

PENERAPAN DATA MINING NILAI PENJAJAKAN UJIAN NASIONAL MENGGUNAKAN METODE CLUSTERING STUDI KASUS DI SMP “X” DI KOTA SEMARANG

Sulastri¹, Arief Jananto², Sri Eniyati³

^{1,2}Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Stikubank

³Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Stikubank

e-mail: ¹ sul4stri66@gmail.com, ² ajananto09@gmail.com, ³ eniyati03@gmail.com

ABSTRAK

Untuk menghadapi Ujian Nasional memerlukan persiapan yang baik agar siswa mendapat nilai yang maksimal. Salah satu cara untuk menyiapkan siswanya, sekolah dan dinas terkait melakukan Penjajakan Ujian Nasional (PUN). SMP “X” di kota Semarang juga melakukan PUN, nilai PUN yang diperoleh kemudian diranking untuk pengelompokan kelas tambahan. Tiap kelas tambahan mempunyai 30 siswa. Siswa yang ranking 1-30 menempati kelas A, siswa yang ranking 31-60 di kelas B dan seterusnya.

Tujuan penelitian ini menganalisa nilai PUN dari 288 siswa dengan menggunakan metoda K-mean untuk mendapatkan cluster-cluster kelas tambahan sesuai dengan kebutuhan. Nilai PUN yang dianalisa adalah nilai Bahasa Indonesia, Matematika, Bahasa Inggris dan IPA. Dari analisa yang sudah dikerjakan didapat bahwa jumlah cluster dapat ditentukan sesuai kebutuhan bidang akademik dan tiap cluster mempunyai cacah anggota yang berbeda.

Kata kunci: *Penjajakan Nilai Ujian, Clustering, K-Means*

1. PENDAHULUAN

Pendidikan merupakan bagian yang sangat penting dalam kehidupan masyarakat. Salah satu komponen yang ada pada sistem pendidikan tersebut adalah evaluasi yang dapat dilakukan oleh guru, administrator pendidikan, atau siapapun yang mempunyai komitmen untuk memajukan dunia pendidikan. Kegiatan evaluasi harus dilakukan berdasarkan prinsip-prinsip dan prosedur evaluasi yang benar. Evaluasi hasil belajar dilakukan untuk mengetahui tingkat pencapaian prestasi belajar siswa setelah mengikuti kegiatan pembelajaran. Salah satu contoh bentuk evaluasi adalah ujian nasional (UN).

Di setiap sekolah sebelum dilaksanakan Ujian Nasional (UN), sekolah-sekolah tersebut melakukan Penjajakan Ujian Nasional (PUN), baik yang dilakukan oleh pihak sekolah maupun oleh Dinas Pendidikan. Kegiatan tersebut bertujuan untuk melatih para siswa dalam menghadapi Ujian Nasional (UN) yang sesungguhnya. Sehingga pada waktu siswa menghadapi Ujian Nasional (UN) dapat mengerjakan dengan baik dan mendapatkan nilai yang maksimal. SMP “X” kota Semarang merupakan SMP Negeri favorit di kota Semarang yang selalu mempunyai nilai tertinggi pada setiap UN. Tentunya menjadi permasalahan setiap tahunnya yaitu bagaimana mempertahankan prestasi tersebut. Untuk itu diperlukan suatu strategi di dalam pelaksanaan UN agar dapat mempertahankan prestasinya. Salah satu strategi yang digunakan yaitu melakukan perangkingan terhadap nilai Penjajakan Ujian Nasional (PUN). Perangkingan ini berdasarkan jumlah nilai yang diperoleh siswa dan kemudian digunakan untuk pembagian kelas jam tambahan. Jam tambahan bertujuan memberi tambahan materi terhadap 4 mata pelajaran yang diUNkan. Kemudian kelas jam tambahan dibagi menjadi 9 kelas dari A sampai dengan I. Kelas A ditempati oleh siswa-siswa yang mempunyai ranking 1-30, kelas B oleh siswa yang ditempati oleh anak yang ranking 31-60 dan seterusnya. Di kelas A, siswa-siswa diberi latihan cukup banyak dan cepat serta kelas ini ditarget mempunyai nilai UN minimal 36. Permasalahan yang timbul apakah perangkingan yang dilakukan sudah sesuai dengan tingkat kemampuan siswa-siswa yang sesungguhnya. Karena jika dilihat beberapa nilai Penjajakan Ujian Nasional (PUN), ada beberapa siswa kelompok jam tambahan tidak stabil/berpindah-pindah.

