



Penerapan K-Nearest Neighbour Dalam Penerimaan Peserta Didik Dengan Sistem Zonasi

Denni Kurniawan, Ade Saputra *

Magister Ilmu Komputer Universitas Budi Luhur

Naskah Diterima : 22 Oktober 2019; Diterima Publikasi : 8 November 2019

DOI : 10.21456/vol9iss2pp212-219

Abstract

Admission of new students is a routine activity at the beginning of each new meeting year in all formal educational institutions. At the moment the acceptance of new students uses the zoning system and has been regulated by Permendikbud No. 20 in 2019. This zoning system will accept students where their residence enters through the user area with the school environment. With this Permendikbud the government hopes that there will be an evenness in the quality of education in all schools, so that schools will no longer get the title of superior and non-superior schools. But in a system, the zoning improves anxieties in the school environment. This research supports to help the participating school students will be accepted in accordance with the provisions of the Ministry of Education and Culture. In overcoming problems that arise in the school environment there needs to be a system that can overcome that problem. In this study using the K-Nearest Neighbor (K-NN) method. Where the K-NN method will do the classification of new learners' residence with the school. In determining the classification using the K-NN method used for zoning and non-zoning areas, it is seen based on the closest K value. In finding the optimal value in this study using the Rapidminer application. The optimal high-level test results at K 5 where the value of this K is 83.36%.

Keywords : Data Mining; K-NN; Rapidminer; Zonasi

Abstrak

Penerimaan peserta didik baru merupakan kegiatan rutinitas di setiap awal tahun ajaran baru di semua Lembaga Pendidikan formal. Saat ini penerimaan Peserta didik baru menggunakan sistem zonasi dan sudah diatur oleh Permendikbud No. 20 tahun 2019. Sistem zonasi ini akan menerima peserta didik dimana tempat tinggalnya masuk kedalam area terdekat dengan lingkungan sekolah. Dengan Permendikbud tersebut pemerintah mengharapkan adanya kerataan mutu pendidikan disemua sekolah, sehingga sekolah tidak lagi mendapatkan predikat sekolah unggulan dan sekolah non unggulan. Tetapi pada kenyataannya sistem zonasi tersebut mengalami suatu keresahan di lingkungan sekolah. Penelitian ini bertujuan untuk membantu pihak sekolah dalam menetapkan peserta didik yang akan diterima sesuai dengan ketentuan dari Permendikbud. Dalam mengatasi permasalahan yang timbul di lingkungan sekolah perlu adanya suatu sistem yang dapat menangani permasalahan itu. Pada penelitian ini menggunakan metode K-Nearest Neighbour (K-NN). Dimana metode K-NN ini akan melakukan pengklasifikasian terhadap tempat tinggal peserta didik baru dengan sekolah. Dalam menentukan klasifikasi dengan metode K-NN yang digunakan untuk wilayah zonasi dan non zonasi dilihat berdasarkan nilai K terdekat. Dalam mencari nilai akurasi optimal pada penelitian ini menggunakan aplikasi Rapidminer. Hasil pengujian tingkat akurasi yang optimal terdapat pada K 5 dimana Nilai dari K tersebut bernilai 83,36 %.

Kata kunci : Data Mining; K-NN; Rapidminer; Zonasi

1. Pendahuluan

Sekolah merupakan salah satu lembaga pendidikan yang sangat penting untuk membangkitkan potensi yang ada dalam diri manusia itu sendiri. Selain itu, sekolah dapat mempersiapkan manusia dalam menghadapi masa depan agar hidup mereka dapat lebih sejahtera, baik secara individu maupun dalam bermasyarakat, berbangsa dan bernegara (Umaedi *et al.*, 2016). Untuk membangkitkan potensi dari peserta didik akan memilih sekolah yang sesuai dengan bakat dan

minatny. Penerimaan peserta didik baru merupakan suatu agenda rutinitas yang dilakukan oleh lembaga pendidikan nasional disemua jenjang. Untuk memilih sekolah yang menjadi minat dari peserta didik baru adalah mereka akan mencari sekolah yang dianggap sekolah favorit. Sehingga sekolah yang tidak memiliki predikat tidak favorit akan mengalami kesulitan dalam mendapatkan peserta didik. Hal inilah yang membuat Menteri Pendidikan dan Kebudayaan Muhajir Effendy membuat suatu peraturan Penerimaan Peserta Didik Baru (PPDB) dengan sistem zonasi dengan Permendikbud No. 51

