

Pemanfaatan Metode K-Means *Clustering* dalam Penentuan Penjurusan Siswa SMA

Chandra Purnamaningsih

Informatika, Fakultas MIPA,
Universitas Sebelas Maret Surakarta
Jl. Ir. Sutami No 36 A Surakarta

cpurnamaningsih@gmail.com

Ristu Saptono

Informatika, Fakultas MIPA,
Universitas Sebelas Maret Surakarta
Jl. Ir. Sutami No 36 A Surakarta

r_saptono@uns.ac.id

Abdul Aziz

Informatika, Fakultas MIPA,
Universitas Sebelas Maret Surakarta
Jl. Ir. Sutami No 36 A Surakarta

abdul_7733@yahoo.com

ABSTRAK

Penentuan penjurusan siswa SMA dilakukan berdasarkan kriteria nilai akademik yang menjadi ciri dari masing-masing jurusan IPA/IPS, hal ini memungkinkan bagi seorang siswa untuk memenuhi kriteria diterima di kedua jurusan atau ditolak di keduanya. Usaha untuk mengurangi kemungkinan tersebut adalah dengan cara memperhatikan pertimbangan kriteria lainnya seperti ditinjau dari nilai IQ dan minat siswa.

Proses penentuan penjurusan siswa SMA dengan teknik clustering menggunakan metode K-Means clustering. Pada penelitian ini dilakukan clustering sendiri-sendiri untuk IPA/IPS dan dipaparkan perbandingan hasil clustering K-Means kriteria nilai akademik, nilai IQ, minat siswa dengan clustering K-Means nilai akademik. Data siswa dikelompokkan sendiri-sendiri sesuai jurusan masing-masing. Jurusan IPA dikelompokkan menjadi dua yaitu diterima IPA dan ditolak IPA. Untuk jurusan IPS dikelompokkan menjadi dua yaitu diterima IPS dan ditolak IPS. Kemudian setiap cluster diklasifikasikan berdasarkan kriteria mana yang lebih diprioritaskan. Cluster dengan nilai terbesar pada centroid akhir merupakan cluster yang diterima IPA/IPS, sedangkan cluster dengan nilai terkecil pada centroid akhir merupakan cluster yang ditolak IPA/IPS.

Hasil penelitian pengujian terbaik pada *praprocessing* clustering K-Means IPA dengan hasil akurasi 0.905882, tingkat kesesuaian hasil prediksi dengan data sebenarnya (*recall*) 1, ketepatan hasil pengujian dalam memprediksi clustering (*sensitivity*) 0.876923, kesesuaian prediksi negatif terhadap aktual negatif (*specificity*) 0.714285. Sedangkan pengujian terbaik juga pada *praprocessing* clustering K-Means IPS didapatkan akurasi 0.905882, *recall* 0.714285, *sensitivity* 1, dan *specificity* 1. Hasil perbandingan clustering terbaik pada *praprocessing* clustering K-Means IPA dengan *praprocessing* clustering K-Means IPS menunjukkan bahwa tidak ada siswa yang diterima di dua jurusan IPA/IPS atau siswa ditolak di keduanya.

Kata kunci

Centroid, Clustering, IPA, IPS, K-Means, Penjurusan

1. PENDAHULUAN

Penentuan penjurusan siswa SMA dilakukan berdasarkan kriteria nilai akademik yang menjadi ciri dari masing-masing jurusan IPA/IPS, dengan demikian dapat diketemukan banyak kemungkinan seorang siswa memenuhi semua atau sebagian kriteria penjurusan dari sekolah. Artinya, dimungkinkan bagi seorang siswa, diterima dua jurusan IPA/IPS atau ditolak keduanya. Kemungkinan yang akan terjadi jika siswa mengalami kesalahan dalam penempatan jurusan yang tidak sesuai yang seharusnya siswa diterima IPA tetapi kenyatannya tidak, hal ini akan menyebabkan penurunan motivasi, begitu juga sebaliknya akan menyebabkan keterbelakangan prestasi. Proses penjurusan di

SMA ini cukup rumit, dengan jumlah siswa yang banyak dan kriteria nilai yang digunakan yang menjadi ciri tiap jurusan berbeda-beda, tentu penentuan penjurusan Sekolah Menengah Atas (SMA) akan memakan waktu yang lama..