Data mining merupakan serangkaian proses eksplorasi dan analisis dalam sebuah basis data, agar memiliki nilai tambah berupa informasi yang sebelumnya tidak diketahui atau tersembunyi dari suatu basis data. Konsep Data mining merupakan konsep yang berasal dari pengembangan berbagai bidang ilmu, antara lain basis data, information retrieval, statistika, machine learning dan sebagainya.

Data Mining memiliki banyak metode dengan berbagai kegunaan masing-masing. Data mining dapat diterapkan di berbagai bidang, seperti bisnis, kesehatan, asuransi, pemasaran dan perbankan.

Data mining mendukung task/fungsionalitas yang meliputi :

1. Prediktive

Menghasilkan model berdasarkan sekumpulan data yang dapat digunakan untuk memperkirakan nilai data yang lain. Metode yang termasuk prediktive data mining :

- a. Klasifikasi : pembagian data ke dalam beberapa kelompok/kelas yang telah ditentukan sebelumnya
- b. Regresi : memetakan data ke suatu prediction variable
- c. Time Series Analysis : pengamatan perubahan nilai atribut dari waktu ke waktu

2. Deskriptive

Mengidentifikasi pola atau hubungan dalam data untuk menghasilkan informasi baru. Metode yang termasuk deskriptive data mining :

- a. Clustering : mengelompokkan beberapa objek yang serupa ke dalam sebuah cluster, dan yang tidak serupa ke cluster yang lain
- b. Association rules : identifikasi hubungan antara data yang satu dengan yang lainnya.
- c. Summarization : pemetaan data ke dalam subset dengan deskripsi sederhana.
- d. Sequence discovery : identifikasi pola sekuensial dalam data

K-means merupakan salah satu algoritma clustering. Tujuan algoritma ini yaitu untuk membagi data menjadi beberapa kelompok/cluster. Algoritma ini menerima masukan berupa data tanpa label kelas. Hal ini berbeda dengan supervised learning yang menerima masukan berupa vektor $(-x-1, y1)$, $(-x-2, y2)$, ..., $(-x-i, yi)$, di mana x_i merupakan data dari suatu data pelatihan dan y_i merupakan label kelas untuk x_i .

Pada algoritma pembelajaran ini, komputer mengelompokkan sendiri data-data yang menjadi masukannya tanpa mengetahui terlebih dulu target kelasnya. Pembelajaran ini termasuk dalam unsupervised learning. Masukan yang diterima adalah data atau objek dan k buah kelompok (cluster) yang diinginkan. Algoritma ini akan mengelompokkan data atau objek ke dalam k buah kelompok tersebut. Pada setiap cluster terdapat titik pusat (centroid) yang merepresentasikan cluster tersebut.

Algoritma untuk melakukan k-Means clustering adalah sebagai berikut:

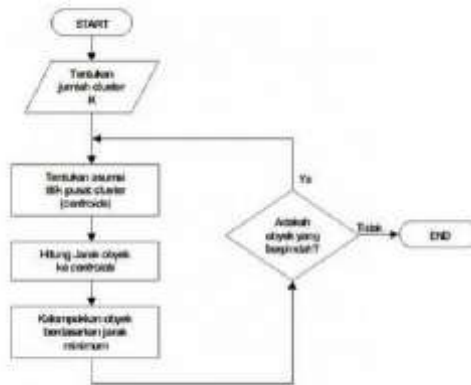
1. Pilih K buah titik centroid secara acak
2. Kelompokkan data sehingga terbentuk K buah cluster dengan titik centroid dari setiap cluster merupakan titik centroid yang telah dipilih sebelumnya
3. Perbaharui nilai titik centroid
4. Ulangi langkah 2 dan 3 sampai nilai dari titik centroid tidak lagi berubah

