

Sistem Penjurusan IPA/IPS Menggunakan Algoritma K-Nearest Neighbor Pada SMA Muhammadiyah 13 Jakarta

Bayu Agung Rahmadi, Mufti*

Fakultas Teknologi Informasi, Program Studi Teknik Informatika, Universitas Budi Luhur, Jakarta, Indonesia
Email: ¹bayuagungg666@gmail.com, ^{2*}muftyhayat@gmail.com

Abstrak

Saat ini perkembangan teknologi dan informasi sangat pesat di era globalisasi, banyak manfaat yang telah membuat pekerjaan manusia lebih mudah dengan teknologi, salah satunya dengan pemanfaatan data mining. Data Mining merupakan proses ataupun kegiatan untuk mengumpulkan data yang berukuran besar kemudian mengekstraksi data tersebut menjadi informasi – informasi yang nantinya dapat digunakan. SMA Muhammadiyah 13 Jakarta merupakan salah satu Sekolah Menengah Atas Negeri di Kota Jakarta yang sudah menerapkan Kurikulum 2013 pada saat menentukan penjurusan. Dimana pada Kurikulum 2013 ini proses untuk menentukan jurusan dilakukan pada saat peserta didik memasuki tahun ajaran pertamanya di kelas X. Dalam proses penentuan jurusan, Bagian Kurikulum masih menggunakan cara manual yaitu dengan melakukan suatu perhitungan dengan beberapa pertimbangan dari para guru koordinator jurusan dan guru BK yang memerlukan banyak waktu dan dianggap tidak efisien. Selanjutnya proses penentuan jurusan, Bagian Kurikulum masih menggunakan cara perhitungan rata-rata penjurusan siswa dengan menggunakan Microsoft Excel. Tujuan dari penelitian ini adalah untuk membantu SMA Muhammadiyah 13 Jakarta dalam menentukan jurusan siswa-siswinya dengan konsep menggali informasi yang terdapat pada suatu data yang besar, pendekatan data mining diharapkan dapat menjawab permasalahan tersebut. Penulis menggunakan metode K - Nearest Neighbor dengan pengukuran jarak Euclidean Distance dan pengurutan menggunakan Insertion Sorting untuk mengklasifikasikan siswa ke dalam dua jurusan yaitu IPA dan IPS. K - Nearest Neighbor merupakan salah satu metode klasifikasi yang menghitung jarak terdekat untuk mengklasifikasikan objek baru berdasarkan kedekatannya dengan objek lama yang sudah diklasifikasikan sebelumnya. Sistem ini dapat memprediksi siswa terkait untuk menentukan jurusan. Waktu yang dibutuhkan dalam proses penentuan jurusan terbilang cukup cepat, yaitu hanya membutuhkan waktu paling lama 50.093 detik dengan jumlah data 306 record pada tahun ajaran 2017/2018.

Kata Kunci: *Data Mining*, klasifikasi, *K-Nearest Neighbor*, *Euclidean Distance*, Penjurusan

1. PENDAHULUAN

Saat ini masih banyak siswa Sekolah Menengah Atas (SMA) yang masih ragu dalam menentukan jurusan apa yang sesuai dengan kemampuan mereka masing-masing dan pihak sekolah masih melakukan proses penjurusan dengan cara manual dan membutuhkan waktu yang lama. Masalah ini sangat penting karena berpengaruh pada saat siswa memilih jurusan tersebut. Biasanya untuk mengatasi masalah tersebut siswa dapat berkonsultasi langsung kepada guru wali kelas, guru BK, maupun orang tua masing-masing.