Salah satu cara untuk mempermudah penentuan penjurusan SMA adalah dengan cara mengelompokkan (*clustering*) data siswa yaitu, untuk clustering IPA dikelompokkan menjadi 2 kelompok antara lain diterima IPA, ditolak IPA. Untuk clustering IPS dikelompokkan menjadi 2 kelompok antara lain diterima IPS dan ditolak IPS. Untuk mengelompokkan siswa ini, metode yang digunakan adalah *K-Means Clustering*. Algoritma *K-Means* dipilih karena memiliki ketelitian yang cukup tinggi terhadap ukuran objek, sehingga algoritma ini relatif lebih terukur dan efisien untuk pengolahan objek dalam jumlah besar. Selain itu algoritma *K-Means* ini tidak terpengaruh terhadap urutan objek[1]. Metode ini berusaha untuk meminimalkan variasi antar data yang ada di dalam suatu cluster dan memaksimalkan variasi dengan data yang ada di cluster lainnya[2].

Untuk kriteria yang digunakan dalam penelitian ini adalah kriteria nilai akademik, nilai IQ dan minat siswa. Pada proses clustering metode K-Means dapat dilakukan pada atribut-atribut numerik kontinu, sedangkan minat merupakan data non numerik yang harus di transformasikan ke bentuk numerik yang menjadi data diskrit. Untuk data nilai IQ juga sama dilakukan di transformasikan terlebih dahulu ke bentuk numerik yang berupa data diskrit. Oleh karena kriteria data minat dan nilai IQ, data yang digunakan masih berupa numerik diskrit maka dilakukan tahap *praprocessing* terlebih dahulu.

Nanjaya [3] pada penelitian sebelumnya, melakukan pembahasan mengenai penggunaan metode *K-Means* pada suatu clustering data non-numerik (*categorical*) untuk studi kasus biro jodoh. Dari penelitian tersebut didapatkan bahwa clustering dapat dilakukan pada atribut-atribut kategorikal yang ditransformasikan terlebih dahulu ke dalam bentuk numerik. Oleh karena itu metode K-Means Clustering dipilih untuk diterapkan dalam kasus penentuan penjurusan siswa SMA.

2. DASAR TEORI

2.1 Clustering

Proses pengelompokan sekumpulan objek fisik ataupun abstrak kedalam kelas-kelas yang mempunyai kemiripan[4].

2.2 K-Means

K-Means merupakan salah satu metode data *clustering* non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster/kelompok. Metode ini mempartisi data ke dalam cluster/kelompok sehingga data yang memiliki karakteristik sama dikelompokkan ke dalam satu cluster yang sama[3].

Menurut MacQueen JB[5], Berikut adalah langkah-langkah dari algoritma *K-Means* :

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

$$d(x_i, \mu_i) = \sqrt{(x_i - \mu_i)^2} \tag{1}$$

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

$$C_k = \frac{1}{n_k} \sum d_i \tag{2}$$

dimana:

n_k = jumlah data dalam cluster

d_i = jumlah dari nilai jarak yang masuk dalam masing-masing cluster

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

2.3 Confusion Matrix

Confusion matrix mengandung informasi tentang kelas data aktual dan kelas data hasil prediksi yang direpresentasikan pada baris matriks[6].

Tabel 2. Confusion Matrik[8]

Data		True Class	
		P	N
Hypothesized Class	Y	True Positives (TP)	False Positives (FP)
	N	False Negatives (FN)	True Negatives (TN)
Total		P	N

Beberapa parameter pengukur kinerja ditunjukkan dengan Persamaan (3) sampai dengan Persamaan (6)[7].

Metrik yang akan dihitung berdasarkan data *confusion matrik* :

- Akurasi (AC)

$$AC = \frac{TP+TN}{TP+FP+FN+TN} \tag{3}$$

- *Recall* merupakan tingkat kesesuaian hasil prediksi dengan data sebenarnya.

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

- *Sensitivity* merupakan ukuran ketepatan hasil pengujian dalam memprediksi clustering.

$$Sensitivity = \frac{TP}{TP + FP} \tag{5}$$

- *Specificity* merupakan suatu ukuran kesesuaian prediksi negatif terhadap aktual negatif.

$$Specificity = \frac{TN}{TN + FP} \tag{6}$$

3. METODOLOGI PENELITIAN

3.1 Pengumpulan Data

Tahap Pengumpulan data dilakukan dengan wawancara, studi literature, dan telaah dokumen. Dari metode pengumpulan data ini diperoleh 255 data siswa satu angkatan.