Proses pengelompokkan data ke dalam suatu cluster dapat dilakukan dengan cara menghitung jarak terdekat dari suatu data ke sebuah titik centroid.

2. METODE PENELITIAN

Obyek penelitian ini adalah nilai Penjajakan Ujian Nasional (PUN) siswa SMP "X" di Kota Semarang yang terangkum dalam 288 record dengan atribut No urut, No Peserta, Nama Siswa, Jenis Kelamin, Kelas, Nilai Bahasa Indonesia, Nilai Matematika, Nilai Bahasa Inggris dan Nilai Ilmu Pengetahuan Alam (IPA). Atribut yang digunakan dalam clustering adalah atribut Nilai Bahasa Indonesia, Nilai Matematika, Nilai Bahasa Inggris, Nilai Ilmu Pengetahuan Alam (IPA), Total dan Rata-rata. Pengumpulan data dilakukan dengan cara melakukan survey ke lapangan dan tanya jawab dengan pihak sekolah untuk mendapatkan data nilai tersebut.

Adapun Flowchat analisa clustering sebagai berikut :



Gambar 1 Algoritma K-means

3. HASIL DAN PEMBAHASAN

3.1 Persiapan Data

1. Data mentah atau data asli berupa data yang disimpan dalam bentuk excel sebagai berikut :

Tabel 1. Tabel Data Mentah

No	No_Peserta	Nama Siswa	L/P	KLS	BINDO	MAT	BING	IPA	TOTAL	RATA2
1	01-002-001	Aditya Gusti Pangestu	L	IX A	7.8	9.25	7.6	8	32.65	8.16
2	01-002-002	Aditya Kusuma Aji	L	IX A	8.8	8.75	9.36	9.5	36.41	9.10
...
175	01-002-175	Shinta Rayi Kinasih	P	IXF		8	8.94	7	23.94	5.99
....
258	01-002-258	Muhammad ilzzuddin Amin	L	IXI	7.2				7.2	1.8
....

Dari Data Mentah terdapat 2 record yang tidak lengkap karena siswa tidak mengikuti Penjajakan Ujian Nasional (PUN) yaitu record ke 175 tidak mengikuti mata pelajaran Bahasa Indonesia. Record 258 tidak mengikut mata pelajaran Matematika, Bahasa Inggris dan Ilmu Pengetahuan Alam.

2. Data Cleaning bertujuan melengkapi data yang belum lengkap/noise karena bila tidak lengkap maka tidak dapat dianalisa akan ada pesan error. Metode yang digunakan adalah mengisi data yang kosong dengan mengatinya dengan nilai rata-rata. Data yang dicleaning adalah sebagai berikut :

175	01-002-175	Shinta Rayi Kinasih	P	IX F	7.98	8	8.94	7	31.92	7.98
258	01-002-258	Muhammad ilzzuddin Amin	L	IX I	7.2	7.2	7.2	7.2	28.8	7.20

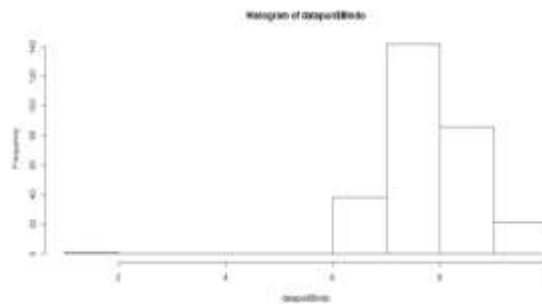
3. Menghilangkan atribut-atribut yang tidak digunakan analisa yaitu No Peserta, Nama Siswa, Jenis Kelamin, Total dan Rata2 diignore dahulu. Sehingga data yang siap diolah adalah sebagai berikut :

