

APLIKASI PREDIKSI KELULUSAN UJIAN NASIONAL MENGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR* DENGAN PENGUKURAN JARAK *MANHATTAN DISTANCE*

Ary Maulana Malik¹⁾, Alexander J.P. Sibarani²⁾

^{1,2}Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur

^{1,2}Jl. Raya Ciledug, Petukangan Utara, Kebayoran Lama, Jakarta Selatan 12260

E-mail : arymalik25@gmail.com¹⁾, alexander.sibarani@budiluhur.ac.id²⁾

Abstrak

Ujian Nasional merupakan evaluasi standar yang dilakukan secara nasional oleh pusat pendidikan untuk dijadikan syarat kelulusan seorang siswa/i. Sebelum menempuh ujian nasional siswa/i akan melakukan ujian Try Out dimana hal itu dilakukan sebagai tolak ukur sebelum menempuh ujian nasional yang sebenarnya. Pemanfaatan data kelulusan belum efisien dan maksimal, hal ini membuat tingkat kelulusan siswa/i belum diketahui dengan mudah dan cepat. Untuk memprediksi tingkat kelulusan, dapat memanfaatkan data-data yang ada khususnya data kelulusan. Banyaknya data membuat proses prediksi memakan waktu yang cukup lama, sehingga diperlukan sebuah sistem yang bisa meningkatkan waktu prediksi tingkat kelulusan siswa/i. Pada penelitian ini membahas sistem prediksi kelulusan dengan mengimplementasikan algoritma *K-Nearest Neighbor* dengan pengukuran jarak *Manhattan Distance*. Hasil dari penelitian ini akan digunakan sebagai pendukung keputusan agar siswa/i siap untuk menghadapi ujian nasional sejak dini. Hasil uji coba dengan menggunakan set data siswa dari 3 tahun terakhir yang berjumlah 124 jurusan IPA dan 270 jurusan IPS. Kemudian dengan parameter (*K*) yaitu $K=9$ dan $K=13$, maka dihasilkan rata-rata akurasi pada jurusan IPA 68,29% pada tahun ajaran 2013/2014, 82,52% pada tahun ajaran 2014/2015, 84,88% pada tahun ajaran 2015/2016, dan untuk jurusan IPS 74,71% pada tahun ajaran 2013/2014, 77,89% pada tahun ajaran 2014/2015, 70,23% pada tahun ajaran 2015/2016. Berdasarkan hasil dari serangkaian proses pengujian yang dilakukan dapat disimpulkan bahwa proses prediksi yang dilakukan menjadi lebih cepat dibandingkan menggunakan perhitungan manual pada *Microsoft Excel*.

Kata kunci: *K-Nearest Neighbor*, *Manhattan Distance*, Prediksi, Ujian Nasional.

1. PENDAHULUAN

Perkembangan teknologi yang semakin pesat pada era globalisasi ini telah memberikan manfaat dalam kemajuan disegala macam aspek. Berkembangannya teknologi ini harus diikuti dengan perkembangan pada sumber daya manusia. Manusia sebagai pengguna dari teknologi tersebut harus bisa memanfaatkan teknologi yang sudah ada saat ini, maupun perkembangan teknologi yang akan terjadi selanjutnya. Teknologi baru yang telah berkembang harus dapat diadaptasi untuk bidang pendidikan. Dengan begitu, teknologi dan pendidikan akan mampu berkembang seiring dengan adanya generasi baru sebagai generasi lama.

Sistem yang dibuat manusia bertujuan untuk meringankan pekerjaan manusia itu sendiri, meningkatkan produktivitas, dan menghemat waktu saat melakukan pekerjaan yang dibutuhkan. Saat ini beberapa pekerjaan yang dilakukan manusia menggunakan komputer, misalnya pertahanan negara, ilmu pengetahuan, pengobatan dan sebagainya. Ujian nasional adalah evaluasi standar pendidikan secara nasional yang dilakukan oleh pusat pendidikan yang dijadikan sebagai salah satu syarat kelulusan. Sebelum melakukan ujian nasional para siswa/i akan melakukan ujian Try Out dimana hal ini dilakukan sebagai uji coba sebelum melaksanakan ujian nasional. Data kelulusan yang sudah ada belum dimanfaatkan dengan maksimal dan efisien, hal ini

membuat tingkat kelulusan siswa/i belum diketahui dengan mudah dan cepat. Untuk memprediksi tingkat kelulusan, maka dapat memanfaatkan data-data yang ada khususnya data kelulusan. Peneliti mencoba untuk mendapatkan informasi yang berguna untuk mendukung proses prediksi tingkat kelulusan siswa/i.