*) Penulis korespondensi: 234tkj@gmail.com

Tahun 2018 tentang penerimaan peserta didik baru pada Taman Kanak-Kanak, Sekolah Dasar, Sekolah Menengah Pertama, Sekolah Menengah Atas dan Sekolah Menengah Kejuruan. Pada Permendikbud No. 51 Pasal 20 ayat (1) telah dijelaskan. Penetapan zonasi dilakukan pada setiap jenjang oleh Pemerintah Daerah sesuai dengan kewenangannya, dengan prinsip mendekatkan domisili peserta didik dengan Sekolah (Mendikbud, 2018). Tujuan dari Permendikbud No. 51 tersebut merupakan strategi pemerintah dalam menyederakan mutu dalam pendidikan. Sehingga tidak ada lagi kesan antara sekolah favorit dengan sekolah tidak favorit. Tetapi pada kenyataannya sistem zonasi ini menimbulkan banyak permasalahan mengingat jarak antara sekolah satu dengan sekolah yang lainnya berjauhan, sehingga para peserta didik baru akan mengalami kesulitan mengingat jarak antara sekolah dengan tempat tinggalnya berjauhan dan jumlah kuota siswa baru yang telah disediakan pihak sekolah melebihi dari jumlah peserta didik yang mendaftar.

Penelitian yang pernah dilakukan (Kartika *et al.*, 2017) yang berjudul Penentuan Siswa Berprestasi Menggunakan Metode *K-Nearest Neighbor* dan *Weighted Product* (Studi Kasus SMP Negeri 3 Mejayan), Permasalahan pada penelitian ini belum seimbang dalam melakukan penilaian dikarenakan nilai akademik masih menjadi pertimbangan yang utama dibandingkan nilai *non-akademik*. Nilai *non akademik* hanya dijadikan data penunjang yang tidak jelas dalam pembobotan nilai. Sehingga dirasa kurang adil dalam penentuan siswa berprestasi. Pengujian perankingan pertama memiliki tingkat akurasi 66.67%. Pengujian perankingan kedua memiliki tingkat akurasi sebesar 88.89%. Pengujian perankingan ketiga memiliki tingkat akurasi sebesar 100%. Dengan nilai rapor, ekstrakurikuler, kepribadian, dan kedisiplinan menjadi prioritas utama dalam penentuan siswa berprestasi. Pada penelitian lainnya yang dilakukan (Sumarlin, 2015) dengan judul Implementasi Algoritma *K-Nearest Neighbor* Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM. Penelitian ini membahas tentang klasifikasi beasiswa peningkatan prestasi akademik (PPA) dan bantuan belajar mahasiswa (BBM) berdasarkan variabel-variabel yang telah ditentukan dengan menerapkan algoritma *K-Nearest Neighbor*. Hasil klasifikasi dari sistem ini akan digunakan sebagai keputusan dalam pemberian beasiswa bagi mahasiswa yang mengajukannya. Hasil *testing* untuk mengukur performa algoritma *K-Nearest Neighbor* menggunakan metode *cross validation*, *Confusion Matrix* dan kurva *Receiver Operating Characteristic (ROC)*, akurasi yang diperoleh untuk beasiswa peningkatan prestasi akademik mencapai 88,33% dengan nilai *Area Under Curva (AUC)* 0,925 dari 227 record dataset, sedangkan akurasi yang diperoleh untuk beasiswa bantuan belajar mahasiswa mencapai

90% dengan nilai AUC 0,937 dari 183 *record* dataset, akurasi yang diperoleh untuk gabungannya beasiswa peningkatan prestasi akademik dan bantuan belajar mahasiswa mencapai 85,56% dan nilai AUC 0,958.

Dari beberapa penelitian yang telah dilakukan sebelumnya akan dijadikan acuan atau pedoman dalam penelitian ini. Adapun yang menjadi pembeda dari penelitian ini dengan penelitian sebelumnya adalah dari segi atribut yang digunakan. Pada penelitian ini menggunakan dua atribut yaitu latitude dan longitude. Atribut latitude diambil dari garis lintang dimana garis tersebut diambil dari titik tempat tinggal peserta didik, sedangkan atribut longitude diambil garis bujur tempat tinggal peserta didik itu.