Oleh sebab itu, untuk mendapatkan keputusan yang tepat, SMA Muhammadiyah 13 Jakarta ini membutuhkan suatu sistem penjurusan yang dapat membantu bagian kurikulum untuk menentukan jurusan siswa sesuai dengan nilai, minat, kecerdasan / IQ, kepribadian, dan kemampuan yang dimiliki siswa. SMA Muhammadiyah 13 Jakarta adalah salah satu Sekolah Menengah Atas yang ada di Kota Jakarta yang sudah menerapkan Kurikulum 2013 pada sistem penjurusannya. SMA Muhammadiyah 13 Jakarta setiap tahunnya memiliki siswa sebanyak 300 siswa lebih dengan masing-masing kelas jurusan IPA dan jurusan IPS. Dimana pada Kurikulum yang digunakan pada saat Sekolah Menengah Atas (SMA) saat ini adalah kurikulum 2013.

Perkembangan kurikulum yang masih baru ini terdapat beberapa perbedaan dengan kurikulum sebelumnya, salah satunya adalah perbedaan dalam menentukan penjurusan pada siswa-siswi SMA. Kurikulum yang baru penjurusannya dilakukan pada saat awal masuk, yaitu pada kelas X. Perubahan kurikulum dimaksudkan agar memungkinkan penyesuaian program pendidikan pada satuan pendidikan dengan kondisi dan potensi yang ada di daerah peserta didik[1].

Jika siswa mengalami kesalahan dalam penjurusan kemungkinan yang akan terjadi adalah rendahnya prestasi belajar siswa atau dapat menyebabkan terjadinya ketidakcocokan dengan jurusan yang sudah dipilih oleh siswa atau siswi sebelumnya. Proses yang dilakukan dalam penentuan jurusan tersebut masih dilakukan dengan menggunakan *Microsoft Excel* oleh bagian kurikulum dibantu dengan guru koordinator jurusan IPA dan IPS dengan melihat parameter atau karakteristik tertentu dan membutuhkan waktu yang lama bisa berhari-hari.

Pada kasus ini penulis menggunakan klasifikasi dengan algoritma *K-Nearest Neighbor* dan metode pengukuran jarak yang digunakan pada penulisan tugas akhir ini menggunakan metode pengukuran jarak *Euclidean Distance* untuk memprediksi siswa-siswi dalam menentukan pilihan jurusan yang akan mereka ambil. Diharapkan algoritma *K-Nearest Neighbor* dengan metode pengukuran jarak *Euclidean Distance* ini dapat menghasilkan tingkat akurasi yang baik dalam pengimplementasiannya pada *data mining* untuk menentukan jurusan siswa-siswi SMA Muhammadiyah 13 Jakarta.

2. METODE PENELITIAN

Metode penelitian yang diusulkan dalam penelitian ini melingkupi aktifitas berikut:

2.1 Sumber Data

Data merupakan kunci utama dari *datamining*, dimana data dan informasi dapat dikatakan baik apabila dapat mewakili keadaan objek yang sedang diteliti dan sesuai dengan pokok permasalahan yang dihadapi.

Pengumpulan data pada penelitian ini dilakukan dengan memperoleh dataset penelitian. Dataset penelitian pada kasus ini diperoleh dari data hasil penjurusan siswa SMA Muhammadiyah 13 Jakarta. Dataset yang digunakan adalah data hasil penjurusan siswa selama empat tahun terakhir, detail dari dataset dapat dilihat pada Tabel 1.

Tabel 1. Dataset Penelitian

No	Tahun Ajaran	Jumlah Data
1	2014/2015	303
2	2015/2016	396
3	2016/2017	354

2.2 Data Cleaning (Pembersihan Data)

Pada proses ini dilakukan pembersihan data yang dianggap sebagai *outlier* atau data yang mengandung *noise*. Kriteria data yang dianggap sebagai *outlier* diantaranya adalah:

- Data dengan nilai Psikotes dan nilai *Meant Test* bernilai 0 atau *null*.
- Duplicate Data (duplikasi data).