Konsep penelitian dengan algoritma *K-Nearest Neighbor* sudah banyak dilakukan oleh peneliti-peneliti sebelumnya diantaranya adalah *K-Nearest Neighbor* merupakan teknik yang sangat sederhana, efektif, dan efisien dalam bidang pengenalan pola, kategori teks, pengolahan objek dan lain-lain, karena kesederhanaan pengolahannya dan mampu melakukan *training* data dalam jumlah besar [2]. Salah satu masalah dari algoritma ini adalah efek yang sama dari semua atribut yang terdapat pada data baru dan data lama dalam dataset pelatihan [7]. Pada penelitian ini digunakan algoritma *K-Nearest Neighbor* untuk melakukan klasifikasi data dan untuk mengukur akurasi dari algoritma *K-Nearest Neighbor*, digunakan metode pengukuran jarak *Manhattan Distance* karena metode ini sesuai untuk mengukur kedekatan data yang bersifat numerikal.

Permasalahan yang muncul adalah banyaknya data siswa/i dan memakan waktu yang cukup lama, sehingga diperlukan sebuah sistem yang dapat meningkatkan waktu prediksi kelulusan berdasarkan data *training* yang diambil dari data siswa/i yang sudah lulus sebelumnya. Sehingga pihak

penyelenggara pendidikan dapat memperbaiki nilai siswa/i sejak dini. Penggunaan algoritma *K-Nearest Neighbor* diharapkan mampu meningkatkan efektifitas dan efisiensi dalam proses prediksi tingkat kelulusan dan menghasilkan tingkat akurasi yang baik.

Tujuan penelitian ini adalah mengembangkan sebuah sistem yang dapat meningkatkan waktu dalam proses prediksi tingkat kelulusan ujian nasional, dan mengetahui kinerja algoritma *K-Nearest Neighbor* dengan pengukuran jarak *Manhattan Distance* pada evaluasi *Confusion Matrix*.

2. METODE PENELITIAN

2.1. Metode Pengumpulan Data

Dalam penelitian ini untuk mendapatkan informasi yang dibutuhkan dalam menyelesaikan masalah yang ditemui. Adapun metode tersebut sebagai berikut:

a. Studi Literatur

Metode ini menggunakan pembelajaran dengan cara mengumpulkan, membaca, dan memahami referensi dengan baik dalam bentuk laporan penelitian, jurnal ilmiah, artikel diinternet dan lain-lain .

b. Observasi

Metode ini menggunakan pembelajaran dengan cara melakukan wawancara untuk memperoleh informasi tentang parameter apa saja yang digunakan sebagai prediksi kelulusan ujian nasional

c. Pengembangan Prototype

Metode ini merupakan salah satu metode pengembangan yang banyak digunakan. Metode ini dimulai dengan pengumpulan data yang dibutuhkan, mengidentifikasi segala kebutuhan, melakukan perancangan kilat yang difokuskan pada penyajian aspek yang diperlukan.

2.2. Sumber Data

Pengumpulan data pada penelitian ini dilakukan dengan memperoleh data primer dan sekunder. Data primer diperoleh dari data hasil ujian *Try Out* dan Ujian Nasional. Sedangkan data sekunder diperoleh dari buku, ebook, laporan penelitian, jurnal ilmiah dan lain-lain sebagainya untuk bahan referensi dan memperoleh kajian teoritis yang berkaitan dengan algoritma *K-Nearest Neighbor*.