Penelitian ini bertujuan untuk mempermudah kinerja dari Panitia Penerimaan Peserta Didik Baru (PPDB) didalam mengklasifikasi calon peserta didik baru yang akan diterima sesuai dengan wilayah zonasi.

2. Kerangka Teori

2.1. Data Mining

Data mining adalah proses yang menggunakan Teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. *Data mining* merupakan salah satu cabang ilmu komputer yang relative baru (Novita, 2016). Secara umum semua operasi dalam *data mining* dapat dikelompokkan menjadi dua metode yaitu metode deskriptif dan metode prediktif. Metode deskriptif bertujuan untuk menemukan pola, relasi, atau *anomaly* dalam data yang mudah dipahami oleh manusia. (Adinugroho dan Sari 2018). Teknik dalam *data mining* dibagi menjadi beberapa bagian kelompok diantaranya, (Kartika *et al.*, 2017) (Adinugroho dan Sari 2018), teknik *Classification* ini akan membentuk suatu klasifikasi yang data baru dengan memanipulasi data yang sudah ada. Proses pengelompokkan menggunakan acuan data yang telah diketahui kelompok atau kelasnya. Data yang belum memiliki kelompok dapat ditentukan kelompoknya melalui proses perbandingan dengan data yang sudah diketahui sebelumnya. Algoritma klasifikasi yang umum digunakan meliputi : *Decision Tree*, *Naïve Bayes*, *K-Nearest Neighbor*, jaringan saraf tiruan dan *Support Vector Machine*. Teknik dalam *data mining* selanjutnya adalah *Association* ini digunakan untuk mengenali kelakuan dari kejadian-kejadian khusus atau proses dimana hubungan asosiasi muncul pada setiap kejadian. *Association* merupakan metode pencarian pola relasi antar data dalam sebuah kumpulan data. Berdasarkan pola tersebut kemunculan suatu data dapat diprediksi berdasarkan kemunculan data lainnya. Algoritma ini meliputi *Apriori*, *Eclat*, *FP-growth*. Teknik yang paling sering digunakan dalam *data mining* adalah

Clustering yang digunakan untuk menganalisa pengelompokan berbeda terhadap data, *clustering* ini mirip dengan klasifikasi, namun pengelompokan belum didefinisikan sebelum dijalankannya *tool data mining*. Biasanya menggunakan metode *Neural Network* atau statistik. Pengelompokan data pada *clustering* berlangsung otonom dengan cara membandingkan semua data yang belum memiliki kelas dan membaginya kedalam beberapa kelas berdasarkan kemiripan antar data. Algoritma *clustering* meliputi: *K-Means*, *Fuzzy C-Means*, *Hierarchical clustering*, *Gaussian Mixtures*. Tugas dari klasifikasi adalah memprediksi keluaran variabel/*class* yang bernilai kategorikal atau *polynomial* (Saifudin, 2018). Proses klasifikasi didasarkan pada empat komponen diantaranya (Sumarlin, 2015) :

1. Kelas Variabel dependen yang berupa kategorikal yang merepresentasikan 'label' yang terdapat pada objek. Contohnya: resiko penyakit jantung, resiko kredit, customer loyalty, jenis gempa.
2. Predictor Variabel independen yang direpresentasikan oleh karakteristik (atribut) data. Contohnya: merokok, minum alkohol, tekanan darah, tabungan, aset, gaji.
3. Training dataset Satu set data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan predictor.
4. Testing dataset Berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi.

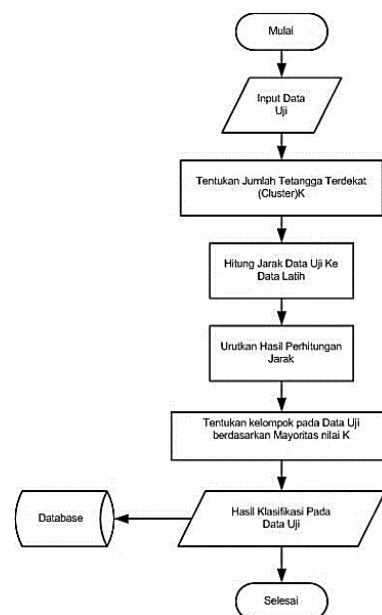
2.2. Algoritma *K-Nearest Neighbor (K-NN)*

