

IMPLEMENTASI ALGORITMA *K-NEAREST NEIGHBOR* DENGAN METODE KLASIFIKASI DAN PENGUKURAN JARAK *MANHATTAN DISTANCE* UNTUK PREDIKSI KELULUSAN UN BERDASARKAN HASIL NILAI *TRYOUT* BERBASIS *JAVA DESKTOP* PADA SMA HARAPAN JAYA 2

Erlangga Dwi Kurniawan, Mufti

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur

Jl. Raya Ciledug, Petukangan Utara, Kebayoran Lama, Jakarta Selatan 12260

Telp. (021) 5853753, Fax. (021) 5866369

E-mail : erlanggaadk12@gmail.com , muftyhayat@gmail.com

Abstrak

SMA Harapan Jaya 2 setiap tahunnya pihak sekolah selalu mengadakan tryout sebanyak dua (2) kali untuk melakukan persiapan menjelang ujian nasional sebagai langkah untuk para siswa dalam mempersiapkan kematangan materi menjelang ujian nasional. Nantinya pihak sekolah akan memprediksi siswa untuk kelulusan ujian nasional dari hasil tryout tersebut. Siswa yang mendapatkan nilai tryout dibawah standar maka diprediksi tidak lulus ujian nasional. Dalam memprediksi kelulusan UN pihak sekolah masih melakukannya dengan cara manual yaitu dengan menggunakan Microsoft Excel, yaitu dengan mengumpulkan semua nilai hasil tryout seluruh siswa dan menjumlah seluruh nilai. Berdasarkan hal tersebut prediksi kelulusan UN dengan cara manual dianggap kurang efisien dan efektif. Penulis menggunakan algoritma K-Nearest Neighbor dengan metode pengukuran jarak Manhattan Distance untuk mengklasifikasi siswa menjadi dua hasil kelulusan yaitu Lulus dan Tidak Lulus dengan atribut yang mempengaruhi antara lain, nilai hasil tryout seluruh mata pelajaran tiap siswa. Algoritma K-Nearest Neighbor merupakan metode klasifikasi yang menghitung jarak terdekat untuk mengklasifikasikan objek baru berdasarkan kedekatannya dengan objek lama yang sudah di klasifikasikan sebelumnya dengan menggunakan dataset siswa selama 3 tahun terakhir. Maka disimpulkan sistem dapat mengklasifikasi dataset siswa SMA Harapan Jaya 2 dengan baik dan cepat dibandingkan dengan proses prediksi kelulusan UN yang dilakukan secara manual menggunakan microsoft excel.

Kata Kunci : *K-Nearest Neighbor, Manhattan Distance, Klasifikasi, Prediksi, Tryout*

1. PENDAHULUAN

Perkembangan teknologi saat ini semakin pesat, baik di negara maju maupun di negara berkembang. Perkembangan ini memudahkan pengguna komputer dalam melakukan aktifitas bisnis, pendidikan dan aktifitas lainnya. Sehingga dapat meningkatkan produktivitas dan meringankan pekerjaan manusia, ini dapat memungkinkan beberapa pekerjaan manusia bisa digantikan dengan komputer yang disesuaikan dengan kebutuhan manusia.

SMA Harapan Jaya 2 adalah sekolah menengah atas yang berlokasi di Provinsi Banten dan berakreditasi B. Untuk setiap tahunnya SMA Harapan Jaya 2 mengadakan *TryOut* sebanyak 3 kali untuk siswa/siswi yang akan menghadapi Ujian Nasional (UN). Dari hasil *TryOut* tersebut pihak sekolah akan mem-prediksi siswa/siswi lulus atau tidak dalam Ujian Nasional berdasarkan hasil ujian *TryOut* yang telah dilaksanakan. Bagi siswa/siswi yang tidak lulus akan diberikan materi tambahan agar lebih matang lagi dalam menghadapi Ujian Nasional (UN).

Dalam penerapannya pihak sekolah masih kesulitan untuk menentukan siswa/siswi yang diprediksi akan lulus dalam menghadapi Ujian Nasional karena masih melakukan penghitungan manual dengan cara mengumpulkan nilai-nilai hasil *TryOut* yang telah dilaksanakan oleh siswa/siswi tersebut. Sehingga proses tersebut membutuhkan waktu sekitar 2-4 hari. Berdasarkan hal tersebut, penerapan sistem prediksi ini akan membantu meningkatkan efektivitas dan efisiensi dalam menentukan kelulusan yang diharapkan.