2.3. Pembersihan Data

Pembersihan data dilakukan dengan membersihkan data yang dianggap sebagai *Outlier* atau data yang mengandung *Noise*. Kriteria yang dianggap sebagai *Outlier* diantaranya sebagai berikut:

- a. Data dengan nilai *Try Out* bernilai 0 atau null
- b. Duplikasi data

2.4. Pemilihan Data

Proses pemilihan data dilakukan dengan tujuan memilih atribut yang relevan pada data. Pemilihan atribut didasari oleh hasil diskusi dengan bidang kurikulum pihak penyelenggara pendidikan yang menghasilkan beberapa atribut (variabel) yang berpengaruh pada kelulusan ujian nasional. Atribut tersebut sebagai berikut:

- a. Nilai *Try Out* pertama Matematika
- b. Nilai *Try Out* pertama Bahasa Indonesia
- c. Nilai *Try Out* pertama Bahasa Inggris
- d. Nilai *Try Out* pertama IPA/IPS (rata-rata)
- e. Nilai *Try Out* kedua Matematika
- f. Nilai *Try Out* kedua Bahasa Indonesia
- g. Nilai *Try Out* kedua Bahasa Inggris
- h. Nilai *Try Out* kedua IPA/IPS (rata-rata)
- i. Nilai *Try Out* Keterangan
- j. Nilai *Try Out* Hasil Ujian Nasional

2.5. Transformasi Data

Transformasi data dilakukan dengan tujuan merubah data kedalam bentuk yang sesuai dengan algoritma klasifikasi yang digunakan. Transformasi pada atribut Keterangan menjadi atribut yang bersifat numerikal.

Tabel 1 : Transformasi Data Atribut Keterangan

Atribut Keterangan	Hasil Transformasi
LULUS	1
TIDAK	0

Nilai ujian *Try Out* pertama dan kedua juga ditransformasi menjadi nilai yang sesuai dengan kebijakan yang ada, yaitu dengan merata-rata kan nilai *Try Out* pertama dan kedua seperti persamaan dibawah ini:

$$\text{Mean1}=(\text{MTK1}+\text{BINDO1}+\text{BING1}+\text{IPA1 atau IPS1})/4$$

$$\text{Mean2}=(\text{MTK2}+\text{BINDO2}+\text{BING2}+\text{IPA2 atau IPS2})/4$$

2.6. Algoritma *K-Nearest Neighbor*

K-Nearest Neighbor merupakan sebuah metode yang melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan keruang berdimensi banyak, dimana masing-masing dimensi mempresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi data pembelajaran. Nilai K data, secara umumnya, nilai K yang tinggi akan mengurangi efek noise pada klasifikasi, tetapi membuat batasan antara setiap klasifikasi menjadi lebih kabur [10].

Langkah-langkah untuk menghitung metode *K-Nearest Neighbor* antara lain [11] :

- a. Menentukan parameter K

- b. Menghitung jarak antara data yang akan dievaluasi
- c. Mengurutkan jarak yang terbentuk
- d. Menentukan jarak terdekat sampai urutan K
- e. Memasangkan kelas yang bersesuaian
- f. Mencari jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi

Algoritma *K-Nearest Neighbor* termasuk algoritma *supervised learning* dimana hasil dari *query instance* yang baru diklasifikasikan berdasarkan mayoritas dari kategori. *K-Nearest Neighbor* menghitung tingkat kemiripan (jarak) suatu kasus terhadap kasus lain berdasarkan beberapa atribut yang didefinisikan berdasarkan pada pembobotan tertentu dan kemudian tingkat (jarak) dari keseluruhan atribut akan dijumlahkan. *K-Nearest Neighbor* didefinisikan berdasarkan persamaan sebagai berikut [5]:

$$Similarity(T, S) = \sum_{i=1}^n f(T_i, S_i) W_i$$

Keterangan :

- T : Kasus target / baru
- S : Kasus sumber / lama / pembanding
- N : Jumlah atribut dalam setiap kasus
- I : Atribut individu dari 1 sampai n
- F : kemiripan atribut i dalam kasus T dan S
- W : Bobot atribut i

2.7. Manhattan Distance

Manhattan Distance digunakan untuk mengambil kasus yang cocok dari basis kasus dengan menghitung jumlah bobot absolut dari perbedaan antara kasus yang sekarang dan kasus yang lain dalam basis kasus untuk menghitung bobot yang digunakan sebagai berikut:

$$D_{ij} = \sum W_k |X_{ik} - C_{jk}|$$

Dimana diketahui D_{ij} adalah jarak antara kasus i dan j dengan parameternya. W mempresentasikan jumlah dari bobot. X adalah kasus yang baru dikurangi dengan C yaitu *history* (kasus ada dalam *Case Base*). *Manhattan distance* adalah pengukuran *similarity* / kemiripan yang paling cocok untuk *approval project* yang mempresentasikan kasus yang relevan dengan angka yang natural atau dengan data yang bersifat *kuantitative* dan menganalisa data[5]. Data hasil prediksi kelulusan akan memiliki acuan untuk pengambilan keputusan menjadi lebih akurat.

