

PEMILIHAN SEKOLAH TERBAIK DENGAN MENGGUNAKAN METODE K-NEAREST NEIGHBORS DAN TAXONOMIC MATCHER

Bidari Ayu Lestari¹⁾; Muhammad Hasbi²⁾; Teguh Susyanto³⁾

^{1, 3)}Program Studi Sistem Informasi, STMIK Sinar Nusantara Surakarta

²⁾Program Studi Teknik Informatika, STMIK Sinar Nusantara Surakarta

¹⁾bidariayu92@gmail.com; ²⁾mhasbi@sinus.ac.id; ³⁾teguh@sinus.ac.id

ABSTRACT

The accuracy of choosing the right school is what every prospective student and parent wants. But in making the decision to choose the right school is not easy to do, because many aspects that are not simple must be taken into account. Mistakes in making decisions for prospective students must risk the loss of opportunities. Calculations in choosing a prospective student must be able to measure rationally the level of ability themselves with the quality of the school to be chosen. The quality of the school is determined based on the school's favorite level, the value of school accreditation, facilities owned, and achievements that have been achieved by the destination school. The purpose of this study was to apply the K-Nearest Neighbors (KNN) and Taxonomic Matcher methods to the creation of a system for selecting schools. The results of the development of the school selection application have been running in accordance with its functions and the results of the user acceptance test have been approved because it has a higher value than the answer agreed on the Likert scale which is 4.188571 on a scale of 1-5.

Kata kunci : K-Nearest Neighbors (KNN), Taxonomic Matcher, Pemilihan sekolah

I. PENDAHULUAN

Ketepatan memilih sekolah yang sesuai adalah hal yang diinginkan setiap calon siswa dan orang tua. Namun dalam pengambilan keputusan memilih sekolah yang tepat tidaklah mudah dilakukan, karena banyaknya aspek-aspek yang tidak sederhana yang harus diperhitungkan [1]. Kesalahan dalam mengambil keputusan calon siswa harus menanggung resiko hilangnya peluang.

Perhitungan dalam memilih sekolah calon siswa harus mampu mengukur secara rasional kadar kemampuan/prestasi yang dimiliki dengan kualitas sekolah tujuan. Kualitas sekolah ditentukan berdasarkan tingkatan favorit sekolah, nilai akreditasi sekolah, fasilitas yang dimiliki, dan prestasi-prestasi yang pernah diraih sekolah tujuan.

Penelitian terkait dengan pemilihan sekolah telah diusulkan dalam paper [1, 2, 3, 4, 5, 6]. Paper [1, 2, 3, 4] memanfaatkan metode algoritma Analytical Hierarchy Process (AHP), sedangkan paper penelitian [6] menggunakan metode *Technique for Order Preference by Similarity to Ideal Solution* (TOPSIS) dan paper [5] menggunakan metode Naive Bayes.

Pada penelitian kali ini yang membedakan dengan usulan penelitian di atas adalah penggunaan algoritma taxonomic matcher [7]. Dengan algoritma ini pemberian

nilai variabel (kriteria) ditentukan dengan model kelas-kelas yang tersusun hirarki.

Penelitian ini mengusulkan sebuah prototipe aplikasi perekomendasi untuk memilih sekolah terbaik dengan menerapkan algoritma *K-Nearest Neighbors (KNN)* dan *Taxonomic Matcher*.

II. TINJAUAN PUSTAKA

2.1 Sistem Pendukung Keputusan

Sistem Pendukung Keputusan (SPK) biasanya dibangun untuk mendukung solusi atas suatu masalah atau untuk suatu peluang. Aplikasi Sistem Pendukung Keputusan (SPK) digunakan dalam pengambilan keputusan.

Aplikasi Sistem Pendukung Keputusan (SPK) menggunakan CBIS (*Computer Based Information System*) yang fleksibel, interaktif, dan diadaptasi, yang dikembangkan untuk mendukung solusi atas masalah manajemen spesifik yang tidak teratur [8].