Berdasarkan uraian diatas, penulis bermaksud untuk membangun suatu sistem yang dapat memprediksi kelulusan siswa/siswi SMA Harapan Jaya 2 dengan mengimplementasikan bidang ilmu klasifikasi dan evaluasi menggunakan algoritma *K-Nearest Neighbor* dan *Manhattan Distance* sebagai pencarian terdekatnya.

Algoritma *K-Nearest Neighbor* pada dasarnya akan mengevaluasi kedekatan antara data satu dengan data lainnya menggunakan metode pengukuran jarak, yang kemudian kedekatan inilah yang mengidentifikasi kelas atau label data tersebut [7].

Algoritma *K-Nearest Neighbor* dengan metode pengukuran jarak *Manhattan Distance* ini diharapkan dapat menghasilkan tingkat akurasi yang baik dalam implementasinya untuk prediksi kelulusan Ujian Nasional siswa/siswi SMA Harapan Jaya 2.

Masalah yang muncul adalah banyaknya data siswa yang ada setiap tahunnya selalu bertambah dan itu membuat proses prediksi tidak efisien sehingga pihak sekolah membutuhkan sistem yang dapat mengefisiensikan waktu dalam proses prediksi kelulusan.

Tujuan yang ingin dicapai pada penelitian ini untuk mengetahui unjuk kerja pada algoritma *K-Nearest Neighbor* dan pengukur jarak *Manhattan Distance* yang di implementasikan pada sistem prediksi kelulusan ujian nasional yang diharapkan dapat meningkatkan waktu pada proses prediksi kelulusan.

2. METODE PENELITIAN

2.1. Studi Literatur

Metode ini dilakukan dengan mempelajari dan membaca dari berbagai referensi yang telah dikumpulkan baik dalam bentuk buku, *e-book*, jurnal, artikel dan lain-lain. Tujuan untuk memperoleh informasi yang akan digunakan dalam pengembangan sistem dan segala hal yang berkaitan dengan algoritma *K-Nearest Neighbor*.

2.2. Observasi

Kegiatan observasi dilakukan untuk mendapatkan data siswa/siswi SMA Harapan Jaya 2 dari bapak Surya Subhan, S.Pd. selaku kepala sekolah serta melakukan wawancara dengan beliau untuk mendapatkan informasi tentang parameter apa saja yang digunakan sebagai prediksi kelulusan Ujian Nasional di SMA Harapan Jaya 2.

2.3. Sumber Data

Pengumpulan data dalam penelitian ini dilakukan dengan mengumpulkan data primer dan data sekunder. Data primer untuk kasus ini diperoleh dari nilai hasil Ujian *TryOut* siswa/siswi SMA Harapan Jaya. *Dataset* yang digunakan adalah data siswa/siswi yang sudah melaksanakan ujian *TryOut* dan Ujian Nasional selama tiga tahun terakhir, berikut detail dari dataset :

Tabel 1. Data Penelitian

No	Tahun Ajaran	IPA	IPS
1	2014/2015	71	70
2	2015/2016	63	65
3	2016/2017	67	75

Sedangkan data sekunder diperoleh dari buku, *e-book*, laporan penelitian, jurnal ilmiah dan lain-lain yang digunakan sebagai referensi untuk mendapatkan kajian teoritis yang berkaitan dengan algoritma *K-Nearest Neighbor*.

2.4. Pembersihan Data

Proses ini melakukan pembersihan data yang dianggap sebagai *outlier* atau data yang mengandung *noise*. Berikut adalah data yang dianggap sebagai *outlier* :

- Data nilai *TryOut* yang bernilai 0 atau null.
- Duplikasi Data.