2.8. Selection Sorting

Selection Sorting adalah metode pengukuran yang membandingkan elemen yang sekarang dengan

elemen berikutnya sampai ke elemen yang terakhir. Jika ditemukan elemen lain yang lebih kecil dari elemen sekarang maka dicatat posisinya dan langsung ditukar, perbandingan dilakukan terus menerus sampai tidak ada lagi pertukaran [9]. Bila diketahui data awal berupa 44, 55, 12, 42, 94, 18, 6, 67, maka langkah pengurutan dengan metode *selection sorting*[12].

44	55	12	42	94	18	06	67	Data Awal
06	55	12	42	94	18	44	67	Tukarkan data ke 1 dengan data ke 7
06	12	55	42	94	18	44	67	Tukarkan data ke 2 dengan data ke 3
06	12	18	42	94	55	44	67	Tukarkan data ke 3 dengan data ke 6
06	12	18	42	94	55	44	67	Data ke 4 tidak ditukarkan
06	12	18	42	44	55	94	67	Data ke 5 ditukarkan dengan data ke 7
06	12	18	42	44	55	94	67	Data ke 6 tidak ditukarkan
06	12	18	42	44	55	67	94	Data ke 7 ditukarkan dengan data ke 8
06	12	18	42	44	55	67	94	Data setelah terurut

Gambar 1 : Langkah Pengurutan *Selection Sorting*

2.9. Majority Voting

Metode pelabelan data yang selama ini digunakan adalah menggunakan seorang pakar untuk melakukan pelabelan. Hal ini tentu sangat sulit atau bahkan tidak mungkin dilakukan untuk *dataset* berukuran besar yang membutuhkan lebih banyak pakar. Salah satu teknik pelabelan data adalah *Majority Voting*. Metode ini menggunakan konsep pengambilan keputusan hasil vote yang diperoleh dari jumlah terbesar dari masing-masing pilihan vote yang ada. Metode ini menggunakan rumus sebagai berikut (asumsi menggunakan 3 pelabelan)[1] :

$$C(X) = \text{mode} \{h1(X), h2(X), h3(X), \dots\}$$

Keterangan :

- C(X) : class/label X
- H1(X) : hasil vote pelabel 1
- H2(X) : hasil vote pelabel 2
- H3(X) : hasil vote pelabel 3

2.10. Confusion Matrix

Confusion Matrix adalah salah satu alat ukur berbentuk matrix 2x2 yang digunakan untuk mendapatkan jumlah ketepatan klasifikasi *dataset* terhadap kelas lulus dan tidak lulus pada algoritma yang di pakai [4]. Contoh tabel *Confusion Matrix* dimana jika *dataset* hanya terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya negatif [6].

Tabel 2: Model Confusion Matrix [11]

Klasifikasi yang benar	Diklasifikasikan sebagai	
	+	-
+	<i>true positives</i>	<i>false negatives</i>
-	<i>false positives</i>	<i>true negatives</i>

True positives adalah jumlah record positif yang diklasifikasi sebagai positif, *false positives* adalah

jumlah record negatif yang diklasifikasi sebagai positif, *false negatives* adalah record positif yang diklasifikasi sebagai negatif, *true negatives* adalah record negatif yang diklasifikasikan sebagai negatif. Kemudian dari hal tersebut dapat diperoleh beberapa pengukuran evaluasi diantaranya *accuracy*, *precision*, dan *recall (sensitivity)* [5] sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$

$$Precision = \frac{TP}{TP+FP}$$

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

Keterangan :

TP = *true positives*

FP = *false positives*

TN = *true negatives*

FN = *false negatives*

Accuracy adalah jumlah data yang diklasifikasikan dengan benar. *Recall* berhubungan dengan kemampuan sistem untuk memanggil dokumen relevan dengan *query*. Sedangkan *precision* berkaitan dengan kemampuan sistem untuk tidak memanggil dokumen yang tidak relevan dengan kebutuhan user [3].