Tabel 2 Tabel Data Cleaning

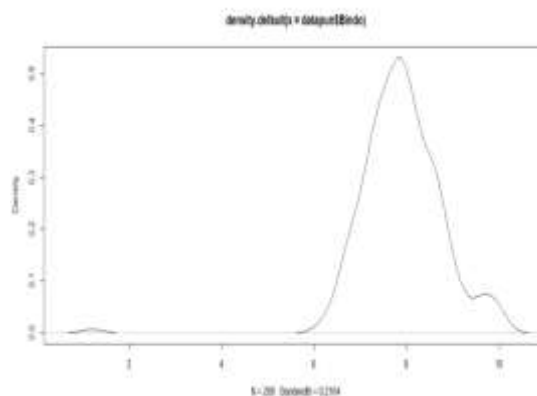
No	Bindo	Mat	Bing	IPA
1	7.80	9.25	7.60	8.00
2	8.80	8.75	9.36	9.50
3	6.80	3.50	5.80	6.75
4	8.80	7.00	8.94	9.50
...

Data Mentah disimpan di excel dengan nama pun.xls. Agar dapat diolah/dipanggil di Bahasa R maka ekstention diubah dalam bentuk txt yaitu menjadi pun.txt. Untuk pemanggilannya dengan menggunakan perintah :

- a. `> data=read.table(file.choose(), header=T)`, kemudian pilih file pun.txt.
 - b. Untuk mengecek apakah data sudah siap diolah/dianalisa menggunakan perintah : `> data`
 - c. Data yang sudah siap dianalisa kemudian dibuat aliasnya yaitu `datapun`. Hal ini bertujuan supaya data asli tidak berubah.
4. Melakukan Eksplorasi Datapun bertujuan untuk mengetahui atribut-atribut apa saja yang ada, jumlah record, nilai-nilai statistic, dimensi data, distribusi dan densitas datapun secara visual yaitu dengan menggunakan histogram dan fungsi densitas.
- a. Visualisasi Datapun Bahasa Indonesia :



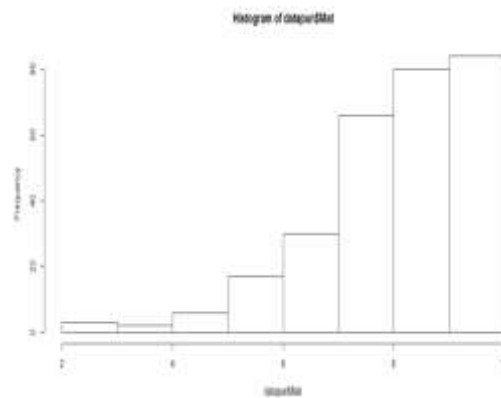
Gambar 2. Histogram Datapun MaPel Bahasa Indonesia



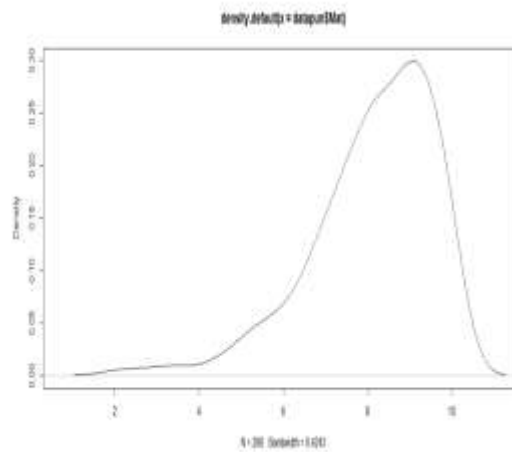
Gambar 3. Densitas Datapun Bahasa Indonesia

