, 4, 6, 8, 14, 15, 18, 23, 25, 29, 39, 43, 46, 47, dan 48. Pusat cluster baru untuk kelompok pertama dihitung berdasarkan rata-rata dari ke 16 anggota tersebut, yaitu :

$$C_{1IPK} = (2,87 + 2,99 + 2,84 + 2,78 + 3,05 + 3,32 + 3,14 + 2,98 + 2,90 + 3,01 + 3,01 + 3,12 + 2,98 + 2,84 + 3,03) / 16 = \frac{47,88}{16} = 2,99$$

- b. Untuk parameter IPK cluster 2 memiliki 23 anggota, 3, 5, 9, 10, 11, 12, 13, 17, 19, 20, 21, 26, 27, 30, 31, 34, 36, 37, 38, 40, 41, 45, dan 49 , pusat cluster baru untuk kelompok kedua adalah:

$$C_{2IPK} = (3,24 + 3,42 + 3,28 + 3,04 + 3,21 + 3,00 + 2,98 + 2,95 + 3,50 + 3,52 + 3,11 + 3,63 + 3,02 + 3,09 + 3,57 + 3,42 + 3,28 + 3,32 + 3,42 + 3,27 + 3,46 + 3,38 + 3,03) / 23 = \frac{75,14}{23} = 3,27$$

- c. Untuk parameter IPK cluster 3 memiliki 11 anggota, 7, 16, 22, 24, 28, 32, 33, 35, 42, 44, dan 50, pusat cluster baru untuk kelompok ketiga adalah :

$$C_{3IPK} = (3,55 + 3,50 + 3,52 + 3,67 + 3,84 + 3,50 + 3,62 + 3,26 + 3,44 + 3,63 + 3,59) / 11 = \frac{39,12}{11} = 3,56$$

- d. Untuk parameter kunjungan ke perpustakaan, pusat cluster baru dari masing-masing cluster adalah :

$$C_{1kunjungan} = (0+2+0+0+1+0+0+0+0+0+0+1+2+0+2+1) / 16 = 9 / 16 = 0,6$$

$$C_{2kunjungan} = (5+9+8+3+7+6+4+7+9+6+6+9+5+4+9+9+6+3+5+6+8+8+3) / 23 = 145 / 23 = 6,3$$

$$C_{3kunjungan} = (10+12+10+10+13+12+13+10+10+13+11) / 11 = 124 / 11 = 11,3$$

- e. Mengulangi langkah kedua yaitu menghitung jarak masing-masing objek ke centroid yang baru, pengulangan ini disebut iterasi. Perulangan akan dihentikan jika tidak ada lagi perpindahan anggota setiap cluster ke cluster yang lain. Jika masih ada perpindahan anggota cluster, maka perulangan diteruskan sampai ke iterasi ke-n.

6. Iterasi ke-2

**Tabel 4.3 Jarak antar data mahasiswa terhadap clusternya dan cluster dari masing-masing mahasiswa pada iterasi ke-2**

Mahasiswa Ke	C1	C2	C3	C1	C2	C3
1	0,57568	6,31683	11,29360	V		
2	1,43750	4,31325	9,29001	V		
3	4,44440	1,30463	6,28070		V	
4	0,58281	6,31879	11,29547	V		
5	8,44832	2,69999	2,27681			V
6	0,60130	6,32313	11,29943	V		
7	9,45395	3,70648	1,27274			V
8	0,44126	5,30878	10,28520	V		
9	7,44305	1,69570	3,28438		V	
10	2,43796	3,31213	8,28883	V		
11	6,44117	0,69798	4,28674		V	
12	5,43751	0,40484	5,30200		V	
13	3,43752	2,32215	7,29553		V	
14	0,65089	6,30457	11,27521	V		
15	0,58152	6,30563	11,28041	V		
16	11,44875	5,70042	0,72945			V
17	6,43764	0,76446	4,31554		V	
18	0,56264	6,31088	11,28745	V		
19	8,45275	2,70571	2,27343			V
20	5,46303	0,39580	5,27285		V	
21	5,43877	0,34244	5,29159		V	
22	9,45223	3,70431	1,27325			V
23	0,57005	6,31502	11,29182	V		
24	9,46179	3,71756	1,27779			V
25	0,56277	6,30958	11,28596	V		
26	8,46155	2,71999	2,27392			V
27	4,43759	1,32752	6,29562		V	
28	12,46634	6,72013	1,75041			V
29	0,56277	6,30958	11,28596	V		
30	3,43888	2,31113	7,28766		V	
31	8,45724	2,71263	2,27277			V
32	11,44875	5,70042	0,72945			V
33	12,45332	6,70495	1,72844			V
34	8,44832	2,69999	2,27681			V
35	9,44129	3,69566	1,30678			V
36	5,44510	0,30463	5,27996		V	
37	2,45940	3,30477	8,27610	V		
38	4,45804	1,31330	6,27421		V	
39	0,43836	5,31009	10,28672	V		
40	5,44458	0,30436	5,28050		V	
41	7,45218	1,70661	3,27415		V	
42	9,44810	3,69970	1,27804			V
43	1,44314	4,30686	9,28299	V		
44	12,45383	6,70549	1,72884			V
45	7,44759	1,69942	3,27748		V	
46	0,56264	6,31088	11,28745	V		