Secara umum data mining memiliki teknik-teknik yang digunakan dalam melakukan pengklasifikasian. Pendekatan dari teknik-teknik itu terdiri dari 2 teknik, yaitu teknik *Supervised learning* dan *unsupervised learning*. *Supervised learning* adalah sebuah pendekatan dimana Algoritma *K-Nearest Neighbor* merupakan suatu metode yang menggunakan algoritma *supervised*. Perbedaan antara *supervised learning* dengan *unsupervised learning* adalah pada *supervised learning* bertujuan untuk menemukan pola baru dalam data dengan menghubungkan pola data yang sudah ada dengan data yang baru. Sedangkan pada *unsupervised learning* data belum memiliki pola apapun, dan tujuan *unsupervised learning* untuk menemukan pola dalam sebuah data. Dalam penelitian penerimaan peserta didik baru ini menggunakan algoritma *supervised learning*. Metode *K-Nearest Neighbor* bekerja berdasarkan asumsi bahwa suatu data akan memiliki kelas atau kategori yang sama dengan data yang berada disekitarnya. Konsep ini dikenal dengan konsep ketetanggaan (Adinugroho dan Sari, 2018). Menurut Gorunescu (Badu, 2016), *K-Nearest Neighbor* merupakan sebuah metode untuk melakukan klasifikasi terhadap objek baru berdasarkan (K) tetangga terdekat. Untuk

mendukung pengambilan keputusan tersebut, akan terlihat mayoritas dari keputusan teman atau tetangga. Teman atau tetangga tersebut dapat dipilih berdasarkan dengan kedekatannya. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan *Euclidean Distance*, atau dapat juga menggunakan rumus jarak yang lain. Dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori *K-Nearest Neighbor*. Dengan demikian *K-Nearest Neighbor* dari sebuah *instance x* didefinisikan sebagai *K instance* yang memiliki jarak terkecil (kedekatan terbesar, *nearest*) dengan *x*. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan *Euclidean Distance* yang direpresentasikan pada persamaan 1 sebagai berikut (Kartika *et al.*, 2017):

$$D(a, b) = \sqrt{\sum_{k=1}^d (a_k - b_k)^2} \quad (1)$$

dimana matriks *D (a,b)* adalah jarak skalar dari kedua vector *a* dan *b* dari matriks. Dalam menentukan nilai *K* sebaiknya dilihat dari jumlah klasifikasi bila jumlahnya genap maka sebaiknya menggunakan nilai *K* yang ganjil, dan sebaliknya jika jumlah klasifikasi jumlahnya ganjil maka sebaiknya dalam menggunakan nilai *K* yang genap, karena jika tidak begitu maka sistem kemungkinan tidak akan mendapatkan jawaban. Berikut ini adalah proses dari metode *K-Nearest Neighbor* yang ditunjukkan pada Gambar 1 (Noviansyah *et al.*, 2018) :



Gambar 1. Proses Metode *K-Nearest Neighbor*

2.3. *RapidMiner*

RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. *RapidMiner* memiliki sifat yang *open source* serta digunakan untuk berbagai teknik

yang bersifat deskriptif dan prediksi dalam membuat suatu keputusan (Aprillia *et al.*, 2013). Dalam penelitian ini menggunakan *confussion matrix* sebagai alat ukur performa algoritma klasifikasi. *Confussion matrix* atau matrik kebingungan merupakan sebuah perhitungan yang membandingkan dataset dengan hasil klasifikasi sesuai dengan data sebenarnya dengan jumlah keseluruhan data. Hasil akhir dari matrik ini adalah tingkat akurasi dengan satuan persen (%). Tingkat akurasi ini yang nantinya dijadikan acuan para peneliti terkait performa algoritma klasifikasi tersebut (Hasanah *et al.*, 2019). Berikut ini adalah model dari *Confussion matrix* yang terlihat pada tabel 1 dibawah ini (Hasanah *et al.*, 2019) :

Tabel 1. Model *Confussion matrix*

Correct Clasification	Classified as	
	+	-
+	True positive (A)	False negatives (B)
-	False positives (C)	True negatives (D)

Adapun rumus yang digunakan dalam model *Confussion matrix* adalah (Hasanah *et al.*, 2019) :

$$\text{Akurasi} = (A+D)/(A+B+C+D) \tag{2}$$

Hasil perhitungan tersebut dapat dijadikan nilai akurasi yang optimal untuk digunakan.