3.2 Tahap Pemodelan Data

Memodelkan data siswa yang telah didapatkan menjadi model data yang terstruktur. Kriteria yang digunakan dalam penentuan penjurusan ini adalah nilai akademik, minat, dan nilai IQ siswa. Namun data minat siswa dan nilai IQ belum dapat digunakan dalam proses clustering maka perlu adanya proses konversi data non numerik minat siswa dan nilai IQ siswa menjadi data numerik, oleh karena itu kriteria minat dan IQ siswa dikategorikan terlebih dahulu.

1. Kategori minat siswa

Tabel3.1 Pengkategorian Minat

Kategori	Kodifikasi
IPA	1
IPS	0

2. IQ siswa.

Tabel3.2 Pengkategorian IQ

Kategori	Kodifikasi
IPA > 110	1
IPS ≤ 110	0

3. Pembobotan IPA, Pembobotan IPS

Perolehan hasil tes IQ yang didapatkan dalam penjurusan ini berdasarkan tes kemampuan kuantitatif, dimana kemampuan kuantitatif diukur berdasarkan kemampuan numerik, aritmatika, dan analisis.

Tabel3.4 Pembobotan IPA

Minat	Nilai IQ	Pembobotan IPA	Pembobotan IPS
1	1	1	1/4
1	0	3/4	1/2
0	1	1/2	3/4
0	0	1/4	1

4. Perkalian masing-masing kriteria

Pada masing-masing kriteria jurusan IPA (matematika, fisika, biologi, kimia) dikalikan dengan hasil pembobotan IPA begitu juga dengan masing-masing kriteria jurusan IPS (sejarah, geografi, ekonomi, sosiologi) dikalikan dengan hasil pembobotan IPS.

3.3 Proses Clustering

Pada tahap ini proses *clustering* menggunakan metode *K-Means clustering* dimana proses clustering dilakukan sendiri-sendiri pada masing-masing jurusan IPA/IPS. Data yang digunakan dalam proses clustering adalah nilai kriteria akademik yang menjadi ciri dari masing-masing jurusan IPA(matematika, fisika, biologi, dan kimia) dan IPS (sejarah, geografi, ekonomi dan sosiologi) yang akan diterapkan dalam metode K-Means clustering yaitu dengan menentukan jumlah cluster sebanyak 2 (dua) diterima IPA dan ditolak IPA untuk clustering IPA sedangkan proses clustering untuk IPS jumlah cluster sebanyak 2 (dua) yakni diterima IPS dan ditolak IPS.

3.4 Klasifikasi Hasil Clustering

- *Cluster* yang direkomendasikan diterima IPA/IPS adalah cluster yang memiliki nilai pusat cluster terbesar dan jarak cluster pada pusat cluster yang paling kecil.

- Cluster yang direkomendasikan ditolak IPA/IPS adalah cluster yang memiliki nilai pusat cluster terkecil dan jarak cluster pada pusat cluster besar.

3.5 Pengujian dan Validasi Hasil

3.5.1 Tahap Pengujian berdasarkan Confusion Matrik.

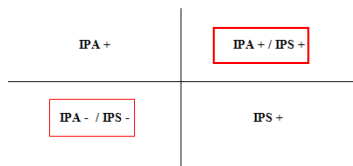
Pengujian dilakukan pada masing-masing clustering K-Means IPA, K-Means IPS, praprocessing clustering K-Means IPA dan praprocessing clustering K-Means IPS dengan menghitung nilai akurasi, recall, sensitivity, specificity dari confusion matrik setiap percobaan.

Tabel 3.5 Confusion Matrik Data Dua Kelas

Data		Aktual	
		Kelas Positif	Kelas Negative
Hasil Prediksi	Kelas Postive	True Positives (TP)	False Positives (FP)
	Kelas Negative	False Negatives(FN)	True Negatives (TN)
Total		P	N

3.5.2 Tahap Pengujian Perbandingan berdasarkan Hasil Clustering.

- Perbandingan hasil pengujian pada clustering K-Means IPA dengan K-Means IPS dengan data kriteria nilai akademik.. Ilustrasi hasil pengelompokan penjurusan ditunjukkan pada Gambar 1.



Gambar 1. Ilustrasi Hasil Pengelompokan Penjurusan

Keterangan :

IPA + : Diterima penjurusan IPA

IPA - : Ditolak penjurusan IPA

IPS + : Diterima penjurusan IPS

IPS - : Ditolak penjurusan IPA

- Perbandingan hasil pada clustering antara praprocessing clustering K-Means IPA dengan praprocessing clustering K-Means IPS apakah dapat mengurangi kemungkinan seorang siswa diterima dua jurusan IPA/IPS atau ditolak keduanya.