47	1,44557	4,32547	9,30036	V		
48	0,43910	5,30964	10,28620	V		
49	2,43779	3,31283	8,28946	V		
50	10,45459	4,70675	0,27479			V
Total				19	15	16

**Tabel 4.4 Pusat cluster pada iterasi ke-2**

Cluster ke	IPK	Kunjungan
1	3,01	0,9
2	3,21	6,1
3	3,54	10,6

7. Iterasi ke-3

**Tabel 4.5 Jarak antar data mahasiswa terhadap clusternya dan cluster dari masing-masing mahasiswa pada iterasi ke-3**

Mahasiswa Ke	C1	C2	C3	C1	C2	C3
1	0,95828	6,07641	10,58381	V		
2	1,05291	4,07283	8,58023	V		
3	4,05892	1,06698	5,57065		V	
4	0,96325	6,07818	10,58575	V		
5	8,06285	2,94056	1,56720			V
6	0,97589	6,08217	10,58990	V		
7	9,06847	3,94766	0,56257			V
8	0,06365	5,06932	9,57511	V		
9	7,05764	1,93446	2,57578		V	
10	2,05279	3,07160	7,57909	V		
11	6,05580	0,93334	3,57787		V	
12	5,05265	0,22414	4,59449		V	
13	3,05282	2,07987	6,58646		V	
14	0,99550	6,06759	10,56482	V		
15	0,95568	6,06712	10,57012	V		
16	11,06330	5,94022	1,43809			V
17	6,05297	0,96995	3,61123		V	
18	0,94799	6,07118	10,57740	V		
19	8,06727	2,94724	1,56304			V
20	5,07788	0,31318	4,56255		V	
21	5,05354	0,12353	4,58284		V	
22	9,06675	3,94522	0,56290			V
23	0,95423	6,07479	10,58195	V		
24	9,07635	3,95968	0,57705			V
25	0,94738	6,07010	10,57585	V		
26	8,07614	2,96268	1,56502			V
27	4,05264	1,08417	5,58687		V	
28	12,08089	6,96154	2,45574			V
29	0,94738	6,07010	10,57585	V		
30	3,05357	2,07038	6,57800		V	
31	8,07179	2,95486	1,56276			V
32	11,06330	5,94022	1,43809			V

33	12,06785	6,94521	2,43877			V
34	8,06285	2,94056	1,56720			V
35	9,05597	3,93360	0,62889			V
36	5,05962	0,09381	4,56997		V	
37	2,07528	3,06850	7,56574	V		
38	4,07290	1,08638	5,56382		V	
39	0,05295	5,07038	9,57670	V		
40	5,05910	0,08707	4,57056		V	
41	7,06671	1,94892	2,56379		V	
42	9,06264	3,93982	0,57154			V
43	1,05793	4,06775	8,57286	V		
44	12,06835	6,94580	2,43912			V
45	7,06211	1,94045	2,56757		V	
46	0,94799	6,07118	10,57740	V		
47	1,06695	4,08383	8,59117	V		
48	0,05495	5,07001	9,57616	V		
49	2,05269	3,07218	7,57976	V		
50	10,06911	4,94764	0,44021			V
Total				19	15	16

**Tabel 4.6 Pusat cluster pada iterasi ke-3**

Cluster ke	IPK	Kunjungan
1	3,01	0,9
2	3,21	6,1
3	3,54	10,6

Pengulangan dihentikan karena hasil perhitungan menunjukkan adanya angka pusat cluster yang sama pada iterasi ke-2 dan ke-3. Dapat disimpulkan bahwa tidak ada lagi anggota cluster yang berpindah.