Dari histogram dan fungsi densitas di atas terlihat bahwa nilai PUN Bahasa Indonesia ada di sisi kanan sehingga dapat dikatakan bahwa nilai untuk mata pelajaran Bahasa Indonesia cukup baik karena diatas 6. Namun ada nilai yang sangat jelek yaitu 2 sehingga perlu tindakan supaya diatas nilai 8.

b. Visualisasi Datapun Matematika :



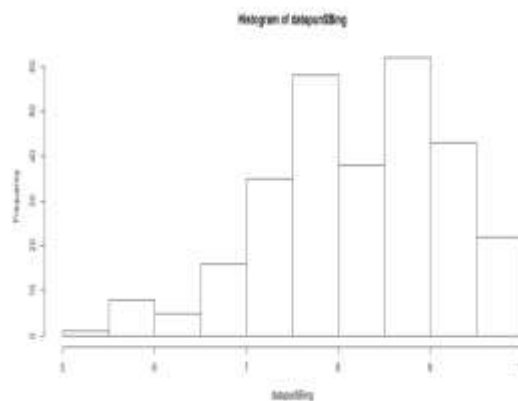
Gambar 4 Histogram Matematika



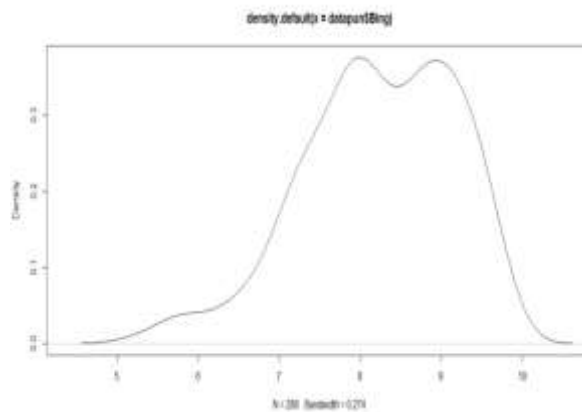
Gambar 5 Densitas Datapun Matematika

Dari histogram di atas terlihat bahwa nilai PUN Matematika tersebar dari nilai 2 sampai dengan 10 sehingga dapat dikatakan bahwa nilai untuk mata pelajaran Matematika belum cukup baik karena ada nilai 2 sampai dengan 6, sehingga perlu suatu tindakan tertentu agar nilai mata pelajaran matematika di atas 8.

c. Visualisasi Datapun Bahasa Inggris



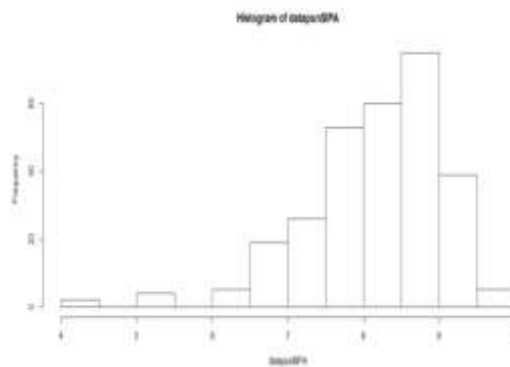
Gambar 6 Histogram Bahasa Inggris



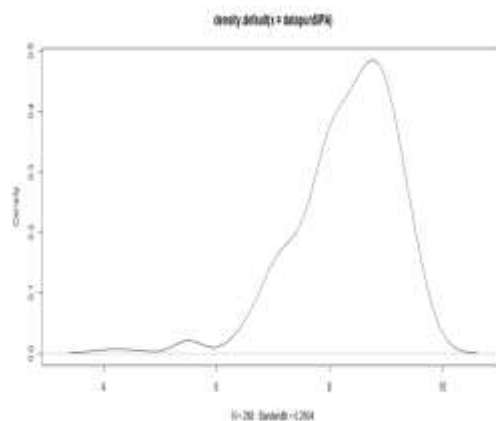
Gambar 7 Densitas Datapun Bahasa Inggris

Dari histogram di atas terlihat bahwa nilai PUN Bahasa Inggris tersebar dari nilai 5 sampai dengan 10 sehingga dapat dikatakan bahwa nilai untuk mata pelajaran Bahasa Inggris belum cukup baik karena ada nilai 5 sampai dengan 6, sehingga perlu suatu tindakan tertentu agar nilai mata pelajaran Bahasa Inggris di atas 8.