2.11. Proses Prediksi dan Klasifikasi

Tabel 3 : *Dataset Model*

NIS	MTK1	BINDO1	BING1	IPS1	MTK2	BINDO2	BING2	IPS2	Ket	Hasil UN
141510001	70	70	73	78	68	68	77	62	LULUS	LULUS
141510005	83	80	56	83	62	82	30	63	TIDAK	TIDAK
141510008	78	82	78	84	76	85	70	73	LULUS	LULUS
141510010	70	75	50	79	64	78	30	58	TIDAK	TIDAK
141510014	70	70	40	71	65	68	34	51	TIDAK	TIDAK
141510015	35	58	37	23	55	70	60	43	TIDAK	TIDAK

Tabel 4 : *Dataset Uji*

NIS	MTK1	BINDO1	BING1	IPS1	MTK2	BINDO2	BING2	IPS2	Ket	HASIL UN
131410042	76	74	66	65	68	68	77	62	LULUS	?

Langkah-langkah yang dilakukan saat proses prediksi dan klasifikasi:

- Menentukan nilai K, sebagai parameter tetangga terdekat, pada ilustrasi ini menggunakan nilai K=3
- Transformasi data, merata-rata nilai ujian *Try Out* pertama dan kedua, serta mengubah atribut keterangan
- Menghitung jarak dengan *Manhattan Distance*
- Mengurutkan hasil perhitungan secara ASC dengan *Selection Sorting*
- Melakukan voting terhadap data untuk menentukan target kelasnya dengan *Majority Voting*.

Melakukan Transformasi data pada *dataset model* dan *dataset uji*.

Tabel 5 : Transformasi *Dataset Model*.

Mean1	Mean2	Ket
72,75	68,75	1
75,5	59,25	0
80,5	76	1
68,5	57,5	0
62,75	54,54	0
38,25	57	0

Tabel 6 : Transformasi *Dataset Uji*

Mean1	Mean2	Ket
70,25	68,75	1

Menghitung jarak dengan *Manhattan Distance*, dimana kasus baru dengan kasus lama $D_{ij} = \sum W_k |X_{ik} - C_{jk}|$.

Matrix baris pertama kolom pertama

$$|72,75-70,5| + |68,75-68,75| + |1-1| = 2,5$$

Matrix baris pertama kolom kedua

$$|75,5-70,5| + |59,25-68,75| + |1-1| = 15,75$$

Matrix baris pertama kolom ketiga

$$|80,5-70,5| + |76-68,75| + |1-1| = 17,5$$

Matrix baris pertama kolom keempat

$$|68,5-70,5| + |57,5-68,75| + |1-1| = 14,5$$

Matrix baris pertama kolom kelima

$$|62,5-70,5| + |54,5-68,75| + |1-1| = 22,75$$

Matrix baris pertama kolom ketujuh

$$|38,25-70,5| + |57-68,75| + |1-1| = 44,75$$

Mengurutkan hasil perhitungan tersebut secara *ascending* dengan metode *selection sorting*.

- 2,5 LULUS
- 14,5 TIDAK
- 15,75 TIDAK
- 17,5 LULUS
- 22,75 TIDAK
- 44,75 TIDAK

Memilih 3 data terdekat sesuai dengan nilai tetangga terdekat yaitu K=3 dengan *majority voting* untuk menentukan target kelas *dataset uji*.

- 2,5 LULUS
- 14,5 TIDAK
- 15,75 TIDAK

Jumlah target kelas LULUS = 1

Jumlah target kelas TIDAK = 2

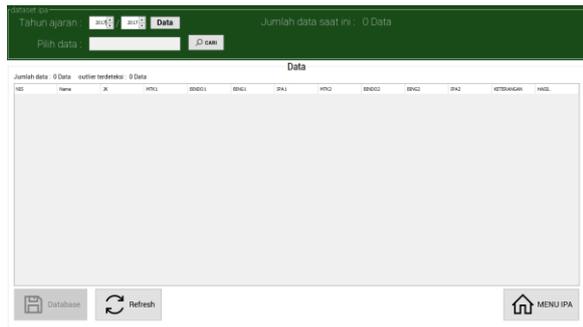
Dari hasil target kelas diatas dapat disimpulkan *dataset uji* tersebut memiliki hasil TIDAK.