4. HASIL DAN PEMBAHASAN

4.1 Pemodelan Data

Data siswa kriteria nilai IQ dan minat siswa dilakukan pemodelan data terlebih dahulu dengan cara mengkategorikan data IQ dan minat siswa dari ketegorial (non numerik) menjadi numerik, selanjutnya dilakukan pembobotan berdasarkan pembobotan pada masing-masing jurusan IPA/IPS. Pada masing-masing kriteria jurusan IPA/IPS dikalikan dengan hasil pembobotan IPA/IPS.

Tabel 4.1 Pemodelan Data jurusan IPA

NIS	Minat	IQ	Bobot IPA	Mat	Fis	Bio	Kim
8811	1	0	3/4	58.5	58.5	63	64.5
8812	1	0	3/4	63.75	61.5	56.25	66
8813	0	1	1/2	35	38	38.5	38.5
8814	1	0	3/4	65.25	66	63	66

Tabel 4.1 Pemodelan Data jurusan IPA

NIS	Minat	IQ	Bobot IPA	Mat	Fis	Bio	Kim
8815	1	1	1	71	77	78	76
8816	1	0	3/4	57	65.25	57	61.5
8817	1	1	1	78	84	80	87
8818	1	0	3/4	63.75	60.75	63	60.75
8819	0	0	1/4	17.5	20.25	19.75	18.75
8820	1	1	1	90	82	83	87

Tabel 4.2 Pemodelan Data Jurusan IPS

NIS	Minat	IQ	Bobot IPS	Sej	Geo	Eko	Sos
8811	1	0	1/2	42.5	36	40	42.5
8812	1	0	1/2	39.5	41	37.5	40
8813	0	1	3/4	60	61.5	56.25	59.25
8814	1	0	1/2	40	42.5	40	42
8815	1	1	1/4	21	19.5	19.25	19.75
8816	1	0	1/2	42.5	40	37	41.5
8817	1	1	1/4	21.25	22.25	22.25	21.25
8818	1	0	1/2	41	36	40.5	40
8819	0	0	1	82	80	81	83
8820	1	1	1/4	20.5	20.75	19.75	21.75

4.2 Proses Clustering

Proses clustering dengan menggunakan metode K-Means akan dilakukan terhadap 255 data sampel siswa.

1. Mula-mula sistem akan mengambil pusat cluster (centroid) awal secara random. Pada pembahasan ini diambil contoh satu percobaan. Dari hasil pengambilan secara random, diperoleh centroid untuk masing-masing cluster.

Clustering K-Means IPA

Pusat cluster IPA diperoleh :

Centroid kriteria 1 adalah matematika, centroid kriteria 2 adalah fisika, centroid kriteria 3 adalah biologi, centroid kriteria 4 adalah kimia.

$$C1 = (74, 77, 78, 82)$$

$$C2 = (75, 87, 75, 78)$$

Clustering K-Means IPS

Pusat cluster IPS diperoleh :

Centroid kriteria 1 adalah sejarah, centroid kriteria 2 adalah geografi, centroid kriteria 3 adalah ekonomi, centroid kriteria 4 adalah sosiologi.

$$C1 = (80, 81, 76, 82)$$

$$C2 = (85, 84, 74, 86)$$

Praprocessing Clustering K-Means IPA

Pusat cluster IPA diperoleh

$$C1 = (55,5, 57,75, 58,5, 58,5)$$

$$C2 = (37, 38,5, 37,5, 37,5)$$

Praprocessing Clustering K-Means IPS

Pusat cluster IPS diperoleh :

$$C1 = (45, 45, 39,5, 45)$$

$$C2 = (40, 37,5, 39,5, 44)$$

2. Kemudian akan dihitung jarak dari setiap data yang ada terhadap setiap pusat cluster awal.
3. Dari hasil perhitungan jarak, setiap data akan menjadi anggota suatu cluster yang memiliki jarak terdekat dari pusat cluster
4. Perhitungan pusat *cluster* baru ini dilakukan dengan menghitung nilai rata-rata masing-masing kriteria dari seluruh anggota yang menjadi anggota masing-masing *cluster*.
5. Proses 2 sampai 4 akan terus berulang hingga posisi data sudah tidak mengalami perubahan. Hasil pengelompokkan iterasi terakhir dibandingkan dengan hasil sebelumnya.