- d. Visualisasi Mata Pelajaran Ilmu Pengetahuan Alam



Gambar 8 Histogram IPA



Gambar 9. Densitas Datapun IPA

Dari histogram di atas terlihat bahwa nilai PUN IPA tersebar dari nilai 4 sampai dengan 10 sehingga dapat dikatakan bahwa nilai untuk mata pelajaran Bahasa Inggris belum cukup baik karena ada nilai 4 sampai dengan 6, sehingga perlu suatu tindakan tertentu agar nilai mata pelajaran IPA di atas 8.

- e. Clustering dengan jumlah cluster = 9, codenya adalah sebagai berikut :

> (kmeans.result <- kmeans(datapun, 9))

Hasilnya adalah sebagai berikut :

1	7	9	5	6	7	1	9	6	2	8	2	4	2	3	5	6	2	6	7	8	3	6	3	6	8	6	1	6	1	6	1	6	8	5	3	2
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

K-means clustering with 9 clusters of sizes 48, 42, 36, 18, 29, 60, 26, 19, 10. Hal ini menunjukkan bahwa Cluster 1 anggotanya 48, Cluster 2 anggotanya 42 dan seterusnya. Jumlah anggota cluster berbeda untuk tiap clusternya tergantung analisa jarak Euclidiannya. Hal ini tentunya berbeda dengan jumlah kelas tambahan untuk kelas IX yang tiap-tiap kelas jumlah siswanya sama yaitu lebih kurang 30. Dari hasil ini dapat digunakan untuk masukan bagi bidang akademik bahwa kelas tambahan jumlah siswanya tidak harus sama tergantung dari hasil analisa.

Cluster means:

Bindo Mat Bing IPA

1	8.041667	8.541667	7.590208	8.718750
2	7.628095	8.363095	9.158810	8.023810
3	7.805556	7.193056	7.228611	7.776389
4	6.622222	9.055556	7.576111	7.569444
5	8.662069	7.017241	8.398966	8.422414
6	7.703333	9.583333	8.892000	8.920333
7	9.207692	8.884615	8.656923	9.038462
8	7.494737	5.736842	8.312105	7.513158
9	7.500000	3.875000	6.292000	6.025000

Artinya bahwa tiap-tiap cluster mempunyai nilai rata-rata sebagai berfungsi sebagai pusat cluster yaitu cluster 1 rata-ratanya untuk Bahasa Indonesia 8.04, Matematika 8.54, Bahasa Inggris 7.59 dan IPA 8.71, cluster 2 rata-ratanya untuk Bahasa Indonesia 7.62, Matematika 8.36, Bahasa Inggris 9.15 dan IPA 8.02, dan seterusnya.

1) Clustering vector:

[1]	1	7	9	5	6	7	1	9	6	2	8	2	4	2	3	5	6	2	6	7	8	3	6	3	6	8	6	1	6	1	6	1	6	8	5	3	2
[38]	1	6	1	4	3	6	3	2	1	2	1	8	4	6	5	6	7	3	9	6	1	2	9	2	2	4	1	6	3	4	9	6	3	4	9	6	1
[75]	2	7	1	1	3	1	6	3	6	1	6	7	4	7	7	1	8	1	2	3	2	5	2	1	6	1	4	3	2	3	1	3	2	6	2	1	2
[112]	3	4	1	6	5	5	3	2	3	1	7	6	2	4	8	6	8	3	7	6	5	6	2	6	7	6	7	3	7	6	4	8	3	7	6	5	4
[149]	5	1	5	1	5	2	8	3	2	3	5	6	1	6	4	1	6	7	3	5	9	3	2	8	2	5	2	7	6	9	9	5	6	1	2	5	2
[186]	3	6	7	6	5	1	7	4	1	2	5	2	7	6	5	2	1	7	1	5	3	2	6	2	1	7	4	2	1	5	6	2	4	5	4	2	5
[223]	6	1	3	6	1	3	6	1	6	3	6	3	6	1	2	1	7	1	6	1	2	3	8	1	6	3	4	8	8	3	8	1	8	9	6	3	6
[260]	8	6	6	8	3	6	1	5	1	6	5	6	1	6	6	2	5	2	7	7	1	5	7	6	2	2	7	8	5								

Clustering vector yang dihasilkan menunjukkan data siswa dari 1 – 288 masuk ke cluster yang mana sesuai dengan angka cluster yang dihasilkan. Sebagai contoh baris 1 : siswa nomor urut 1 masuk di cluster 1, siswa nomor urut 2 masuk di cluster 7, siswa nomor urut 3 masuk di cluster 9 dan seterusnya

Karena hasil dalam bentuk text maka untuk memudahkan pembacaan seorang siswa masuk di cluster berapa, maka clustering vector ini selanjutnya dikembalikan ke file excell (Data Mentah). Karena bentuknya text maka harus dipecah menjadi kolom-kolom dengan menggunakan menu text to columns, kemudian matriks yang dihasilkan ditranspose ke data mentah, sehingga akan didapat semua siswa dengan clusternya. Berikut tampilan matriks urutan cluster setelah dipecah di file excell :

Tabel 5.3 Clustering vector cluster = 9 setelah dipindah ke file excel

Tabel 3 Hasil import ke file excell

1	6	1	4	3	6	3	2	1	2	1	8	4	6	5	6	7	3	9	6	1	2	9	2	2	4	1	6	3	4	9	6	3	4	9	6	1
2	7	1	1	3	1	6	3	6	1	6	7	4	7	7	1	8	1	2	3	2	5	2	1	6	1	4	3	2	3	1	3	2	6	2	1	2
3	4	1	6	5	5	3	2	3	1	7	6	2	4	8	6	8	3	7	6	5	6	2	6	7	6	7	3	7	6	4	8	3	7	6	5	4
5	1	5	1	5	2	8	3	2	3	5	6	1	6	4	1	6	7	3	5	9	3	2	8	2	5	2	7	6	9	9	5	6	1	2	5	2
3	6	7	6	5	1	7	4	1	2	5	2	7	6	5	2	1	7	1	5	3	2	6	2	1	7	4	2	1	5	6	2	4	5	4	2	5
6	1	3	6	1	3	6	1	6	3	6	3	6	1	2	1	7	1	6	1	2	3	8	1	6	3	4	8	8	3	8	1	8	9	6	3	6
8	6	6	8	3	6	1	5	1	6	5	6	1	6	6	2	5	2	7	7	1	5	7	6	2	2	7	8	5								

Matriks clustering vector kemudian ditranspose di kolom cluster dari data mentah. Hasilnya :

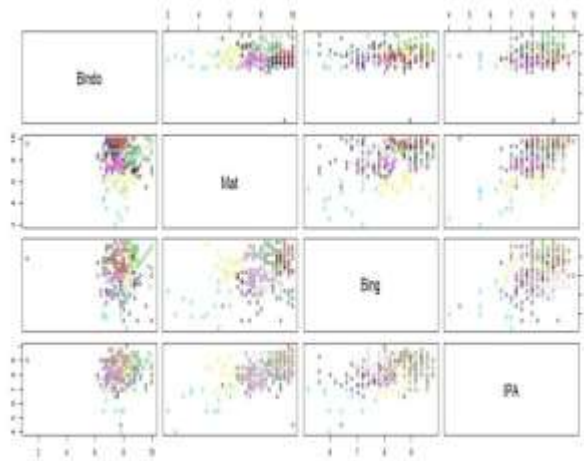
Tabel 4 Tabel Hasil Cluster = 9

No	Bindo	Mat	Bing	IPA	cluster
1	7.80	9.25	7.60	8.00	1
2	8.80	8.75	9.36	9.50	7
.....
287	7.80	6.50	8.80	7.25	8
288	8.40	7.50	8.20	9.50	5