Pusat cluster terakhir akan diambil menjadi hasil dari proses *K-Means* untuk pengelompokan mahasiswa ke perpustakaan pada IAIN Bukittinggi, dan hasilnya :

- a) Cluster 1 adalah mahasiswa dengan IPK rendah dan kunjungan rendah dengan nilai rata-rata IPK 3,01 dan kunjungan 0 - 1 kali.
- b) Cluster 2 adalah mahasiswa dengan IPK sedang dan kunjungan sedang dengan nilai rata-rata IPK 3,21 dan kunjungan rata-rata 6 kali.
- c) Cluster 3 adalah mahasiswa dengan IPK tinggi dan kunjungan tinggi dengan nilai rata-rata IPK 3,54 dan kunjungan rata 10 kali.

**Tabel 4.7 Hasil Clustering**

Cluster	Anggota	Jumlah	Kelompok
1	mahasiswa ke - 1, 2, 4, 6, 8, 10, 14, 15, 18, 23, 25, 29, 37, 39, 43, 46, 47, 48, dan 49	19	Rendah
2	Mahasiswa ke - 3, 9, 11, 12, 13, 17, 20, 21, 27, 30, 36, 38, 40, 41, dan 45	15	Sedang
3	Mahasiswa ke - 5, 7, 16, 19, 22, 24, 26, 28, 31, 32, 33, 34, 35, 42, 44, dan 50	16	Tinggi

Tabel 4.8 diatas memperlihatkan kelompok mahasiswa yang terbentuk dari hasil implementasi *k-means*, sehingga kelompok mahasiswa yang mengunjungi perpustakaan IAIN bukittinggi terbagi menjadi 3 yaitu kelompok 1 dengan anggota berjumlah 19, yaitu kelompok dengan kunjungan ke perpustakaan rendah, dan IPK rendah, kelompok 2 dengan anggota berjumlah 15, merupakan kelompok dengan kunjungan sedang dan IPK sedang, sedangkan kelompok 3 dengan anggota berjumlah 16, merupakan kelompok dengan kunjungan tinggi dan IPK tinggi.

Setelah melalui semua proses k-means secara manual, dilanjutkan dengan pengujian menggunakan aplikasi *RapidMiner* dengan menambah 55 data baru dan total menjadi 105 data, dan dapat diambil sebuah kesimpulan dan fakta bahwa kunjungan mahasiswa ke perpustakaan dapat mempengaruhi prestasi yang akan didapatkannya. Analisa ini didapat dengan melihat hasil dari *centroid* akhir pada pengujian data 105 orang mahasiswa yang dapat dilihat pada tabel dibawah ini.

Tabel 5.4 Pusat cluster akhir 105 data

Cluster ke	IPK	Kunjungan
1	3,09	1
2	3,24	5,9
3	3,55	10,4

Tabel diatas memperlihatkan hasil uji dari 105 data mahasiswa, dan hasilnya semakin sering mahasiswa tersebut ke perpustakaan, maka ada peningkatan terhadap IPK yang diperoleh. Mahasiswa yang mengunjungi perpustakaan sekali, maka rata-rata IPK nya adalah 3,09, jika kunjungan ditingkatkan menjadi 5 sampai 6 kali, rata-rata IPK nya adalah 3,24, dan jika mahasiswa lebih sering mengunjungi perpustakaan 10 kali, maka rata-rata IPK yang diperoleh lebih tinggi yaitu 3,55.

Tabel 5.5 Anggota Masing-masing cluster 148 data

Cluster	Anggota	Jumlah	Kelompok
1	Mahasiswa ke – 37, 39, 43, 46, 47, 48, 49, 51, 52, 53, 56, 57, 58, 60, 64, 67, 73, 79, 80, 83, 84, 89, 90, 94, 95, 96, 97, 98, 99, 100, 101, 103, 104, 105	46	Rendah
2	Mahasiswa ke – 3, 9, 11, 12, 13, 17, 20, 21, 27, 30, 36, 38, 40, 41, 45, 54, 55, 61, 62, 63, 66, 68, 69, 71, 72, 74, 76, 77, 78, 81, 82, 85, 86, 87	34	Sedang
3	Mahasiswa ke- 5, 7, 16, 19, 22, 24, 26, 28, 31, 32, 33, 34, 35, 42, 44, 50, 59, 65, 70, 75, 88, 91, 92, 93, 102	25	Tinggi

Kemudian untuk anggota masing-masing *cluster* terlihat jumlah anggota pada *cluster* 1 dan *cluster* 2 cukup tinggi, sedangkan *cluster* 3 rendah, ini menunjukkan bahwa masih kurangnya minat mahasiswa untuk mengunjungi perpustakaan.