3. Metode

Model Perancangan Sistem Pada penelitian ini menggunakan pendekatan *Systems Development Life Cycle* (SDLC) yang digunakan dalam pengembangan perangkat lunak. Metodologi untuk Pengembangan System merupakan proses standard yang digunakan tim pengembang untuk menghubungkan semua langkah yang diperlukan untuk menganalisa, merancang, mengimplementasi, dan memelihara Sistem informasi (Inggi *et al.*, 2018). Dalam pengembangan framework dibutuhkan beberapa tahapan yang ada pada SDLC yaitu planning, analysis, design, implementation, dan maintenance seperti pada gambar berikut ini:(Inggi *et al.*, 2018).



Gambar 2. Tahapan SDLC

Tahapan pada Gambar 2 dapat dijelaskan sebagai berikut:

- Planning atau tahap perencanaan bertujuan untuk mengidentifikasi dan memprioritaskan sistem apa saja yang akan dikembangkan, dan sasaran-sasaran yang ingin dicapai.
- Analysis atau tahap analisis sistem merupakan tahap penelitian atas system yang telah ada dengan tujuan untuk merancang sistem baru atau memperbaharui sistem yang sudah ada.
- Design atau tahap perancangan sistem merupakan tahap untuk menentukan proses tahapan atau teknik untuk menerapkan sistem baru atau sistem yang dikembangkan dari sistem sebelumnya..
- Implementation atau tahap implentasi sistem merupakan tahap untuk mengimplementasikan rancangan dari tahap-tahap sistem yang dibangun atau dikembangkan serta melakukan uji coba terhadap sistem tersebut.
- Maintenance atau tahap pemeliharaan sistem merupakan proses pemeliharaan sistem selama penggunaan agar tetap mampu beroperasi secara benar

Siklus SDLC dijalankan secara berurutan, mulai dari langkah pertama hingga langkah keenam. Setiap langkah yang telah selesai harus dikaji ulang, kadang-kadang bersama *expert user*, terutama dalam langkah spesifikasi kebutuhan dan perancangan sistem untuk memastikan bahwa langkah telah dikerjakan dengan benar dan sesuai harapan. Jika tidak maka langkah tersebut perlu diulangi lagi atau kembali ke langkah sebelumnya (Inggi *et al.*, 2018).

4. Hasil dan Pembahasan

4.1. Analisis Metode K Nearest Neighbor

Proses perhitungan dengan metode *K-Nearest Neighbor* ini dilakukan untuk memperoleh label dari sistem zonasi berdasarkan kriteria-kriteria yang digunakan untuk penerimaan siswa baru berdasarkan zonasi. Jumlah data yang digunakan pada penelitian ini sebanyak 50 data yang terdiri dari data *training* dan data *testing*. Data training merupakan data yang diambil dari data peserta didik baru di SMKN 5 Kota Tangerang tahun ajaran 2018/2019. Data training dari penelitian ini sebanyak 35 data dari peserta didik baru di tahun itu. dimana data tersebut terdapat 18 data *training* zonasi dan 17 data *training* non zonasi. Atribut yang digunakan dalam data *training* ini adalah atribut *Latitude* dan *longitude*, atribut ini dipilih berdasarkan titik lokasi atau tempat tinggal dari peserta didik. Untuk data *training* dapat dilihat pada Tabel 2 berikut ini:

Tabel 2. Data Training

No	Nama	Latitude	Longitude	Klasifikasi
1	0001-0683	-6.217964	106.689143	zonasi
2	0002-0683	-6.188698	106.673039	zonasi
3	0003-0683	-6.224478	106.689594	zonasi
4	0004-0683	-6.196643	106.678416	zonasi
5	0005-0683	-6.164766	106.595306	zonasi
6	0006-0683	-6.266679	106.597233	zonasi
7	0054-0683	-6.180875	106.620347	non zonasi
8	0055-0683	-6.153566	106.577946	non zonasi
9	0056-0683	-6.2405	106.669329	non zonasi
10	0057-0683	-6.237395	106.65981	non zonasi
11	0058-0683	-6.243802	106.691642	non zonasi
12	0012-0683	-6.215279	106.647383	zonasi
13	0013-0683	-6.186139	106.660354	zonasi
14	0014-0683	-6.164599	106.676606	zonasi
15	0015-0683	-6.284518	106.762555	zonasi
16	0016-0683	-6.222399	106.698367	zonasi
17	0017-0683	-6.176125	106.640527	zonasi
18	0018-0683	-6.199004	106.652896	zonasi
19	0081-0683	-6.241366	106.645128	non zonasi
20	0082-0683	-6.236462	106.655592	non zonasi
21	0083-0683	-6.204738	106.638855	non zonasi
22	0084-0683	-6.185393	106.646851	non zonasi
23	0085-0683	-6.183004	106.681631	non zonasi
24	0086-0683	-6.184342	106.693005	non zonasi
25	0034-0683	-6.210392	106.659624	zonasi
26	0035-0683	-6.212079	106.688969	zonasi
27	0036-0683	-6.221551	106.692330	zonasi
28	0037-0683	-6.216907	106.682250	zonasi
29	0038-0683	-6.212079	106.688969	zonasi
30	0092-0683	-6.22871	106.707249	non zonasi
31	0093-0683	-6.221767	106.713783	non zonasi
32	0094-0683	-6.228821	106.715094	non zonasi
33	0095-0683	-6.220075	106.698915	non zonasi
34	0096-0683	-6.228839	106.704024	non zonasi
35	0097-0683	-6.232508	106.688134	non zonasi