2.2 K-Nearest Neighbor (KNN)

Algoritma K-Nearest Neighbor (KNN) [9] adalah sebuah metode untuk melakukan klasifikasi terhadap obyek berdasarkan data pembelajaran yang jaraknya paling dekat dengan obyek tersebut. Prinsip kerja dari *K-Nearest Neighbors (KNN)* adalah mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga (neighbor) terdekatnya dalam data pelatihan.

Persamaan (1) menunjukkan rumus perhitungan untuk mencari jarak terdekat dengan d adalah jarak dan p adalah dimensi data [9] :

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \dots\dots\dots (1)$$

Dimana x_1 merupakan: sampel data; x_2 merupakan: data uji; d adalah besaran jarak; p merupakan dimensi data; dan i adalah variable data.

Contoh Perhitungan KNN:

Tabel 1 Contoh Data KNN

X1=Daya tahan keasaman (detik)	X2 = Kekuatan (kg/meter persegi)	Klasifikasi
7	7	Jelek
7	4	Jelek
3	4	Bagus
1	4	Bagus

Sebuah pabrik memproduksi kertas *tissue* baru yang memiliki $x_1 = 3$ dan $x_2 = 7$. digunakan algoritma KNN untuk melakukan prediksi termasuk klasifikasi apakah bagus atau jelek, kertas *tissue* yang baru ini.

Contoh perhitungan numerik

1. Tentukan parameter K = jumlah banyaknya tetangga terdekat.
2. Hitung jarak antara data baru dan semua data yang ada di data training. Hasil perhitungan ditunjukkan pada Tabel 2.

Tabel 2. Contoh Perhitungan KNN

X1=Daya tahan keasaman (detik)	X2 = Kekuatan (kg/meter persegi)	Square Distance ke data baru (3, 7)
7	7	$(7^2-3^2)+(7^2-7^2)=16$
7	4	$(7^2-3^2)+(4^2-7^2)=25$
3	4	$(3^2-3^2)+(4^2-7^2)=9$
1	4	$(1^2-3^2)+(4^2-7^2)=13$

3. Urutkan jarak tersebut dan tentukan tetangga mana yang terdekat berdasarkan jarak minimum ke- K . Hasil perangkaian ditunjukkan pada Tabel 3.

Tabel 3. Contoh Perangkaian KNN

X1=Daya tahan keasaman (detik)	X2 = Kekuatan (kg/meter persegi)	Square Distance ke data baru (3, 7)	Urutan (ranking) jarak	Apakah termasuk 3-NN?
7	7	$(7^2-3^2)+(7^2-7^2)=16$	3	YA
7	4	$(7^2-3^2)+(4^2-7^2)=25$	4	TIDAK
3	4	$(3^2-3^2)+(4^2-7^2)=9$	1	YA
1	4	$(1^2-3^2)+(4^2-7^2)=13$	2	YA

4. Tentukan kategori dari tetangga terdekat. Hasil pengkategorian tetangga terdekat diperlihatkan pada Tabel 4.

Tabel 4. Pengkategorian tetangga terdekat

X1=Daya tahan keasaman (detik)	X2 = Kekuatan (kg/meter persegi)	Square Distance ke data baru (3, 7)	Urutan (ranking) jarak	Apakah termasuk 3-NN?	Y = Category of nearest Neighbor
7	7	$(7^2-3^2)+(7^2-7^2)=16$	3	YA	Jelek
7	4	$(7^2-3^2)+(4^2-7^2)=25$	4	TIDAK	-
3	4	$(3^2-3^2)+(4^2-7^2)=9$	1	YA	Bagus
1	4	$(1^2-3^2)+(4^2-7^2)=13$	2	YA	Bagus

5. Gunakan kategori mayoritas yang sederhana dari tetangga yang terdekat tersebut sebagai nilai prediksi dari data yang baru.