3. HASIL DAN PEMBAHASAN

3.1. Tampilan Layar Menu *Dataset*

Tampilan dari menu *Dataset* digunakan untuk meng-*import* dataset kedalam basis data untuk digunakan sebagai dataset uji dan dataset model pada

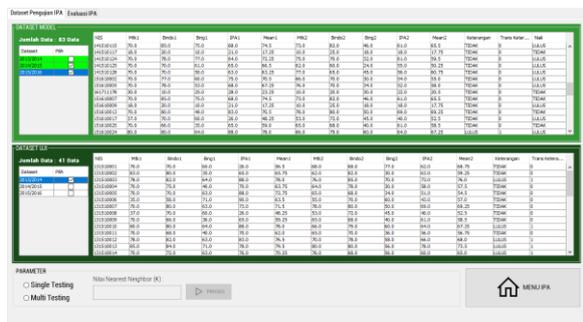
proses prediksi dan klasifikasi. Tampilan layar menu *dataset* sebagai berikut :



Gambar 2 : Tampilan Layar Menu Dataset

3.2. Tampilan Layar Menu Evaluasi

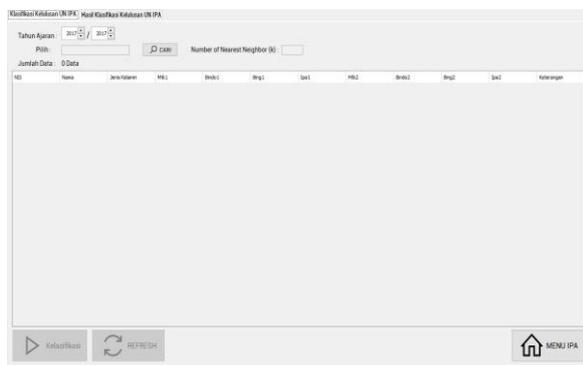
Tampilan menu Evaluasi berfungsi untuk memilih dataset yang mana saja untuk dilakukan proses pengujian. Tampilan Layar Menu Evaluasi sebagai berikut:



Gambar 3 : Tampilan Layar Menu Evaluasi

3.3. Tampilan Layar Menu Klasifikasi

Tampilan menu Klasifikasi berfungsi untuk melakukan proses prediksi dan klasifikasi pada dataset yang ingin diketahui hasil dari prediksi kelulusan ujian nasionalnya. Tampilan layar menu klasifikasi sebagai berikut:



Gambar 4 : Tampilan Layar Menu Klasifikasi

3.4. Hasil Uji Coba

Pada pengujian ini dilakukan 2 macam pengujian untuk mengetahui kinerja sistem berdasarkan nilai akurasi yang dihasilkan dalam memprediksi kelulusan ujian nasional (UN). Pengujian yang dilakukan ada 2 jenis, yaitu *multi*

testing dan *single testing*. Kedua pengujian ini dilakukan untuk mengetahui nilai K terbaik dalam melakukan proses klasifikasi. Pada proses *multi testing* menggunakan 2 parameter K=9 dan K=13. Sedangkan pada *single testing* menggunakan parameter K=9.

a. Pengujian *Multi Testing*

dataset uji adalah tahun ajaran 2013/2014, dan dataset model adalah tahun ajaran 2014/2015 dan 2015/2016.

Tabel 7: Tabel Pengujian *Multi Testing* Pertama IPA

K	Accuracy	Recall	Precision
9	68.29%	100%	68.29%
13	68.29%	100%	68.29%

Tabel 8: Tabel Pengujian *Multi Testing* Pertama IPS

K	Accuracy	Recall	Precision
9	74.72%	85.82%	82.19%
13	73.62%	85.71%	81.08%

dataset uji adalah tahun ajaran 2014/2015, dan dataset model adalah tahun ajaran 2013/2014 dan 2015/2016.