Tabel 4.3 menunjukkan proses berhenti pada iterasi ke-9, untuk *clustering K-Means* IPA dengan hasil *centroid* akhir:

Tabel 4.3 Hasil centroid akhir K-Means IPA

Cluster 1	
<i>centroid</i> kriteria 1 cluster 1	73,62
<i>centroid</i> kriteria 2 cluster 1	78,45
<i>centroid</i> kriteria 3 cluster 1	78,34
<i>centroid</i> kriteria 4 cluster 1	77,87
Cluster 2	
<i>centroid</i> kriteria 1 cluster 2	81,81
<i>centroid</i> kriteria 2 cluster 2	80,32
<i>centroid</i> kriteria 3 cluster 2	81,73
<i>centroid</i> kriteria 4 cluster 2	83,79

Tabel 4.4 menunjukkan proses berhenti pada iterasi ke-7, untuk *clustering K-Means* IPS dengan hasil *centroid* akhir:

Tabel 4.4 Hasil centroid akhir K-Means IPS

Cluster 1	
<i>centroid</i> kriteria 1 cluster 1	81,96
<i>centroid</i> kriteria 2 cluster 1	77,49
<i>centroid</i> kriteria 3 cluster 1	77,38
<i>centroid</i> kriteria 4 cluster 1	81,72
Cluster 2	
<i>centroid</i> kriteria 1 cluster 2	84,66
<i>centroid</i> kriteria 2 cluster 2	85,91
<i>centroid</i> kriteria 3 cluster 2	77,94
<i>centroid</i> kriteria 4 cluster 2	86,12

Tabel 4.5 menunjukkan proses berhenti pada iterasi ke-2, untuk *praprocessing clustering K-Means* IPA dengan hasil *centroid* akhir:

Tabel 4.5 Hasil centroid akhir praprocessing K-Means IPA

Cluster 1	
<i>centroid</i> kriteria 1 cluster 1	64,41
<i>centroid</i> kriteria 2 cluster 1	65,6
<i>centroid</i> kriteria 3 cluster 1	66,19
<i>centroid</i> kriteria 4 cluster 1	67,05
Cluster 2	
<i>centroid</i> kriteria 1 cluster 2	22,23
<i>centroid</i> kriteria 2 cluster 2	23,8
<i>centroid</i> kriteria 3 cluster 2	23,73
<i>centroid</i> kriteria 4 cluster 2	23,36

Tabel 4.6 menunjukkan proses berhenti pada iterasi ke-3, untuk *praprocessing clustering K-Means* IPS dengan hasil *centroid* akhir:

Tabel 4.6 Hasil centroid akhir praprocessing K-Means IPS

Cluster 1	
<i>centroid</i> kriteria 1 cluster 1	77,44
<i>centroid</i> kriteria 2 cluster 1	74,13
<i>centroid</i> kriteria 3 cluster 1	72,96
<i>centroid</i> kriteria 4 cluster 1	78,03
Cluster 2	
<i>centroid</i> kriteria 1 cluster 2	35,48
<i>centroid</i> kriteria 2 cluster 2	34,98
<i>centroid</i> kriteria 3 cluster 2	32,97
<i>centroid</i> kriteria 4 cluster 2	35,69

4.3 Klasifikasi Hasil Clustering

Berikut hasil klasifikasi *cluster K-Means* IPA

Tabel 4.7 Klasifikasi cluster K-Means IPA

Ditolak IPA	Diterima IPA
Cluster 1	Cluster 2
73,62	81,81
78,45	80,32
78,34	81,73
77,87	83,79

Berdasarkan Tabel 4.7 siswa yang diterima IPA adalah siswa yang masuk dalam pengelompokkan *cluster 2*, sedangkan siswa yang direkomendasikan ditolak IPA adalah siswa yang masuk dalam pengelompokkan *cluster 1*.

Berikut hasil klasifikasi *cluster K-Means* IPS berhenti pada iterasi ke-7.

Tabel 4.8 Klasifikasi cluster K-Means IPS

Ditolak IPS	Diterima IPS
Cluster 1	Cluster 2
81,96	84,66
77,49	85,91
77,38	77,94
81,72	86,12

Berdasarkan Tabel 4.8 siswa yang diterima IPS adalah siswa yang masuk dalam pengelompokkan *cluster 2*, sedangkan siswa yang direkomendasikan ditolak IPS adalah siswa yang masuk dalam pengelompokkan *cluster 1*.