Terdapat 2 kategori Bagus dan 1 kategori Jelek, karena $2 > 1$ maka dapat disimpulkan bahwa kertas tissue baru tadi yang memiliki $X_1=3$ dan $X_2=7$ termasuk dalam kategori Bagus.

2.3 Taxonomic Matcher

Metode KNN mempunyai beberapa kekurangan yaitu: metode KNN perlu menunjukkan parameter K (jumlah tetangga terdekat), tidak jelas perhitungan jarak mana yang sebaiknya digunakan dan atribut mana yang memberikan hasil yang baik, nilai komputasi metode KNN tinggi, karena beberapa kekurangan dari metode KNN diatas, maka perlu adanya metode lain yang membantu kinerja metode KNN.

Inti dari algoritma *Taxonomic matcher* adalah menghitung nilai kemiripan antara dua kelas c_1 dan c_2 berdasarkan jarak $d_c(c_1, c_2)$ antar keduanya [7].

Berikut adalah rumus *Taxonomic Matcher* yang dinyatakan pada persamaan (2), (3), (4) dan (5).

$$sim_c(c_1, c_2) = 1 - dc(c_1, c_2) \dots\dots\dots (2)$$

$$d_c(c_1, c_2) = d_c(c_1, ccp) + d_c(c_2, ccp) \dots\dots\dots (3)$$

$$d_c(c, ccp) = milestone(ccp) - milestone(c) \dots\dots (4)$$

$$milestone(n) = \frac{1}{k^{l(n)}} \dots\dots\dots (5)$$

Dimana Sim_c merupakan derajat kesamaan class 1 dan class 2; d_c adalah besaran jarak class 1 dan class 2; $milestone(n)$ merupakan jarak class parent dan class; $l(n)$ adalah depth/level class; k adalah konstanta ($k > 1$).

III. METODE PENELITIAN

Teknik penelitian diterapkan dalam paper ini adalah teknik rancang bangun perangkat lunak untuk memfasilitasi pemilihan sekolah dengan teknik pencarian multi kriteria dengan menggunakan kombinasi algoritma K-Nearest Neighbors dan *Taxonomic matcher*. Adapun

tahapan pengembangan perangkat lunak meliputi:

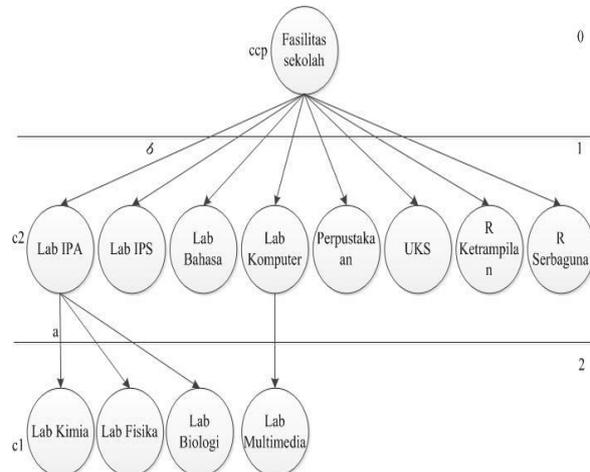
1. Pengumpulan Data
Pengumpulan data dilaksanakan dengan menggunakan teknik observasi, wawancara, dan studi pustaka. Adapun data digunakan dalam penelitian ini yaitu: (i). Dalam penelitian ini yang dijadikan data primer meliputi nama sekolah, alamat sekolah, no telepon sekolah, status sekolah, jenis sekolah, jarak atau wilayah, fasilitas, ekstra kurikuler, kondisi sekolah. (ii). Data sekunder dalam riset ini adalah sekolah yang sesuai dengan kriteria pencarian pengguna.
2. Analisis Sistem
Analisis sistem adalah penguraian dari sistem informasi yang utuh kedalam bagian-bagian komponennya dengan maksud untuk mengidentifikasi dan mengevaluasi permasalahan-permasalahan dan hambatan-hambatan yang terjadi dan kebutuhan-kebutuhan yang diharapkan dapat diusulkan perbaikan. Dan dapat memberikan makna yang berguna dalam memecahkan masalah-masalah penelitian. Analisa sistem yang digunakan meliputi pengkajian terhadap keseluruhan sistem yang ada di bagian administrasi.
3. Perancangan Sistem
Dalam perancangan sistem ini digunakan tool: Diagram Alir Data, Entity Relationship Data, Desain Input dan Output.
4. Implementasi
Aplikasi perangkat lunak yang dikembangkan menggunakan aplikasi berbasis web dengan bahasa pemrograman PHP, sedangkan untuk mengimplementasikan data diterapkan sistem basis data MySQL.
5. Pengujian
Untuk menguji keabsahan aplikasi yang dikembangkan digunakan dua pengujian yaitu teknik *blackbox* dan validitas. Teknik *blackbox* untuk memastikan kesesuaian secara fungsional aplikasi yang dikembangkan dengan perancangan aplikasi. Sedangkan uji validitas untuk menguji tingkat valid hasil output aplikasi dengan fakta nyata dari pandangan pengguna.