- 2) Within cluster sum of squares by cluster:
 [1] 45.38456 51.06018 57.93551 68.29688 36.33926 47.08902 28.17380 29.54641
 [9] 28.28806
 (between_SS / total_SS = 70.3 %)
- 3) Available components:
 [1] "cluster" "centers" "totss" "withinss" "tot.withinss"
 [6] "betweenss" "size" "iter" "ifault"



Gambar 10 Code Clustering dengan k-means dengan jumlah cluster = 9



Gambar 11 Plot cluster dengan jumlah cluster = 9 untuk 4 mata pelajaran

- f. Analisa untuk cluster yang lain dapat dilakukan sesuai dengan kebutuhan bidang akademik.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Setelah melakukan analisa clustering terhadap nilai Penjajakan Ujian Nasional (PUN) yaitu nilai Bahasa Indonesia, Matematika, Bahasa Inggris dan Ilmu Pengetahuan terhadap 288 siswa pada SMP “X” di Kota Semarang maka dapat disimpulkan sebagai berikut :

1. Jumlah Cluster yang terbentuk dapat ditentukan oleh pihak sekolah dalam hal ini bidang akademik sesuai dengan kebutuhan misal jumlah guru, ketersediaan ruang atau hal yang lain.
2. Cluster yang terbentuk berisi informasi anggota cluster berupa cluster mean, clustering vector, sum of square dan visualisasi cluster dalam bentuk diagram plot.

Cluster yang dihasilkan dapat menjadi alternative dalam pengadaan kelas tambahan yang diberikan pada SMP “X” di Kota Semarang mengenai jumlah cluster dan anggota clusternya.

4.2 Saran

1. Perlu pengembangan analisa kembali dengan nilai-nilai Penjajakan Nilai Ujian (PUN) yang lain agar didapat cluster yang betul-betul baik sesuai dengan jumlah kelas dan nilai.
2. Perlu dilakukan analisa perbandingan dengan metode clustering yang lain agar didapat cluster yang baik

DAFTAR PUSTAKA

- [1] Charu C Aggarwal, 2015, Data Mining The Texbook, New York, Springer.
- [2] Jiawei Han dan Micheline Kamber dan Jian Pei, 2012, Data Mining Concepts and Techniques, New York, Morgan Kaufmann Publisher.
- [3] Ian H Witten dan Eibe Frank dan Mark A Hall, 2011, Data Mining Practical Machine Learning Tools and Techniques, Burlington MA 01803 USA , Morgan Kaufmann Publishers is an imprint of Elsevier 30 Corporate Drive, Suite 400.
- [4] Nur Rosyid Muhtada’I, Mike Yuliana, Beni Ilham Priyambodo, 2011, Analisa Perbandingan Clustering Metode Manual Dan Metode Single Linkage Untuk Menentukan Kinerja Agent Pada Call Centre Berbasis Asterisk For Java, Electronic Engineering Polytechnic Institute of Surabaya (EEPIS), Surabaya Indonesia.
- [5] Turban , Efraim & Aronson, Jay E. 2005. *Decision Support Systems and Intelligent Systems. 6th edition*. Prentice Hall: Upper Saddle River, NJ.
- [6] <https://www.scribd.com/doc/51963978/ANALISA-METODE-K-MEANS-PARTITIONAL-CLUSTERING-07560292>
- [7] https://www.academia.edu/9762272/LANDASAN_TEORI_1._Pengertian_Data_Mining_Data_mining