Iterasi pada percobaan *praprocessing clustering K-Means* IPA berhenti pada iterasi ke-2.

Tabel 4.9 Klasifikasi cluster praprocessing K-Means IPA

Diterima IPA	Ditolak IPA
Cluster 1	Cluster 2
64,41	22,23
65,6	23,8
66,19	23,73
67,05	23,36

Berdasarkan Tabel 4.9 siswa yang diterima IPA dengan kriteria nilai akademik, IQ, dan minat adalah *cluster 1*, sedangkan siswa yang direkomendasikan ditolak IPA adalah *cluster 2*.

Iterasi pada percobaan *praprocessing clustering K-Means* IPS berhenti pada iterasi ke-2.

Tabel 4.10 Klasifikasi praprocessing clustering K-Means IPS

Diterima IPS	Ditolak IPS
Cluster 1	Cluster 2
77,44	35,48
74,13	34,98
72,96	32,97
78,03	35,69

Berdasarkan Tabel 4.10 siswa yang diterima IPS dengan kriteria nilai akademik, IQ, dan minat adalah *cluster 1*, sedangkan siswa yang direkomendasikan ditolak IPS adalah *cluster 2*.

4.4 Hasil Pengujian Clustering

4.4.1 Hasil Pengujian berdasarkan Confusion Matrik.

Mengetahui keakuratan yang dihasilkan metode *K-Means clustering* diukur dengan menggunakan *confusion matrix*. Tabel 4.11 menunjukkan pengujian dengan *confusion matrix* dengan sample 1 percobaan dari 30 percobaan *clustering K-Means* IPA.

Tabel 4.11 Confusion Matrik K-Means IPA

Data		Aktual	
		Diterima IPA	Ditolak IPA
Hasil Prediksi	Diterima IPA	98 (TP)	8 (FP)
	Ditolak IPA	73 (FN)	76(TN)
Total		171 (P)	84 (N)

Tabel 4.12 menunjukkan pengujian dengan *confusion matrix* dengan sample 1 percobaan dari 30 percobaan *clustering K-Means* IPS.

Tabel 4.12 Confusion MatrikK-Means IPS

Data		Aktual	
		Diterima IPS	Ditolak IPS
Hasil Prediksi	Diterima IPS	19 (TP)	99 (FP)
	Ditolak IPS	65 (FN)	72 (TN)
Total		84 (P)	171 (N)

Tabel 4.13 menunjukkan pengujian dengan *confusion matrix* dengan sample 1 percobaan dari 30 percobaan *praprocessing clustering K-Means* IPA.

Tabel 4.13 Confusion Matrik praprocessing K-Means IPA

Data		Aktual	
		Diterima IPA	Ditolak IPA
Hasil Prediksi	Diterima IPA	171 (TP)	24 (FP)
	Ditolak IPA	0 (FN)	60 (TN)
Total		171 (P)	84 (N)

Dari tabel 4.14 menunjukkan pengujian dengan *confusion matrix* dengan sample 1 percobaan dari 30 percobaan *praprocessing clustering K-Means* IPS.

Tabel 4.14 Confusion Matrik praprocessing K-Means IPS

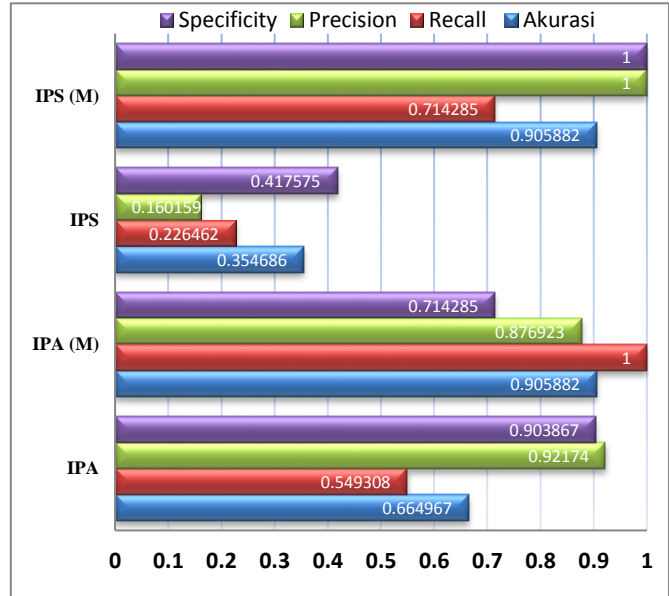
Data		Aktual	
		Diterima IPS	Ditolak IPS
Hasil Prediksi	Diterima IPS	60 (TP)	0 (FP)
	Ditolak IPS	24 (FN)	171 (TN)
Total		84 (P)	171 (N)