Tabel 9: Tabel Pengujian *Multi Testing* Kedua IPA

K	Accuracy	Recall	Precision
9	82.5%	97.05%	84.61%
13	82.5%	97.05%	84.61%

Tabel 10: Tabel Pengujian *Multi Testing* Kedua IPS

K	Accuracy	Recall	Precision
9	78.94%	100%	78.02%
13	76.84%	100%	76.34%

dataset uji adalah tahun ajaran 2015/2016, dan dataset model adalah tahun ajaran 2014/2015 dan 2013/2014.

Tabel 11: Tabel Pengujian *Multi Testing* Ketiga IPA

K	Accuracy	Recall	Precision
9	83.72%	97.29%	85.71%
13	86.04%	100%	86.04%

Tabel 12: Tabel Pengujian *Multi Testing* Ketiga IPS

K	Accuracy	Recall	Precision
9	69.04%	69.11%	90.38%
13	71.42%	77.94%	85.48%

Dari ketiga pengujian *multi testing* diatas, didapatkan parameter nilai K terbaik adalah K=13.

Tabel 13: Tabel Pengujian *Multi Testing* IPA

NO	Dataset		Nilai K	Hasil Evaluasi	
	Uji	Model		Akurasi	Rata-rata
1	2013/2014	2014/2015	9	68.29%	68.29%
		2015/2016	13	68.29%	
2	2014/2015	2013/2014	9	82.52%	82.52%
		2015/2016	13	82.52%	
3	2015/2016	2013/2014	9	83.72%	84.88%
		2014/2015	13	86.04%	

Dari ketiga pengujian *multi testing* diatas, didapatkan parameter nilai K terbaik adalah K=9.

Tabel 14: Tabel Pengujian *Multi Testing* IPS

NO	Dataset		Nilai K	Hasil Evaluasi	
	Uji	Model		Akurasi	Rata-rata
1	2013/2014	2014/2015	9	74.72%	74.71%
		2015/2016	13	73.62%	
2	2014/2015	2013/2014	9	78.94%	77.89%
		2015/2016	13	76.84%	
3	2015/2016	2013/2014	9	69.04%	70.23%
		2014/2015	13	71.42%	

b. Pengujian *Single Testing*

dataset uji adalah tahun ajaran 2013/2014, dan dataset model adalah tahun ajaran 2014/2015 dan 2015/2016.

CONFUSION MATRIX			
	Prediksi Lulus	prediksi Tidak	class recall
Actual LULUS	28	0	100 %
Actual TIDAK	13	0	0 %
Class Precision	68,29 %	NaN %	

*AKURASI
Overall Accuracy is 68,29 %

Gambar 5: Pengujian *Single Testing* Pertama IPA

CONFUSION MATRIX			
	Prediksi Lulus	prediksi Tidak	class recall
Actual LULUS	60	10	85,71 %
Actual TIDAK	13	8	38,1 %
Class Precision	82,19 %	44,44 %	

*AKURASI
Overall Accuracy is 74,73 %

Gambar 6: Pengujian *Single Testing* Pertama IPS

dataset uji adalah tahun ajaran 2014/2015, dan dataset model adalah tahun ajaran 2013/2014 dan 2015/2016

CONFUSION MATRIX			
	Prediksi Lulus	prediksi Tidak	class recall
Actual LULUS	33	1	97,06 %
Actual TIDAK	6	0	0 %
Class Precision	84,62 %	0 %	

*AKURASI
Overall Accuracy is 82,5 %

Gambar 7: Pengujian *Single Testing* Kedua IPA

CONFUSION MATRIX			
	Prediksi Lulus	prediksi Tidak	class recall
Actual LULUS	71	0	100 %
Actual TIDAK	20	4	16,67 %
Class Precision	78,02 %	100 %	

*AKURASI
Overall Accuracy is 78,95 %

Gambar 8: Pengujian *Single Testing* Kedua IPS

dataset uji adalah tahun ajaran 2015/2016, dan dataset model adalah tahun ajaran 2014/2015 dan 2013/2014

CONFUSION MATRIX			
	Prediksi Lulus	prediksi Tidak	class recall
Actual LULUS	30	7	81,08 %
Actual TIDAK	4	2	33,33 %
Class Precision	88,24 %	22,22 %	