Sedangkan sampel dari data *testing* yang digunakan sebanyak 15 data dari peserta didik baru di SMKN 5 Kota Tangerang ditahun ajaran 2019/2020. Atribut yang digunakan pada data *testing* ini juga sama dengan atribut yang digunakan pada data

training yaitu *latitude* dan *longitude*, data *testing* tersebut dapat dilihat pada Tabel 3 di bawah ini.

Tabel 3. Data Testing

No	Nama	Latitude	Longitude
1	100311120	-6.196706	106.662312
2	100721165	-6.195319	106.664415
3	100421103	-6.194487	106.656325
4	100411154	-6.204599	106.650446
5	100111138	-6.207628	106.645768
6	100511131	-6.210828	106.656819
7	100811206	-6.214220	106.642914
8	100410986	-6.215020	106.645285
9	100410801	-6.220235	106.657655
10	100610972	-6.223829	106.660971
11	100610821	-6.223947	106.666281
12	100110895	-6.225845	106.672665
13	100110983	-6.231466	106.675283
14	101010911	-6.235881	106.681731
15	100510820	-6.235124	106.684123

Pada tabel 3 telah diketahui letak *latitude* dan *longitude* dari peserta didik baru yang diukur berdasarkan tempat tinggal dari peserta didik baru. Pada langkah awal adalah dengan menentukan nilai K yang terbaik. K disini dimaksudkan sebagai titik dari objek yang akan diukur. Misalnya telah ditentukan 5 tetangga terdekat ($K=5$), berarti titik objek yang akan diambil sebanyak 5 titik objek. Setelah ditentukan nilai K maka langkah selanjutnya adalah dengan menghitung kuadrat jarak *eucliden* objek terhadap data *training* yang diberikan dengan menggunakan rumus (1), sedangkan Tabel 4 merupakan hasil K 5 terhadap jarak eucliden.

Tabel 4. Hasil K 5 terhadap jarak eucliden

NO	No Peserta	Euclidean Distance	Klasifikasi
1	100311120	0.034231639	zonasi
		0.013386433	zonasi
		0.016104123	zonasi
		0.023829213	zonasi
		0.010746872	zonasi
2	100721165	0.010872489	zonasi
		0.014063462	zonasi
		0.026239105	zonasi
		0.010038133	zonasi
		0.03305055	zonasi
3	100421103	0.017688141	zonasi

NO	No Peserta	Euclidean Distance	Klasifikasi
4	100411154	0.022195959	zonasi
		0.022633308	zonasi
		0.009269409	zonasi
		0.036119406	zonasi
		0.027627621	zonasi
5	100111138	0.029079526	zonasi
		0.034106629	non zonasi
		0.0111110552	zonasi
		0.020950896	zonasi
		0.033197174	zonasi
6	100511131	0.034446511	zonasi
		0.032912795	non zonasi
		0.007819592	zonasi
		0.025971687	zonasi
		0.027437662	zonasi
7	100811206	0.02583882	zonasi
		0.026734838	non zonasi
		0.010433096	zonasi
		0.024940789	zonasi
		0.039482757	zonasi
8	100410986	0.037261114	non zonasi
		0.028680227	non zonasi
		0.00459276	zonasi
		0.033055955	zonasi
		0.037886373	zonasi
9	100410801	0.035033475	non zonasi
		0.026676136	non zonasi
		0.002113926	zonasi
		0.032575864	zonasi
		0.028263387	zonasi
10	100610972	0.028935777	zonasi
		0.021920912	non zonasi
		0.01719923	non zonasi
		0.014463598	zonasi
		0.028776028	zonasi
11	100610821	0.028630357	zonasi
		0.018648818	non zonasi
		0.013615589	non zonasi
		0.016054166	zonasi
		0.018265668	zonasi
		0.016984102	zonasi
		0.015029901	non zonasi
		0.017281595	non zonasi
		0.026126239	non zonasi