2.3 Data Selection (Pemilihan Data)

Proses ini bertujuan untuk memilih atribut yang relevan pada data untuk dapat dilakukan suatu analisis data. Pemilihan atribut didasari oleh hasil diskusi penulis dengan Bidang Kurikulum SMA Muhammadiyah 13 Jakarta yang menghasilkan beberapa keputusan terkait atribut (variabel) yang berpengaruh dalam penentuan jurusan siswa SMA Muhammadiyah 13 Jakarta. Atribut tersebut diantaranya adalah:

- Nilai Ujian Nasional
- Nilai Matematika
- Nilai IPA
- Nilai IPS
- Minat Siswa
- Jurusan

2.4 Data Transformation (Transformasi Data)

Pada proses ini dilakukan perubahan data ke dalam bentuk yang sesuai dengan algoritma klasifikasi yang digunakan. Yang dimana sebelumnya diketahui bahwa algoritma *K-Nearest Neighbor* ini tidak cocok jika dilakukan terhadap atribut yang bertipe kategorikal, sehingga perlu dilakukan transformasi data pada atribut minat menjadi atribut dengan tipe numerikal seperti dibawah ini.

Tabel 2. Transformasi Data

Kategorikal Minat	Hasil Transformasi
IPA	1
IPS	0

Selain itu, mengacu pada kebijakan yang berlaku pada SMA Muhammadiyah 13 Jakarta terkait sistem penentuan jurusannya, maka nilai Psikotes dan nilai *Mean Test* juga di transformasi menjadi nilai yang sesuai dengan kebijakan yang ada. Yaitu dengan menentukan dari nilai Ujian Nasional dan nilai *Mean Test* tersebut seperti yang dituliskan pada persamaan di bawah ini.

Ujian Nasional/UN

$$\text{Rata-rata Test} = (\text{Matematika} * 2 + \text{IPA} + \text{IPS})$$

- Ujian Nasional untuk menghitung jarak kedekatan *Euclidean Distance* antara dataset uji dengan masing-masing data yang terdapat pada dataset model secara berurutan.
- Nilai *mean test* didapat dari nilai matematika, ipa, dan ips.

2.5 Algoritma K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* merupakan salah satu algoritma klasifikasi yang cukup populer dalam *data mining* menurut[9]. Algoritma klasifikasi ini termasuk kedalam algoritma *instance-based classifier* yang dimana proses klasifikasi baru hanya akan dilakukan ketika ada objek yang ingin diketahui kelasnya, sehingga algoritma tidak membangun suatu model dari data pelatihan melainkan data pelatihan itu sendiri yang merepresentasikan suatu *knowledge* dari model klasifikasi tersebut. *Instance-based classifier* sering juga disebut sebagai metode pembelajaran *lazy learner*[10]. Hal ini sangat berbeda dengan konsep pembelajaran *eager learner* yang membangun suatu model dari data pelatihan yang kemudian model tersebut akan digunakan untuk mengklasifikasikan objek baru dan kemudian mengabaikan data pelatihannya. Algoritma *K-Nearest Neighbor* juga merupakan jenis algoritma pembelajaran terawasi (*supervised learning*) dimana algoritma *K-Nearest Neighbor* bertujuan untuk menentukan suatu kelas *query instance* yang baru terhadap data-data yang sudah ditentukan atau diketahui kelasnya.

Alur kerja dari algoritma *K-Nearest Neighbor* menurut[9] adalah sebagai berikut:

- Tentukan nilai K, sebagai parameter ketetanggan terdekat

- b. Hitung jarak antara *query instance* dan semua data pelatihan
- c. Urutkan berdasarkan jarak terkecil
- d. Tentukan label kelas *query instance* berdasarkan label mayoritas pada K objek data latih terdekat

2.6 Euclidean Distance

Pengukuran jarak (*distance measure*) digunakan untuk menentukan tingkat kedekatan dua buah objek, dimana hasil dari perhitungan kedekatan akan menentukan dua buah objek memiliki kesamaan (*similarity*) atau tidak[9]. *Euclidean Distance* mengukur kedekatan antara dua buah objek yang digambarkan sebagai pengukuran garis lurus atau secara langsung[11]. Metode pengukuran ini cocok apabila diimplementasikan terhadap 14 data yang memiliki nilai atribut bersifat *numerikal*, khususnya dengan atribut kontinu[12]. *Euclidean Distance* dihitung dengan persamaan dibawah ini:

$$d_{Euclidean}(x, y) = |x - y| = \sqrt{\sum_i^n (x_i - y_i)^2}$$

Sebagai contoh, terdapat 2 buah objek yang akan diukur tingkat kedekatannya dengan menggunakan *Euclidean Distance*. Dimana 2 buah objek tersebut adalah sebagai berikut:

$$X = [9, 2, 0, 3]$$

$$Y = [1, 3, 7, 5]$$

Maka, jarak antara objek X dan Y jika dihitung adalah:

$$\begin{aligned} d_{Euclidean}(x,y) &= \sqrt{(9-1)^2 + (2-3)^2 + (0-7)^2 + (3-5)^2} \\ &= \sqrt{64 + 1 + 49 + 4} \\ &= 10,86 \end{aligned}$$

Berarti tingkat kedekatan antara objek X dan Y adalah 10,86.

3. ANALISA DAN PEMBAHASAN

3.1 Analisa

Saat ini masih banyak siswa Sekolah Menengah Atas (SMA) Muhammadiyah 13 Jakarta yang masih ragu dalam menentukan jurusan apa yang sesuai dengan kemampuan mereka masing-masing dan pihak sekolah juga masih melakukan proses penjurusan dengan cara manual dan membutuhkan waktu yang lama. Masalah ini sangat penting karena berpengaruh pada saat siswa memilih jurusan tersebut.

Oleh sebab itu, untuk mendapatkan keputusan yang tepat, SMA Muhammadiyah 13 Jakarta ini membutuhkan suatu sistem penjurusan yang dapat membantu bagian kurikulum untuk menentukan jurusan siswa sesuai dengan nilai, minat, kecerdasan / IQ, kepribadian, dan kemampuan yang dimiliki siswa.

3.2 Solusi

Dari permasalahan diatas, maka dibuatlah sebuah sistem penjurusan dalam sebuah aplikasi yang dapat menentukan penjurusan IPA/IPS terhadap siswa-siswi SMA Muhammadiyah 13 Jakarta. Dan berikut adalah tampilan dari aplikasi yang telah dibuat dengan berbasis desktop:

Tampilan Form Dataset

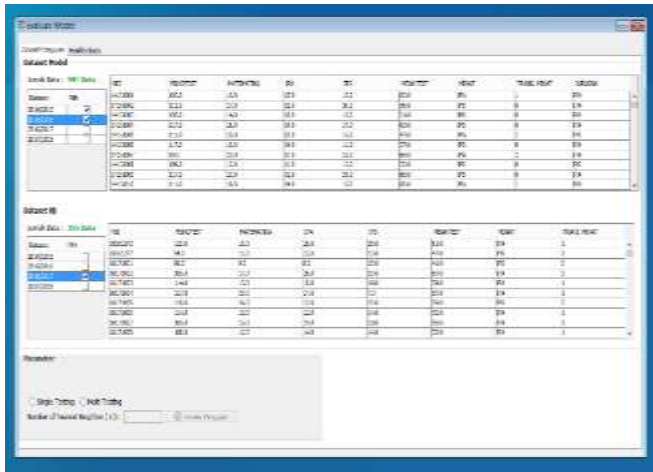
Jika *user* memilih *form* Dataset maka akan muncul tampilan *form* Dataset. Di dalam *form* Dataset terdapat aktivitas *user* yaitu mengirim *port* dataset hasil penjurusan siswa yang sudah pernah dilakukan sebelumnya untuk digunakan sebagai dataset model dalam menentukan jurusan. Tampilan *Form* Dataset dapat dilihat pada gambar 1.