IV. HASIL DAN PEMBAHASAN

4.1. Pemilihan Sekolah Dengan Metode *Taxonomic Matcher* dan *K-Nearest Neighbors (KNN)*.

Penyelesaian kasus dengan pemodelan *Taxonomic Matcher* dan *K-Nearest Neighbors (KNN)* yaitu:

- 1). Menentukan jarak kedekatan antar nilai kriteria.



Gambar 1. *Taxonomic Matcher* fasilitas

Berdasarkan pemetaan hirarki yang diilustrasikan pada Gambar 1, dimisalkan Lab Kimia adalah c_1 dan Lab IPA adalah c_2 . Perhitungan jarak antara c_1 dan c_2 adalah sebagai berikut:

$$milestone(n) = \frac{1}{k^{l(n)}}$$

misal konstanta adalah 2

$$milestone(ccp) = \frac{1}{k^{l(n)}}$$

$$milestone(ccp) = \frac{1}{2^1} = \frac{1}{2} = 0,25$$

$$milestone(c_1) = \frac{1}{k^{l(n)}}$$

$$milestone(c_1) = \frac{1}{2^2} = \frac{1}{4} = 0,125$$

$$milestone(c_2) = \frac{1}{k^{l(n)}}$$

$$milestone(c_2) = \frac{1}{2^1} = \frac{1}{2} = 0,25$$

$$d_c(c_1, ccp) = milestone(ccp) - milestone(c_1)$$

$$d_c(c_1, ccp) = 0,25 - 0,125 = 0,125$$

$$d_c(c_2, ccp) = milestone(ccp) - milestone(c_2)$$

$$d_c(c_2, ccp) = 0,25 - 0,25 = 0$$

$$Sim(c_1, c_2) = 1 - d_c(c_1, c_2)$$

$$Sim(c_1, c_2) = 1 - (0,125 + 0)$$

$$Sim(c_1, c_2) = 1 - 0,125 = 0,875$$

Dari perhitungan di atas, dapat diketahui bahwa kemiripan c_1 dan c_2 adalah 0,875. Dengan cara di atas, dapat dicari nilai kedekatan antar nilai kriteria.

2). Menghitung kemiripan setiap kriteria pilihan dengan kriteria masing-masing sekolah. Seorang siswa ingin mencari sekolah:

- 1) Jenjang sekolah= SMP
- 2) Status sekolah = Swasta
- 3) Akreditasi = A
- 4) Fasilitas = Lab Kimia
- 5) Ekstrakurikuler = OSN
- 6) Kondisi sekolah = Baik

Menghitung kemiripan C1 langsung membandingkan antara C1 alternatif dengan C1 pilihan. Begitu juga dengan menghitung C2 dan C3. Sehingga diperoleh C1=1, C2 = 0,5, C3=1.