2.5. Pemilihan Data

Pada proses ini bertujuan untuk menentukan *attribute* yang tepat untuk proses analisis data. Penentuan *attribute* ini berdasarkan hasil diskusi dengan guru bidang kurikulum SMA Harapan Jaya 2. Berikut *attribute* IPS dan IPS yang didapat dari hasil diskusi :

- Nilai *TryOut* Pertama Matematika
- Nilai *TryOut* Pertama Bahasa Indonesia
- Nilai *TryOut* Pertama Bahasa Inggris
- Nilai *TryOut* Pertama IPA/IPS (Salah Satu Mata Pelajaran IPA/IPS)
- Nilai *TryOut* Kedua Matematika
- Nilai *TryOut* Kedua Bahasa Indonesia
- Nilai *TryOut* Kedua Bahasa Inggris
- Nilai *TryOut* Kedua IPA/IPS (Salah Satu Mata Pelajaran IPA/IPS)
- Keterangan Hasil *TryOut*
- Keterangan Hasil UN (Ujian Nasional)

2.6. Transformasi Data

Proses ini melakukan perubahan data ke bentuk yang sesuai dengan algoritma klasifikasi yang digunakan. Sebelumnya diketahui bahwa algoritma *K-Nearest Neighbor* tidak bisa dilakukan pada *attribute* yang bertipe kategorikal, maka perlu dilakukan transformasi data pada *attribute* hasil *TryOut* menjadi tipe numerikal seperti pada tabel berikut :

Tabel 2. Transformasi Data

Hasil <i>TryOut</i> Kategorikal	Transformasi Numerikal
LULUS	1
TIDAK	0

Sedangkan nilai hasil ujian *TryOut* IPA maupun IPS yang pertama dan kedua juga di transformasi menjadi nilai yang sesuai dengan kebijakan yang ada. Yaitu dengan menentukan nilai rata-rata dari hasil tersebut. Berikut transformasi nilai hasil ujian *tryout* IPA dan IPS :

$$(MTK1+BINDO1+BING1+IPS1 \text{ atau } IPA1)/4 = \text{MeanTO1}$$

$$(MTK2+BINDO2+BING2+IPS2 \text{ atau } IPA2)/4 = \text{MeanTO2}$$

2.7. Algoritma *K-Nearest Neighbor*

Algoritma *K-Nearest Neighbor* adalah salah satu algoritma klasifikasi yang cukup populer. Algoritma ini juga termasuk kedalam algoritma *instance-based classifier* dimana proses klasifikasi baru hanya akan dilakukan saat 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. [10] menyatakan bahwa *instance-base classifier* sering juga disebut sebagai metode pembelajaran *lazy learner* sehingga sangat berbeda 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 pelatihnnya. Algoritma *K-Nearest Neighbor* juga merupakan algoritma pembelajaran terawasi (*supervised learning*) dimana algoritma ini bertujuan untuk menentukan suatu kelas *query instance* yang baru terhadap data-data yang sudah ditentukan atau diketahui kelasnya. Berikut ini Alur kerja dari algoritma *K-Nearest Neighbor* menurut :

- Tentukan nilai K, sebagai parameter ketetanggan terdekat
- Hitung jarak antara *query instance* dan semua data pelatihan
- Urutkan berdasarkan jarak terkecil
- Tentukan label kelas *query instance* berdasarkan label mayoritas pada K objek data latih terdekat

2.8. Manhattan Distance

Manhattan atau *City Distance* digunakan untuk mengambil kasus yang cocok dari basis kasus dengan menghitung jumlah bobot absolute dari perbedaan antara kasus yang sekarang dan kasus yang lain dalam basis kasus. Untuk menghitung bobot digunakan persamaan sebagai berikut :

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

Diketahui d_{ij} adalah jarak antara kasus antara i th dan j th dengan semua parameternya. W merepresentasikan jumlah dari bobot. X adalah kasus yang baru dikurangi dengan C yaitu *history*. *Manhattan Distance* adalah pengukuran *similarity* / kemiripan yang paling cocok untuk *approval project* yang merepresentasikan kasus yang relevan dengan angka yang natural atau dengan data yang bersifat kuantitatif, dengan menganalisa data, panitia lelang akan memiliki acuan untuk pengambilan keputusan persetujuan terhadap pemenang dan membuat prediksi hasil keputusan menjadi lebih akurat [5].

2.9. Confusion Matrix

Menurut [2], *Confusion Matrix* memberikan nilai kinerja model klasifikasi berdasarkan perhitungan objek yang diklasifikasikan dengan benar dan objek yang diklasifikasikan dengan salah. Tabel dibawah ini menggambarkan hasil perhitungan yang ditabulasikan kedalam sebuah tabel.