*AKURASI
Overall Accuracy is 74,42 %

Gambar 9: Pengujian *Single Testing* Ketiga IPA

CONFUSION MATRIX			
	Prediksi Lulus	prediksi Tidak	class recall
Actual LULUS	47	21	69,12 %
Actual TIDAK	5	11	68,75 %
Class Precision	90,38 %	34,38 %	

*AKURASI
Overall Accuracy is 69,05 %

Gambar 10: Pengujian *Single Testing* Ketiga IPS

4. KESIMPULAN

4.1. Kesimpulan

Berdasarkan perancangan, pembuatan, analisa program dan serangkaian uji coba dari sistem ini, maka dapat diambil kesimpulan sebagai berikut:

- Algoritma *K-Nearest Neighbor* dengan metode pengukuran jarak *Manhattan Distance* dapat diimplementasikan kedalam proses prediksi peluang kelulusan ujian nasional
- Waktu proses tergantung jumlah data dan memori ram komputer
- Semakin besar jumlah data akan mempengaruhi tingkat akurasi

4.2. Saran

Berikut adalah saran yang penulis berikan untuk pengembangan sistem dimasa depan:

- Menggunakan algoritma atau metode lain untuk mengembangkan sistem untuk prediksi peluang kelulusan ujian nasional.
- Dalam mencari nilai K terbaik dapat dilakukan secara otomatis tanpa harus manual kembali.

- c. Proses dari mulai dataset, evaluasi, klasifikasi, dan backup untuk jurusan IPA dan IPS dapat disatukan.
- d. Mengganti bahasa pemrograman java desktop menjadi berbasis web.
- e. Melakukan prediksi untuk bidang lainnya.
- f. Dengan sistem prediksi diharapkan dunia pendidikan diindonesia menjadi lebih baik lagi.
- g. Dapat memprediksi per-mata pelajaran ujian nasional.

5. DAFTAR PUSTAKA

- [1] A. R. C *et al.*, "Implementasi Sistem Crowdsourced Labelling Berbasis Web dengan Metode Weighted Majority Voting," vol. VI, no. 2, pp. 76–82, 2015.
- [2] Bhatia, M., Vandana., 2010. Survey of nearest neighbor techniques. *International Journal of Computer Science and Information Security* 8, 1947-5500.
- [3] Hasugian, J. & Utara, U.S., 2006. Penggunaan Bahasa Alamiah dan Kosa Kata Terkendali dalam Sistem Temu Balik Informasi Berbasis Teks Jonner Hasugian Departemen Studi Perpustakaan dan Informasi. , 2(2).
- [4] Kunang, Y.N. et al., 2013. IMPLEMENTASI TEKNIK DATA MINING UNTUK MEMPREDIKSI TINGKAT KELULUSAN MAHASISWA PADA UNIVERSITAS BINA. , 2013(semnasIF), pp.56–63.
- [5] Latifah, K., Kombinasi Algorithma K-NN dan Manhattan Distance untuk Menentukan Pemenang Lelang. , pp.49–58.
- [6] Leidiyana, H., 2013. Penerapan algoritma k-nearest neighbor untuk penentuan resiko kredit kepemilikan bermotor., 1(1), pp.65-76.
- [7] Moradian, M., Baraani, A., 2009. K-Nearest Neighbor Based Association Algorithm. *Journal of Theoretical and Applied Information Technology* 6, 123 – 129.
- [8] Nabil El-Sawalhi , David Eaton , Rifat Rustom ; Contractor pre-qualification model: State-of-the-art; *International Journal of Project Management* 25 (2007) 465–474., Science Direct 2007.
- [9] Rahayuningsih, P.A., 2016. Analisa Perbandingan Kompleksitas Algoritma Pengurutan Nilai (Sorting). , 4.
- [10] Sumarlin,(2015). Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa dan BBM
- [11] Syaliman, K.U., 2017. Pembentukan Prototype Data Dengan Metode K-Means Untuk Klasifikasi dalam Metode K- Nearest Neighbor (K-NN). , pp.185–190.
- [12] Siswa, P. et al., 2015. IMPLEMENTASI METODE SELECTION SORT UNTUK MENENTUKAN NILAI. , 11(1).