NO	No Peserta	Euclidean Distance	Klasifikasi
12	100110895	0.019349512	zonasi
		0.015925981	zonasi
		0.010819578	non zonasi
		0.016570056	non zonasi
		0.02048887	non zonasi
13	100110983	0.019389601	zonasi
		0.01385118	zonasi
		0.013234227	non zonasi
		0.021973221	non zonasi
		0.012687402	non zonasi
14	101010911	0.017879206	zonasi
		0.011969509	zonasi
		0.015740515	non zonasi
		0.024418833	non zonasi
		0.011482293	non zonasi
15	100510820	106.7734218	zonasi
		106.7812781	zonasi
		106.7993521	non zonasi
		106.7554446	non zonasi
		106.8283389	zonasi

Pada tabel 4 merupakan hasil perhitungan dari jarak *eucliden* objek data testing terhadap data *training*. Bila dari 5 tetangga terdekat terdapat 3 label yang termasuk ke dalam wilayah zonasi maka dapat disimpulkan bahwa data yang dicari itu termasuk ke dalam wilayah zonasi. Sebaliknya, jika dari 5 tetangga tersebut terdapat 2 label yang termasuk ke dalam wilayah zonasi maka dapat disimpulkan bahwa data yang dicari itu termasuk ke dalam wilayah *non* zonasi. Hasil dari klasifikasi dengan nilai K 5 dapat dilihat pada tabel 5 dibawah ini.

Tabel 5. Hasil klasifikasi dari K5

No	No. Peserta	Klasifikasi
1	100311120	Zonasi
2	100721165	Zonasi
3	100421103	Zonasi
4	100411154	Zonasi
5	100111138	Zonasi
6	100511131	Zonasi
7	100811206	Zonasi
8	100410986	Non Zonasi
9	100410801	Non Zonasi
10	100610972	Non Zonasi
11	100610821	Zonasi
12	100110895	Zonasi
13	100110983	Non Zonasi
14	101010911	Non Zonasi
15	100510820	Zonasi

4.2. Penentuan Nilai Akurasi

Pada pengujian akurasi nilai K pada penelitian ini menggunakan aplikasi Rapid Miner versi 5.2. Dengan aplikasi tersebut akan menentukan nilai akurasi yang optimal untuk digunakan dalam memilih nilai K terdekat. Hasil akurasi nilai K dengan menggunakan rapid miner dapat dilihat pada gambar 3 berikut ini.

Nearest Neighbour Model (k-NN)	ExampleSet (Read CSV)	Performance Vector (Performance)
5-Nearest Neighbour model for classifi: The model contains 108 examples with 2 Zonasi	Data Table # Source: E:\data\test\k1\beranda\data file: file\test\data\beranda\beranda_rapid miner\data\training1.csv Number of examples = 108 3 attributes: Role Name Type Range Missing Co - latitude real [7..7] no - mean missing - - a? values	Performance Vector (Performance) accuracy: 83.36% +/- 12.05% (akurasi: 83) ConfusionMatrix: True: Zonasi non zonasi Zonasi: 53 10 non zonasi: 8 37

Gambar 3. Pengujian akurasi nilai K

Pada gambar 3 menunjukkan tingkat akurasi dari K5. Adapun kesimpulan dari nilai akurasi yang optimal untuk digunakan dapat dilihat pada Tabel 6 dibawah ini:

Tabel 6. Perbandingan jumlah cluster dan akurasi

Cluster	Akurasi
K 1	78,73 %
K 2	79,64 %
K 3	80,55 %
K 4	77,82 %
K 5	83,36 %
K 6	81,55 %
K 7	82,45 %
K 8	80,55%
K 9	82,45 %
K 10	77,00 %

Dari tabel 6 hasil pengukuran nilai akurasi dengan bantuan aplikasi *rapidminer* yang memiliki nilai persentase paling tinggi berada pada K5 dengan nilai akurasi sebesar 83,36%.