Tabel 3. *Confusion Matrix*

		Predicted class		TOTAL
		YES	NO	
Actual Class	YES	TP	FN	P
	NO	FP	TN	N
TOTAL		P'	N'	P + N

Keterangan :

- P (*Positive*) : Jumlah kelas positif (*YES*)
- N (*Negative*) : Jumlah kelas negatif (*NO*)
- TP (*True Positive*) : Jumlah tupel positif yang terklasifikasi dengan benar
- TN (*True Negative*) : Jumlah tupel negatif yang terklasifikasi dengan benar
- FP (*False Positive*) : Jumlah tupel positif yang terklasifikasi dengan tidak benar.
- FN (*False Negative*) : Jumlah tupel negatif yang terklasifikasi dengan tidak benar.

Dari *Matrix* tersebut dapat diperoleh beberapa pengukuran evaluasi diantaranya *accuracy*, *recall* (*sensitivity*) dan *precision* untuk menghitung nilai *accuracy* digunakan persamaan (2.1), nilai *recall* menggunakan persamaan (2.2) dan nilai *precision* menggunakan persamaan (2.3).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{2.1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2.2}$$

$$Precision = \frac{TP}{TP + FP} \tag{2.3}$$

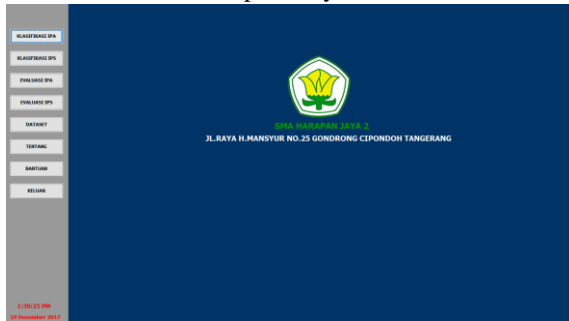
Accuracy adalah persentase dari *dataset* tes yang diklasifikasikan dengan benar oleh *classifier*, sedangkan menurut [4] *recall* berhubungan dengan rasio kemampuan *classifier* dalam memperoleh kelas yang relevan atau benar dan *precision* adalah ukuran tingkat ketepatan *classifier* untuk tidak mengklasifikasikan kelas kedalam kelas yang tidak tepat atau salah.

3. HASIL DAN PEMBAHASAN

3.1. Tampilan Layar Menu Utama

Menu Utama adalah tampilan layar yang pertama kali ditampilkan saat *user* membuka aplikasi ini. Pada tampilan layar menu utama ini ada beberapa *button* yang dapat dipilih oleh *user* yaitu *button* klasifikasi IPA untuk memanggil *form* klasifikasi IPA jika *user* ingin mengklasifikasi data kelulusan ujian nasional siswa IPA berdasarkan nilai ujian *tryout*, *button* klasifikasi IPS jika *user* ingin mengklasifikasi data kelulusan ujian nasional siswa IPS, *button* evaluasi IPA untuk memanggil *form* evaluasi IPA jika *user* ingin melakukan pengujian dan mencari parameter (k) terbaik pada *dataset* uji siswa IPA, *button* evaluasi IPS untuk melakukan pengujian dan mencari parameter (k) terbaik pada *dataset* uji siswa IPS,

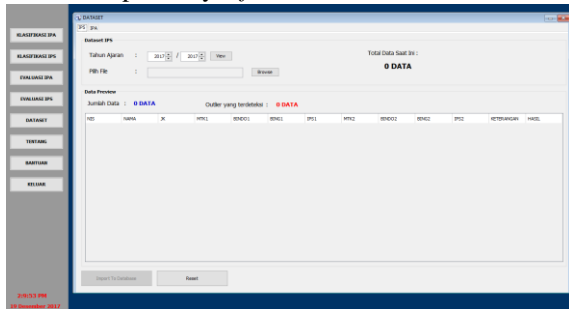
button dataset untuk memanggil form dataset jika user ingin meng-import data yang akan dijadikan dataset model dan dataset uji ke dalam database, button tentang untuk memanggil form tentang jika user ingin melihat informasi dari penulis, button bantuan untuk memanggil form bantuan jika user ingin melihat cara menggunakan aplikasi ini dan button keluar jika user ingin keluar dari aplikasi ini. Berikut ini adalah tampilan layar menu utama :