Setelah didapatkan hasil pengujian data aktual dengan hasil prediksi menggunakan *confusion matrik* maka dapat dilakukan pengujian terhadap akurasi, *recall*, *sensitivity*, *specificity* pada masing-masing *clustering K-Means* IPA, *K-Means* IPS, *praprocessing clustering K-Means* IPA dan *praprocessing clustering K-Means* IPS. Berikut perolehan rata-rata hasil pengujian yang ditunjukkan pada Tabel 4.15

Tabel 4.15 Rata-rata Hasil Pengujian

K-Means	Akurasi	Recall	Precision	Specificity
IPA	0.664967	0.549308	0.921740	0.903867
Pra IPA	0.905882	1	0.876923	0.714285
IPS	0.354686	0.226462	0.160159	0.417575
Pra IPS	0.905882	0.714285	1	1

Berdasarkan gambar 2. diketahui hasil pengujian clustering untuk *praprocessing clustering K-Means* IPA dan *preprocessing clustering K-Means* IPS memberikan hasil grafik yang signifikan dibandingkan dengan clustering *K-Means* IPA dan *K-Means* IPS sehingga berdasarkan grafik diatas *preprocessing clustering K-Means*IPA diperoleh akurasi 0.905882, *recall* 1, *sensitivity* 0.876923, *specivicity* 0.714285. Sedangkan untuk *praprocessing clustering K-Means* IPS akurasi 0.905882, *recall* 0.714285, *sensitivity* 1, dan *specificity* 1.



Gambar 2. Grafik Perbandingan Hasil Pengujian Clustering

4.4.2 Hasil Pengujian Perbandingan berdasarkan Clustering.

Tabel 4.16 menunjukkan perbandingan hasil *clustering K-Means* IPA dengan *K-Means* IPS.

Tabel 4.16 Pengujian K-Means IPA dengan K-Means IPS

Uji	IPA+/IPS-	IPA+/IPS+	IPA-/IPS-	IPA-/IPS+
1	36	70	101	48
2	32	67	104	52
3	32	67	104	52
4	36	70	101	48
5	36	70	100	49
6	32	67	105	51
7	32	67	105	51
8	35	69	101	50
9	36	70	100	49
10	31	68	105	51
11	32	67	104	52
12	36	70	100	49
13	32	67	104	52
14	32	67	104	52
15	35	71	101	48
16	34	67	102	52
17	36	70	101	48
18	34	67	102	52
19	36	70	100	49
20	35	71	101	48

Tabel 4.16 Pengujian K-Means IPA dengan K-Means IPS

Uji	IPA+/IPS-	IPA+/IPS+	IPA-/IPS-	IPA-/IPS+
21	32	67	105	51
22	36	70	100	49
23	32	67	105	51
24	35	71	101	48
25	32	67	105	51
26	36	70	101	48
27	32	67	105	51
28	36	70	101	48
29	31	68	105	51
30	32	67	104	52
Rata-rata	33	68	102	51

Berdasarkan Tabel 4.16 diketahui bahwa masih banyak ditemukan kemungkinan siswa diterima kedua jurusan IPA/IPS maupun ditolak keduanya yang dibuktikan rata-rata siswa diterima di kedua jurusan IPA/IPS sebesar 68 siswa dan ditolak di kedua jurusan rata rata 102 siswa dari 255 siswa.

Tabel 4.17 menunjukkan perbandingan hasil clustering *praprocessing clustering K-Means* IPA dengan *preprocessing clustering K-Means* IPS.