Cara menghitung kemiripan C4 adalah dengan membandingkan setiap fasilitas lalu dicari rata-ratanya.

$$C4 = \frac{\text{nilaikedekatan}}{\text{jumlahfasilitasyangdimiliki}}$$

$$C4 = \frac{0,375 + 0,375 + 0,375 + 0,375}{4}$$

$$C4 = \frac{1,5}{4}$$

$$C4 = 0,375$$

Jadi nilai C4 untuk sekolah alternatif 1 adalah 0,375. Cara menghitung kemiripan C5 adalah dengan membandingkan setiap ekstrakurikuler lalu dicari rata-ratanya.

$$C5 = \frac{\text{nilaikedekatan}}{\text{jumlahekstrayangdimiliki}}$$

$$C5 = \frac{0,375 + 0,375}{2}$$

$$C5 = \frac{0,75}{2}$$

$$C5 = 0,375$$

3). Menentukan jarak kriteria

$$\text{Jarak (d) A1} = C1+C2+C3+C4+ C5+C6$$

$$= 1 + 0,5 + 1 + 0,375 + 0,375 + 0,25$$

$$= 3,5$$

4). Menghitung nilai bobot

$$\text{Bobot A1}$$

$$= 15 \cdot C1 + 15 \cdot C2 + 15 \cdot C3 + 20 \cdot C4 + 20 \cdot C5 + 15 \cdot C6$$

$$= 15 \cdot 1 + 15 \cdot 0,5 + 15 \cdot 1 + 20 \cdot 0,375 + 20 \cdot 0,375 + 15 \cdot 0,25$$

$$= 15 + 7,5 + 15 + 7,5 + 7,5 + 3,75$$

$$= 56,25$$

5). Proses perangkingan

Hasil perhitungan bobot masing-masing nilai kemudian diurutkan berdasarkan nilai kemiripan terbesar. Tabel 6 memperlihatkan urutan rangking dari setiap alternatif sekolah.

Tabel 6. Data Hasil Perangkingan

No	Alternatif	C1	C2	C3	C4	C5	C6	Jumlah	Rangking
1.	A1	1	0,5	1	0,375	0,375	0,25	56,25	92
2.	A2	1	0,5	1	0,75	0,375	0,75	71,25	38
3.	A3	1	0,5	1	0,375	1	0,75	76,25	17
4.	A4	1	0,5	1	0,375	0,375	0,75	63,75	65
5.	A5	1	0,5	1	0,375	0,375	1	67,5	51
6.	A6	1	0,5	1	0,375	0,375	0,25	56,25	92
7.	A7	1	0,5	1	0,375	0,375	0,25	56,25	92
8.	A8	1	0,5	1	0,75	0,375	1	75	19
9.	A9	1	0,5	1	0,375	0,5	1	70	46
10.	A10	1	0,5	1	0,375	0,375	1	67,5	51

4.2 Pengujian Sistem

1). Pengujian Fungsional

Pengujian fungsional dilakukan dengan metode *black box*, yaitu pengujian dengan menguji semua komponen fungsional sistem apakah sudah berjalan sesuai perancangan atau belum.

Hasil pengujian menggunakan pengujian fungsional (*blackbox*) yang telah dilakukan pada masing-masing fungsi sistem diatas dapat ditarik kesimpulan bahwa aplikasi pemilihan sekolah menengah pertama di wilayah sragen ini berjalan sesuai dengan fungsinya.

2). Pengujian *User Acceptance*

Dalam pengujian *user acceptance* pada sistem pemilihan sekolah menengah pertama di wilayah sragen digunakan kuesioner dari responden untuk mengetahui apakah sistem tersebut sesuai dengan kebutuhan siswa sekolah dasar.

Berdasarkan data yang diperoleh diketahui bahwa dari 25 responden yang mengisi kuesioner didapatkan nilai rata-rata sebesar 4,188571. Hal ini menjelaskan bahwa dengan adanya uji *user acceptance* pada sistem pemilihan Sekolah Menengah Pertama dengan menggunakan metode *K-Nearest Neighbor* (KNN) dan *Taxonomic Matcher* di wilayah sragen telah disetujui karena memiliki nilai yang lebih tinggi dari jawaban setuju pada skala *likert*.