Gambar 1. Tampilan Menu Utama

3.2. Tampilan Form Dataset

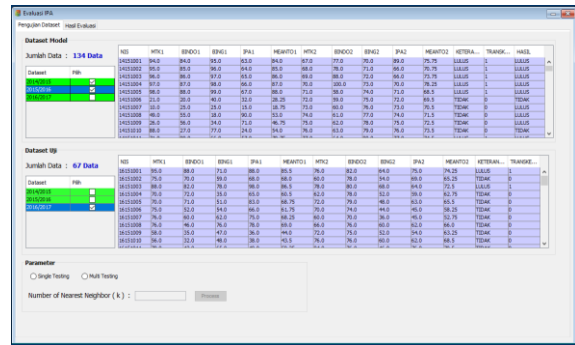
Tampilan Layar Form Dataset akan muncul jika user memilih button dataset pada menu utama. Fungsi form dataset untuk meng-import dataset siswa IPA atau siswa IPS ke dalam database yang nantinya digunakan sebagai dataset model dan dataset uji pada proses prediksi dan klasifikasi kelulusan ujian nasional berdasarkan nilai ujian tryout. Berikut ini adalah tampilan layar form dataset :



Gambar 2. Tampilan Form Dataset

3.3. Tampilan Form Evaluasi

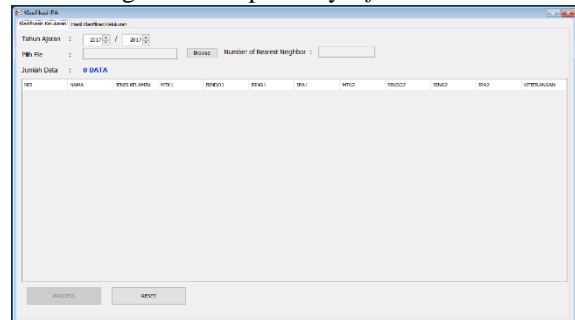
Tampilan Layar Form Evaluasi ini akan tampil saat user mengklik button evaluasi pada menu utama. Pada form ini terdapat dua tab yaitu tab Pengujian Dataset dan Hasil Evaluasi. Tab Pengujian Dataset berfungsi untuk memilih dataset yang telah dikelompokkan yaitu dataset model dan dataset uji yang akan diproses pengujian untuk mencari nilai K terbaik dalam sistem ini. Berikut ini adalah tampilan layar form evaluasi tab pengujian dataset :



Gambar 3. Tampilan Form Evaluasi

3.4. Tampilan Form Klasifikasi

Tampilan Layar Form Klasifikasi adalah tampilan layar yang akan tampil ketika user mengklik button klasifikasi pada menu utama. Form ini berfungsi untuk melakukan prediksi dan klasifikasi pada dataset yang ingin diketahui hasil prediksi kelulusan ujian nasional siswa berdasarkan nilai tryout. Berikut ini adalah gambar tampilan layar form klasifikasi :



Gambar 4. Tampilan Form Klasifikasi

3.5. Hasil Uji Coba

Pada bagian ini akan memaparkan hasil dari proses pengujian untuk mengetahui unjuk kerja sistem berdasarkan nilai akurasi yang dihasilkan dan pengujian untuk mengetahui waktu proses dalam memprediksi dan klasifikasi kelulusan ujian nasional suatu dataset siswa IPA dan IPS.

Pengujian unjuk kerja sistem ini memiliki dua jenis pengujian yaitu pengujian single testing dan pengujian multi testing. Kedua pengujian ini dilakukan untuk menentukan nilai K (Number of Nearest Neighbor) terbaik untuk melakukan proses prediksi dan klasifikasi kelulusan.

a. Pengujian Multi Testing IPA

Pada pengujian ini tiap dataset uji akan diuji dengan seluruh dataset model dan menggunakan nilai K yang sama agar mendapatkan nilai K terbaik dari masing-masing dataset. Berikut hasil uji multi testing IPA :