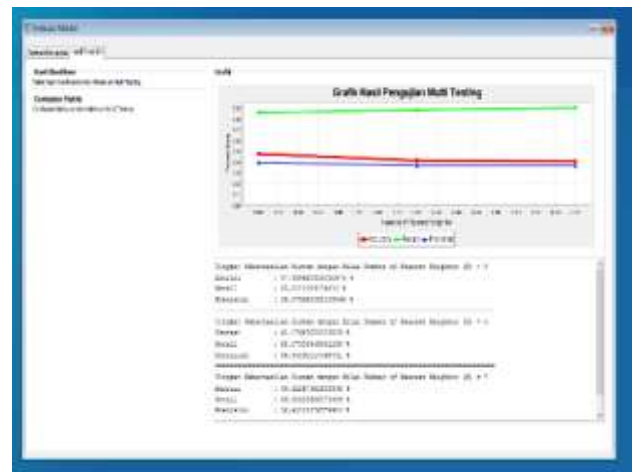
Gambar 1. Tampilan Form Dataset

Tampilan Form Evaluasi Model

Form Evaluasi Model dapat digunakan oleh *user* untuk melakukan pengujian data. Pada tab Dataset Pengujian, *user* dapat memilih dataset mana saja yang akan digunakan dalam proses pengujian melalui dua buah tabel yang telah dikelompokkan berdasarkan fungsinya, yakni tabel dataset model dan tabel dataset uji.



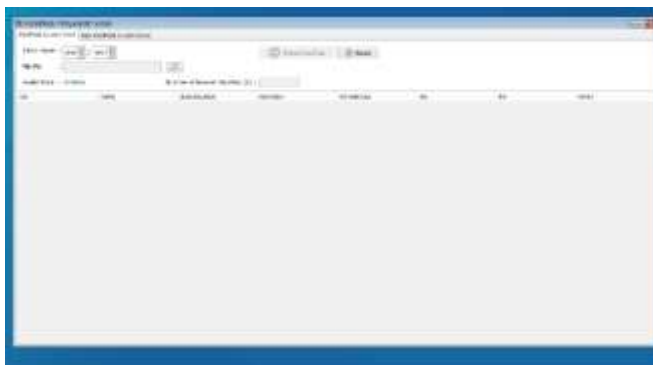
Gambar 2. Tampilan Form Evaluasi Model



Gambar 3. Tampilan Hasil Evaluasi Pengujian

Tampilan Form Klasifikasi Penjurusan Siswa

Berikut adalah tampilan yang akan ditampilkan kepada *user* saat akan melakukan klasifikasi atau penjurusan terhadap dataset siswa yang ingin diketahui jurusannya.