Untuk bagian administrasi perpustakaan, hasil ini dapat bermanfaat untuk menimbulkan keinginan terhadap mahasiswa untuk lebih sering mengunjungi perpustakaan, apalagi bagi mahasiswa yang termasuk kedalam kelompok rendah dan sedang, sehingga kedepannya prestasi yang diperoleh oleh mahasiswa pada kelompok rendah dan sedang dapat meningkat sesuai apa yang mereka inginkan.

## 5. Kesimpulan

Berdasarkan pembahasan, analisa, dan pengujian dari bab sebelumnya, maka dari penelitian ini dapat ditarik kesimpulan sebagai berikut:

1. Penerapan metode K-Means dapat melakukan pengelompokan data mahasiswa yaitu kelompok dengan kunjungan rendah, kunjungan sedang, dan kunjungan tinggi.
2. Hasil penerapan k-means dapat menentukan hubungan antara prestasi mahasiswa dengan kunjungan mahasiswa ke perpustakaan, hasil dari uji data sebanyak 105 mahasiswa, dihasilkan hubungan antara prestasi dan kunjungan mahasiswa, yaitu kunjungan rendah IPK rendah, kunjungan sedang IPK sedang, kunjungan tinggi IPK tinggi.
3. Hasil pengelompokan mahasiswa dapat digunakan oleh pihak institusi sebagai acuan pengambilan keputusan untuk peningkatan kunjungan bagi mahasiswa yang masuk kelompok sedang dan rendah.

## Referensi

- Benri M M dan Herlina L S, 2015** “*Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhila*”. **Jurnal Media Infotama Vol. 11 No. 2. Hal 110-118**
- Gatot Subrata, 2009** “*Perpustakaan Digital*”. **Artikel Pustakawan Perpustakaan Univeritas Negeri Malang**
- Gunawan Abdillah, 2016** “*Penerapan Data Mining Pemakaian Air Pelanggan Untuk Menentukan Klasifikasi Potensi Pemakaian Air Pelanggan Baru Di Pdam Tirta Raharja Menggunakan Algoritma K-Means*”. **Seminar Nasional Teknologi Informasi dan Komunikasi 2016 (SENTIKA 2016) Yogyakarta. Hal 498 – 506**
- Handi K S dan Suhermanto , 2011** “*Penerapan Algoritma K-Means Untuk Menetapkan Kelompok Mutu Karet*”. **Progresif, Vol. 7, No. 2, Agustus 2011. Hal 747 – 754**
- Hari Ram, et al, 2013** “*Improving efficiency of Apriori algorithm using transaction reduction*” **International Journal of Scientific and Research Publications, Volume 3, Issue 1, January 2013**
- Naina Pal, et al, 2014** “*Clustering of Image Data Using K-Means and Fuzzy K-Means*”. **(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 5, No. 7. Hal 160-163**
- Nisha Rani, 2016** “*Knowledge Discovery in Database*”. **International Journal of Enhanced Research in Management & Computer Applications ISSN: 2319-7471, Vol. 5 Issue 6, June-2016**



**Nurul Rohmawati W, et al, 2015** *“Implementasi Algoritma K-Means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa”*. **Jurnal Ilmiah Teknologi Informasi Terapan Volume I, No 2, 30 April 2015. Hal 62 – 68**

**Suparto, 2016** *“Analisis Faktor-Faktor Yang Mempengaruhi Indeks Prestasi ( IP ) Mahasiswa ITATS Jurusan Teknik Industri”*. **SEMINAR NASIONAL INOVASI DAN APLIKASI TEKNOLOGI DI INDUSTRI (SENIATI) 2016**

**Suprpti, 2015** *“Faktor Eksternal Yang Mempengaruhi Tingginya Prestasi Belajar Mahasiswa Tingkat Ii Di Stikes Widya Dharma Husada Tangerang”*. **Jurnal Kesehatan dan Budaya HIKMAH. Volume 08 No. 01. Juni 2015**