4.3. Evaluasi Confussion matrix

Perhitungan menggunakan *confussion matrix* untuk menentukan prosentase data yang sesuai dengan kenyataan dibandingkan jumlah keseluruhan data yang ada. Untuk perhitungan menggunakan model *Confussion matrix* dapat dilihat pada tabel 7 dibawah ini :

Tabel 7 Perhitungan *Confussion matrix*

Correct Classification	Classified as Zonasi	Non Zonasi
Zonasi	53	10
Non Zonasi	8	37

Perhitungan akurasi pada tabel 7 dapat menggunakan rumus (2) dengan hasil akurasi sebesar 83,33%.

5. Kesimpulan

Setelah melakukan analisa penelitian ini, maka dapat disimpulkan bahwa pada penerapan metode *K-Nearest Neighbor* (K-NN) yang digunakan telah mampu digunakan sebagai acuan dalam menentukan peserta didik yang diterima berdasarkan zonasi dan non zonasi sehingga dapat membantu pihak sekolah khususnya panitia penerimaan peserta didik dalam mengklasifikasi calon siswa untuk menentukan wilayah zonasi dan non zonasi. Dari percobaan perhitungan penentuan penerimaan peserta didik baru menggunakan perhitungan model *Confussion matrix* dan Rapidminer, Nilai akurasi yang diusulkan terletak pada K5 dengan tingkat akurasi sebesar 83,36%.

Ucapan Terima Kasih

Peneliti mengucapkan banyak terima kasih kepada semua pihak yang telah membantu peneliti hingga penyusunan dalam terbitan jurnal.

Daftar Pustaka

- Adinugroho, S. dan Y.A. Sari., 2018. Implementasi Data Mining Menggunakan WEKA. 1st ed. Malang: UB Press.
- Aprillia, Dennis, D.A. Baskoro, Lia, A. dan I.W.S. Wicaksana. 2013. Belajar Data Mining dengan RapidMiner. Jakarta: Remi Sanjaya.
- Badu, Z.S., 2016. Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Dana Desa.” (November).
- Hasanah, Riyan, L., Hasan, M. dan Pangesti, W.E., 2019. Klasifikasi penerimaan dana bantuan desa menggunakan metode K-NN (K-Nearest Neighbor). TECHNO Nusa Mandiri 16(1):1–6.
- Inggi, Rahmat, Sugiantoro, B. dan Prayudi, Y., 2018. Penerapan system development life cycle (SDLC) dalam mengembangkan framework audio forensik. SemanTIK 4(2):201–36.
- Kartika, Irjaya, J., Santoso, E. dan Sutrisno, 2017. Penentuan siswa berprestasi menggunakan metode K-Nearest Neighbor dan Weighted Product (Studi Kasus : SMP Negeri 3 Mejayan).” 1(5).
- Mendikbud, 2018. Peraturan Menteri Pendidikan Dan Kebudayaan Republik Indonesia No 51 Tahun 2018.” Retrieved June 12, 2019 (https://jdih.kemdikbud.go.id/arsip/PERMENDI_KBUD_NOMOR_51_TAHUN_2018.pdf).
- Noviansyah, Reza, M., Rismawan, T. dan Midyanti, D.M., 2018. Penerapan data mining menggunakan metode k-nearest neighbor untuk klasifikasi indeks cuaca kebakaran berdasarkan data AWS (Automatic Weather Station) Studi Kasus : Kabupaten Kubu Raya. Jurnal Coding, Sistem Komputer Untan, 06(2):48–56.

Novita, R., 2016. Teknik Data Mining : Algoritma C 4.5. 1–12.

Saifudin, A., 2018. Metode data mining untuk seleksi calon mahasiswa pada penerimaan mahasiswa baru di Universitas Pamulang. *Jurnal Teknologi* 10(1):25–36.

Sumarlin, S., 2015. Implementasi algoritma k-nearest neighbor sebagai pendukung keputusan klasifikasi penerima beasiswa PPA dan BBM. *Jurnal Sistem Informasi Bisnis* 5(1):52–62.

Umaedi, H. dan Siswantari, 2016. *Manajemen Berbasis Sekolah*. Edisi 1. edited by D. Setiawan and A. Suroso. Tangerang Selatan: Universitas Terbuka.