Gambar 4. Tampilan Form Klasifikasi Penjurusan Siswa

No	Nama	Departemen	Nilai	Nilai	Nilai	Nilai	Nilai
1	Adi Nugroho	SI	85	85	85	85	85
2	Adi Nugroho	SI	85	85	85	85	85
3	Adi Nugroho	SI	85	85	85	85	85
4	Adi Nugroho	SI	85	85	85	85	85
5	Adi Nugroho	SI	85	85	85	85	85
6	Adi Nugroho	SI	85	85	85	85	85
7	Adi Nugroho	SI	85	85	85	85	85
8	Adi Nugroho	SI	85	85	85	85	85
9	Adi Nugroho	SI	85	85	85	85	85
10	Adi Nugroho	SI	85	85	85	85	85
11	Adi Nugroho	SI	85	85	85	85	85
12	Adi Nugroho	SI	85	85	85	85	85
13	Adi Nugroho	SI	85	85	85	85	85
14	Adi Nugroho	SI	85	85	85	85	85
15	Adi Nugroho	SI	85	85	85	85	85
16	Adi Nugroho	SI	85	85	85	85	85
17	Adi Nugroho	SI	85	85	85	85	85
18	Adi Nugroho	SI	85	85	85	85	85
19	Adi Nugroho	SI	85	85	85	85	85
20	Adi Nugroho	SI	85	85	85	85	85
21	Adi Nugroho	SI	85	85	85	85	85
22	Adi Nugroho	SI	85	85	85	85	85
23	Adi Nugroho	SI	85	85	85	85	85
24	Adi Nugroho	SI	85	85	85	85	85
25	Adi Nugroho	SI	85	85	85	85	85
26	Adi Nugroho	SI	85	85	85	85	85
27	Adi Nugroho	SI	85	85	85	85	85
28	Adi Nugroho	SI	85	85	85	85	85
29	Adi Nugroho	SI	85	85	85	85	85
30	Adi Nugroho	SI	85	85	85	85	85
31	Adi Nugroho	SI	85	85	85	85	85
32	Adi Nugroho	SI	85	85	85	85	85
33	Adi Nugroho	SI	85	85	85	85	85
34	Adi Nugroho	SI	85	85	85	85	85
35	Adi Nugroho	SI	85	85	85	85	85
36	Adi Nugroho	SI	85	85	85	85	85
37	Adi Nugroho	SI	85	85	85	85	85
38	Adi Nugroho	SI	85	85	85	85	85
39	Adi Nugroho	SI	85	85	85	85	85
40	Adi Nugroho	SI	85	85	85	85	85
41	Adi Nugroho	SI	85	85	85	85	85
42	Adi Nugroho	SI	85	85	85	85	85
43	Adi Nugroho	SI	85	85	85	85	85
44	Adi Nugroho	SI	85	85	85	85	85
45	Adi Nugroho	SI	85	85	85	85	85
46	Adi Nugroho	SI	85	85	85	85	85
47	Adi Nugroho	SI	85	85	85	85	85
48	Adi Nugroho	SI	85	85	85	85	85
49	Adi Nugroho	SI	85	85	85	85	85
50	Adi Nugroho	SI	85	85	85	85	85
51	Adi Nugroho	SI	85	85	85	85	85
52	Adi Nugroho	SI	85	85	85	85	85
53	Adi Nugroho	SI	85	85	85	85	85
54	Adi Nugroho	SI	85	85	85	85	85
55	Adi Nugroho	SI	85	85	85	85	85
56	Adi Nugroho	SI	85	85	85	85	85
57	Adi Nugroho	SI	85	85	85	85	85
58	Adi Nugroho	SI	85	85	85	85	85
59	Adi Nugroho	SI	85	85	85	85	85
60	Adi Nugroho	SI	85	85	85	85	85
61	Adi Nugroho	SI	85	85	85	85	85
62	Adi Nugroho	SI	85	85	85	85	85
63	Adi Nugroho	SI	85	85	85	85	85
64	Adi Nugroho	SI	85	85	85	85	85
65	Adi Nugroho	SI	85	85	85	85	85
66	Adi Nugroho	SI	85	85	85	85	85
67	Adi Nugroho	SI	85	85	85	85	85
68	Adi Nugroho	SI	85	85	85	85	85
69	Adi Nugroho	SI	85	85	85	85	85
70	Adi Nugroho	SI	85	85	85	85	85
71	Adi Nugroho	SI	85	85	85	85	85
72	Adi Nugroho	SI	85	85	85	85	85
73	Adi Nugroho	SI	85	85	85	85	85
74	Adi Nugroho	SI	85	85	85	85	85
75	Adi Nugroho	SI	85	85	85	85	85
76	Adi Nugroho	SI	85	85	85	85	85
77	Adi Nugroho	SI	85	85	85	85	85
78	Adi Nugroho	SI	85	85	85	85	85
79	Adi Nugroho	SI	85	85	85	85	85
80	Adi Nugroho	SI	85	85	85	85	85
81	Adi Nugroho	SI	85	85	85	85	85
82	Adi Nugroho	SI	85	85	85	85	85
83	Adi Nugroho	SI	85	85	85	85	85
84	Adi Nugroho	SI	85	85	85	85	85
85	Adi Nugroho	SI	85	85	85	85	85
86	Adi Nugroho	SI	85	85	85	85	85
87	Adi Nugroho	SI	85	85	85	85	85
88	Adi Nugroho	SI	85	85	85	85	85
89	Adi Nugroho	SI	85	85	85	85	85
90	Adi Nugroho	SI	85	85	85	85	85
91	Adi Nugroho	SI	85	85	85	85	85
92	Adi Nugroho	SI	85	85	85	85	85
93	Adi Nugroho	SI	85	85	85	85	85
94	Adi Nugroho	SI	85	85	85	85	85
95	Adi Nugroho	SI	85	85	85	85	85
96	Adi Nugroho	SI	85	85	85	85	85
97	Adi Nugroho	SI	85	85	85	85	85
98	Adi Nugroho	SI	85	85	85	85	85
99	Adi Nugroho	SI	85	85	85	85	85
100	Adi Nugroho	SI	85	85	85	85	85