Tabel 4.17 Praprocessing K-Means IPA vs Praprocessing IPS

Uji	IPA+/IPS-	IPA+/IPS+	IPA-/IPS-	IPA-/IPS+
1	195	0	0	60
2	195	0	0	60
3	195	0	0	60
4	195	0	0	60
5	195	0	0	60
6	195	0	0	60
7	195	0	0	60
8	195	0	0	60
9	195	0	0	60
10	195	0	0	60
11	195	0	0	60
12	195	0	0	60
13	195	0	0	60
14	195	0	0	60
15	195	0	0	60
16	195	0	0	60
17	195	0	0	60
18	195	0	0	60
19	195	0	0	60
20	195	0	0	60
21	195	0	0	60
22	195	0	0	60

Tabel 4.18 Praprocessing K-Means IPA vs Praprocessing IPS

Uji	IPA+/IPS-	IPA+/IPS+	IPA-/IPS-	IPA-/IPS+
23	195	0	0	60
24	195	0	0	60
25	195	0	0	60
26	195	0	0	60
27	195	0	0	60
28	195	0	0	60
29	195	0	0	60
30	195	0	0	60
Rata-rata	195	0	0	60

Berdasarkan Tabel 4.18 dari hasil rata-rata pengujian, siswa diterima di dua jurusan IPA/IPS atau ditolak kedua jurusan IPA/IPS tidak ada. Jadi *clustering* dengan *praprocessing clustering K-Means* IPA dan *praprocessing clustering K-Means* IPS tidak terjadi kemungkinan siswa diterima di dua jurusan IPA/IPS atau ditolak di dua jurusan.

5. PENUTUP

Dari hasil penelitian, dapat disimpulkan bahwa algoritma *K-Means clustering* dapat digunakan untuk mengelompokkan data siswa sebagai pendukung keputusan penentuan penjurusan siswa SMA. Berdasarkan hasil pengujian terbaik pada *praprocessing clustering K-Means* IPA dengan hasil akurasi 0.905882, tingkat kesesuaian hasil prediksi dengan data sebenarnya (*recall*) 1, ketepatan hasil pengujian dalam memprediksi clustering (*sensitivity*) 0.876923, kesesuaian prediksi negatif terhadap aktual negatif (*specificity*) 0.714285. Sedangkan pengujian terbaik juga pada *praprocessing clustering K-Means* IPS didapatkan akurasi 0.905882, recall 0.714285, sensitivity 1, dan specificity 1. Hasil perbandingan clustering terbaik pada *praprocessing clustering K-Means* IPA dengan *praprocessing clustering K-Means* IPS menunjukkan bahwa tidak ada siswa yang diterima di dua jurusan IPA/IPS atau siswa ditolak di keduanya. Hal ini dikarenakan proses *clustering* mempertimbangkan kriteria nilai akademik, IQ dan minat siswa yang dapat mempengaruhi proses hasil *clustering K-Means* yang dapat mencegah kemungkinan siswa diterima dua jurusan IPA/IPS atau ditolak keduanya.

Saran untuk penelitian selanjutnya dalam penentuan penjurusan siswa SMA sebaiknya mengkatagorikan nilai IQ. Pengkatagorikan nilai IQ berdasarkan komponen tes IQ yang terdiri dari tes kuantitatif yang merupakan saran untuk ke penjurusan IPA, verbal saran untuk ke penjurusan Bahasa, dan kualitatif saran untuk ke penjurusan IPS. Hasil tes IQ untuk setiap komponen tersebut disesuaikan dengan rekomendasi untuk kebutuhan masing-masing jurusan.

6. DAFTAR PUSTAKA

[1] Simamora B. *Analisis Multivariat Pemasaran*. Jakarta: PT. Gramedia Pustaka Utama; 2005.
 [2] Agusta, Y. (2007). K-Means-Penerapan, Permasalahan dan Metode Terkait. *Jurnal Sistem dan Informatika* Vol.3 ,47-60.
 [3] Nanjaya, D. (2005). *Clustering* Data Non-Numerik dengan Pendekatan Algoritma K-Means dan Hamming Distance Studi Kasus Biro Jodoh. *Jurnal Ilmiah Teknologi Informasi*, 46-53.

- [4] J. Han, M. Kamber. (2001: pp34-39). *Data Mining: Concepts and Techniques*, The Morgan Kaufmann Series.
- [5] J. MacQueen. 1967. Some methods for classification and analysis of multivariate observations. *In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Volume I, Statistics*. Edited by Lucien M. Le Cam and Jerzy Neyman. University of California Press.
- [6] Sun Y, Wong AKC, Kamel MS. 2009. Classification of imbalanced data: review. *Internation J Pattern Recognition Artific Intelligen*. 23(4):687-719.
- [7] Fawcett, T. (2006), "An introduction to ROC analysis", *Pattern Recognition Letters*, Vol. 27, hal. 861–874.