Tabel 4. Hasil *Multi Testing* IPA

No.	Dataset		Nilai K	Hasil Evaluasi	
	Uji	Model		Akurasi	Rata – Rata Akurasi
1	2016/2017	2014/2015 2015/2016	3	85.07%	85.81%
			5	86.56%	
			7	86.56%	
			9	85.07%	
2	2015/2016	2014/2015 2016/2017	3	88.88%	88.09%
			5	88.88%	
			7	87.30%	
			9	87.30%	
3	2014/2015	2015/2016 2016/2017	3	88.73%	89.08%
			5	90.14%	
			7	88.73%	
			9	88.73%	

Dapat disimpulkan pada *multi testing* IPA ini nilai K terbaik yaitu 5.

b. Pengujian *Multi Testing* IPS

Pada pengujian ini tiap *dataset* uji akan diuji dengan seluruh *dataset* model dan menggunakan nilai K yang sama agar mendapatkan nilai K terbaik dari masing-masing dataset. Berikut hasil uji *multi testing* IPS :

Tabel 5. Hasil *Multi Testing* IPS

No.	Dataset		Nilai K	Hasil Evaluasi	
	Uji	Model		Akurasi	Rata – Rata Akurasi
1	2016/2017	2014/2015 2015/2016	3	85.33%	82.33%
			5	84.0%	
			7	81.33%	
			9	78.66%	
2	2015/2016	2014/2015 2016/2017	3	98.46%	97.30%
			5	96.92%	
			7	96.92%	
			9	96.92%	
3	2014/2015	2015/2016 2016/2017	3	92.85%	92.85%
			5	94.28%	
			7	91.42%	
			9	92.85%	

Dapat disimpulkan pada *multi testing* IPS ini nilai K terbaik yaitu 3.

c. Pengujian *Single Testing* IPA

Pada pengujian *single testing* IPA dilakukan tiga pengujian, pada pengujian yang pertama ini menggunakan *dataset* uji dengan tahun ajaran 2016/2017 dan untuk *dataset* modelnya menggunakan tahun ajaran 2014/2015 dan 2015/2016.

Gambar 5. Pengujian *Single Testing* IPA Pertama

Pengujian kedua ini menggunakan *dataset* uji dengan tahun ajaran 2015/2016 dan untuk *dataset* modelnya menggunakan tahun ajaran 2014/2015 dan 2016/2017.

Gambar 6. Pengujian *Single Testing* IPA Kedua

Pengujian ketiga ini menggunakan *dataset* uji dengan tahun ajaran 2014/2015 dan untuk *dataset* modelnya menggunakan tahun ajaran 2015/2016 dan 2016/2017.

Gambar 7. Pengujian *Single Testing* IPA Ketiga

d. Pengujian *Single Testing* IPS

Pada pengujian *single testing* IPS juga dilakukan tiga pengujian, pada pengujian yang pertama ini menggunakan *dataset* uji dengan tahun ajaran 2016/2017 dan untuk *dataset* modelnya menggunakan tahun ajaran 2014/2015 dan 2015/2016.

Gambar 8. Pengujian *Single Testing* IPS Pertama

Pengujian kedua ini menggunakan *dataset* uji dengan tahun ajaran 2015/2016 dan untuk *dataset* modelnya menggunakan tahun ajaran 2014/2015 dan 2016/2017.

Gambar 9. Pengujian *Single Testing* IPS Kedua

Pengujian ketiga ini menggunakan *dataset* uji dengan tahun ajaran 2014/2015 dan untuk *dataset* modelnya menggunakan tahun ajaran 2015/2016 dan 2016/2017.

	Predicted LULUS	Predicted TIDAK	Class Recall
Actual LULUS	58	3	95,08 %
Actual TIDAK	2	7	77,78 %
Class Precision	96,67 %	70 %	

* ACCURACY
Overall Accuracy is 92,86 %

Gambar 10. Pengujian *Single Testing* IPS Ketiga

4. KESIMPULAN

4.1. Kesimpulan

- Menerapkan algoritma *K-Nearest Neighbor* dengan metode evaluasi dan klasifikasi dengan pengukuran jarak *Manhattan Distance* dapat diimplementasi ke dalam proses prediksi untuk kelulusan ujian nasional (UN).
- Perbedaan tingkat akurasi yang dihasilkan antar parameter nilai K terlalu signifikan dan masih dalam kategori akurasi yang tinggi.
- Nilai akurasi pada pengujian menggunakan algoritma *K-Nearest Neighbor* dengan pengukur jarak *Manhattan Distance* berdasarkan jumlah uji sebanyak 3 (tiga) kali dengan total 201 data jurusan IPA dan 210 data jurusan IPS selama 3 (tiga) angkatan didapat nilai $K = 5$ untuk jurusan IPA dan nilai $K = 3$ untuk jurusan IPS dengan nilai rata-rata akurasi terbesar 89,08% pada data tahun 2014/2015 untuk jurusan IPA dan nilai rata-rata akurasi terbesar 97,30% pada data tahun 2015/2016 untuk jurusan IPS.