Gambar 5. Tampilan Hasil Klasifikasi Jurusan Siswa

Hasil Uji Coba Program

Pada pengujian unjuk kerja sistem ini memiliki 2 jenis pengujian, yaitu pengujian *multi testing* dan pengujian *single testing*. Dimana kedua pengujian ini juga dilakukan dengan tujuan menentukan nilai K (*number of nearest neighbor*) terbaik dalam melakukan proses klasifikasi.

a. Pengujian Multi Testing

Pengujian *multi testing* berikut dilakukan dengan tujuan untuk mengevaluasi model yang ada menggunakan beberapa parameter nilai ketetanggaan terdekat (*number of nearest neighbor*) atau nilai K. Pada pengujian ini dilakukan 4 kali pengujian, dimana semua dataset masing-masing akan dijadikan sebagai dataset uji (*testing dataset*) dan juga dataset model (*training dataset*), dengan perbandingan jumlah dataset pada tiap pengujian yakni 1:2.

Pengujian Multi Testing Pertama, Dataset model yang digunakan adalah dataset tahun ajaran 2014/2015 dan 2015/2016 dimana sebagai data ujinya digunakan dataset tahun ajaran 2016/2017.

Tabel 3. Multi Testing Pertama

Nilai K	Accuracy	Recall	Precision
3	49.29%	83.89%	44.48%
5	47.32%	86.57%	43.58%
7	47.60%	87.24%	43.77%
9	48.45%	88.59%	44.29%
11	48.73%	90.60%	44.54%

Pengujian Multi Testing Kedua, Dataset model yang digunakan adalah dataset tahun ajaran 2014/2015 dan 2016/2017 dimana sebagai data ujinya digunakan dataset tahun ajaran 2015/2016.

Tabel 4. Multi Testing Kedua

Nilai K	Accuracy	Recall	Precision
3	47.38%	85.32%	39.07%
5	41.17%	88.07%	36.50%
7	40.52%	89.90%	36.43%
9	40.84%	92.66%	36.86%
11	41.17%	91.74%	36.90%

Pengujian Multi Testing Ketiga, Dataset model yang digunakan adalah dataset tahun ajaran 2015/2016 dan 2016/2017 dimana sebagai data ujinya digunakan dataset tahun ajaran 2014/2015.

Tabel 5. Multi Testing Ketiga

Nilai K	Accuracy	Recall	Precision
3	61.16%	53.41%	65.64%
5	58.89%	47.20%	64.40%
7	56.31%	43.47%	61.40%
9	54.04%	40.99%	58.40%
11	54.04%	40.99%	58.40%

b. Pengujian Single Testing

Pengujian *single testing* merupakan pengujian yang dilakukan dengan menggunakan suatu nilai parameter ketetangaan terdekat. Pengujian ini bertujuan untuk melihat secara detail hasil klasifikasi dan hasil evaluasi yang dihasilkan oleh suatu nilai parameter tersebut. penentuan dataset pada pengujian ini menggunakan teknik yang sama dengan pengujian *multi testing* dengan perbandingan jumlah dataset pada tiap pengujian adalah 1:2. Nilai K yang digunakan adalah 3.

Pengujian Single Testing Pertama ini dilakukan dengan menggunakan dataset model tahun ajaran 2014/2015 dan 2015/2016, dimana sebagai data ujinya digunakan dataset tahun ajaran 2016/2017.