4.2. Saran

- Menggunakan metode evaluasi *K-Fold Cross Validation* dalam proses evaluasi model untuk mengetahui unjuk kerja sistem klasifikasi dengan otomatis.
- Menggunakan nilai K yang lebih variatif untuk mengetahui perbedaan nilai akurasi, presisi dan recallnya.
- Sebaiknya program ini dapat dikembangkan dan diaplikasikan serta diterapkan di dunia pendidikan pada umumnya, karena prediksi dan klasifikasi ujian nasional siswa SMA suatu hal yang penting untuk melakukan kontrol nilai pada siswa dan mengendalikan progress dari nilai siswa.

5. DAFTAR PUSTAKA

- [1] Badrul, M., Studi, P. and Informasi, S. (2015) 'PREDIKSI HASIL PEMILU LEGISLATIF DENGAN MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR', XI(2), pp. 152–160.
- [2] Gorunescu, F. (2011). Data Mining-Concepts, Models and Technologies. Journal of Chemical Information and Modeling (Vol. 12). Craiova: Springer. <http://doi.org/10.1017/CBO9781107415324.004>.
- [3] Han, J., Kamber, M., & Pei, J. (2012). Data Mining:

Concepts and Techniques. Journal of Chemical Information and Modeling (3rd ed., Vol. 3). Waltham: Morgan Kaufmann. <http://doi.org/10.1017/CBO9781107415324.004>.

- [4] Hasugian, J. (2006) 'Penggunaan Bahasa Alamiah dan Kosakata Terkendali dalam Sistem Temu Balik Informasi Berbasis Teks Informasi', Pustaka: Jurnal Studi Perpustakaan dan Informasi, 2(2), pp. 73–80. Available at: <http://repository.usu.ac.id/bitstream/123456789/17059/1/pus-des2006-1.pdf>.
- [5] Latifah, K. (2015) 'Kombinasi Algoritma 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 Kendaraan Bermotor. Jurnal Penelitian Ilmu Komputer, System Embedded & Logic, 1(1), 65–76.
- [7] Nursalim, Suprapedi, & H.Himawan. (2014). Klasifikasi Bidang Kerja Lulusan Menggunakan Algoritma K-Nearest Neighbor. Jurnal Teknologi Informasi, 10(April), 31–43.
- [8] Rahmawati, T. D., Adnan, F. N. and Cs, M. (1978) 'Penentuan Produk Asuransi Bpjs Berdasarkan Profil Pelanggan Dengan Pendekatan K-Nearest', (x), pp. 1–15.
- [9] Rich, E., & Knight, K. (1991). Artificial Intelligence (Third Edition) (3rd ed.). New York: Tata McGraw-Hill.
- [10] Sillueta, C. Y. (2016). Implementasi Data Mining Untuk Memprediksi Kelulusan Mahasiswa Dengan Metode Klasifikasi Dan Algoritma K-Nearest Neighbor Berbasis Desktop (Studi Kasus: Fakultas Teknologi Informasi, Program Studi Teknik Informatika-Universitas Budi Luhur).
- [11] Siswanto. (2000). Kecerdasan Tiruan (2nd ed., pp. 1–134). Graha Ilmu.
- [12] Sumarlin (2015) 'Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM', Jurnal Sistem Informasi Bisnis, 1, pp. 52–62.
- [13] Wu, X., Kumar, V., Ross, Q. J., Ghosh, J., Yang, Q., Motoda, H., ... Steinberg, D. (2008).
- [14] Top 10 Algorithms in Data Mining. Knowledge and Information Systems (Vol. 14). <http://doi.org/10.1007/s10115-007-0114-2>.