Confusion Matrix

	Predicted IPA	Predicted IPS	Class Recall
Actual IPA	125	24	83.89 %
Actual IPS	156	50	24.27 %
Class Precision	44.48 %	67.57 %	

***Accuracy**
Overall Accuracy is 49.3 %

Gambar 6. Single Testing Pertama

Pengujian Single Testing Kedua ini dilakukan dengan menggunakan dataset model tahun ajaran 2014/2015 dan 2016/2017, dimana sebagai data ujinya digunakan dataset tahun ajaran 2015/2016.

Confusion Matrix

	Predicted IPA	Predicted IPS	Class Recall
Actual IPA	78	111	41.27 %
Actual IPS	36	57	51.29 %
Class Precision	68.42 %	33.93 %	

***Accuracy**
Overall Accuracy is 47.87 %

Gambar 7. Single Testing Kedua

Pengujian Single Testing Ketiga ini dilakukan dengan menggunakan dataset model tahun ajaran 2015/2016 dan 2016/2017, dimana sebagai data ujinya digunakan dataset tahun ajaran 2014/2015.

Confusion Matrix

	Predicted IPA	Predicted IPS	Class Recall
Actual IPA	75	34	68.81 %
Actual IPS	103	94	47.72 %
Class Precision	42.13 %	73.44 %	

***Accuracy**
Overall Accuracy is 55.23 %

Gambar 8. Single Testing Ketiga

4. KESIMPULAN

Melalui proses pengerjaan dan pengujian dalam penelitian ini, maka dapat disimpulkan beberapa hal, yaitu:

- a. Waktu yang dibutuhkan dalam proses penentuan jurusan siswa terbilang cukup cepat apabila dibandingkan dengan proses penentuan jurusan sebelumnya yang memakan waktu hingga 3-4 hari, yaitu hanya membutuhkan waktu paling lama 50.093 detik dengan jumlah data 306 *record*.
- b. Algoritma *K-Nearest Neighbor* dengan metode pengukuran jarak *Euclidean Distance* dapat diimplementasikan ke dalam proses sistem prediksi jurusan siswa.

REFERENCES

- [1] Kemendikbud_RI, "Permendikbud RI Nomor 81A tahun 2013 tentang Implementasi Kurikulum," p. 88, 2013.
- [2] Obbie Kristanto, "Penerapan Algoritma Klasifikasi Data Mining Id3 Untuk Menentukan Penjurusan Siswa SMAN 6," pp. 6–7, 2013.
- [3] H. Leidiyana, "Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor," *J. Penelit. Ilmu Komputer, Syst. Embed. Log.*, vol. 1, no. 1, pp. 65–76, 2013.
- [4] Y. S. Nugroho and S. N. Haryati, "Klasifikasi dan Klastering Penjurusan Siswa SMA Negeri 3 Boyolali," *Khazanah Inform.*, vol. 1, no. 1, pp. 3–8, 2015.
- [5] S. M. A. P. Banjarbaru, "Penerapan Metode Learning Vector Quantization (Lvq) Pada Prediksi Jurusan Di," vol. 4, no. 1, pp. 11–20, 2016.
- [6] K. Latifah, "Kombinasi Algorithma K-NN dan Manhattan Distance untuk Menentukan Pemenang Lelang," pp. 49–58.
- [7] R. K. Niswatin, "Penempatan Jurusan Mahasiswa Baru Menggunakan Metode K-Nearest," vol. 1, no. 1, pp. 55–67, 2013.
- [8] A. Sulistyio, "Penentuan Jurusan Sekolah Menengah Atas Menggunakan Metode K-Nearest Neighbor Classifier Pada SMAN 16 Semarang," pp. 1–5, 2015.
- [9] "No Title," 2017.
- [10] Sillueta, "IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI KELULUSAN Perpustakaan Universitas Budi Luhur," vol. 1, no. Huda 2010, 2013.
- [11] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*. 2012.
- [12] F. Gorunescu, *Data Mining: Concepts, Models and Techniques*. 2011.