Tabel 5. Data Perhitungan Nilai Pengamatan Kuesioner

NO	Pernyataan							Jumlah
	P1	P2	P3	P4	P5	P6	P7	
1	4	4	4	3	3	4	3	25
2	3	4	4	4	4	3	4	26
3	4	4	4	4	4	4	4	28
4	4	4	4	4	4	4	4	28
5	4	4	4	4	4	4	3	27
6	5	4	5	4	5	4	5	32
7	5	5	5	4	5	4	4	32
8	5	5	3	3	5	5	4	30
9	5	5	5	5	5	5	5	35
10	4	4	4	3	3	4	4	26
11	3	4	4	4	4	4	4	27
12	4	4	3	4	3	4	4	26
13	4	4	4	3	3	4	4	26
14	4	3	4	4	4	4	4	27
15	4	4	3	3	4	3	3	24
16	4	4	4	4	4	4	4	28
17	5	5	4	3	4	4	4	29
18	5	5	5	5	5	4	4	33
19	5	5	5	5	5	5	5	35
20	5	5	5	5	3	5	4	32
21	5	5	5	4	5	5	4	33
22	5	5	5	5	4	5	5	34
23	5	5	5	4	4	5	4	32
24	5	5	5	5	4	4	4	32
25	5	5	4	4	4	4	4	30
Jumlah	111	111	107	100	102	105	101	737
Rata - Rata	4,44	4,44	4,28	4	4,08	4,2	4,04	4,211429

Hal ini membuktikan bahwa sistem pemilihan sekolah menengah pertama dengan metode metode *K-Nearest Neighbor* (KNN) dan *Taxonomic Matcher* di wilayah sragen memiliki kinerja sistem yang baik sehingga layak untuk digunakan oleh siswa.

3). Pengujian *Preprocessing* (*Taxonomic Matcher*)

Pengujian *preprocessing* dilakukan dengan membandingkan kekomplekan pohon *taxonomic matcher*.

Berdasarkan hasil pengujian menunjukkan bahwa perbandingan perhitungan program dengan perhitungan uji coba menggunakan metode *Taxonomic Matcher* pada pemilihan sekolah dari 20 sampel data memiliki 9 data yang berbeda dan mempunyai prosentase tingkat perbedaan hasil sebesar 80%. Ini menunjukkan bahwa semakin kompleks pohon *Taxonomic Matcher* maka hasil yang diperoleh akan semakin akurat

V. PENUTUP

5.1. Kesimpulan

Berdasarkan pengujian aplikasi pada masing-masing fungsi, telah berjalan sesuai dengan perancangan.

Kualitas output yang dihasilkan pada sistem pemilihan sekolah dengan menggunakan metode *K-Nearest Neighbor*

(KNN) dan *Taxonomic Matcher* telah dianggap relevan karena memiliki nilai rata-rata di atas 4 pada skala *likert*.

5.2 Saran

Data fasilitas sekolah dan data ekstra masih terbatas satu fasilitas dan satu ekstra sehingga pemilihan belum bisa akurat, untuk kedepannya bisa dikembangkan dengan penambahan fasilitas dan ekstra sekolah sehingga pemilihan akan lebih akurat.

DAFTAR PUSTAKA

[1] Faisal and S. D. H. Permana, "Sistem Penunjang Keputusan Pemilihan Sekolah Menengah Kejuruan Teknik Komputer Dan Jaringan Yang Terfavorit Dengan Menggunakan Multi-Criteria Decision Making," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 1, pp. 11–19, 2015.

[2] A. A. Harahap, "Sistem Pendukung Keputusan Penentuan Jurusan Dengan Metode Analytical Hierarchy Process (Studi Kasus : Smk Swasta Kartini Utama Sei Rampah)," *Pelita Inform. Budi Darma*, vol. 9, no. 2, pp. 13–20, 2015.

[3] M. Arifin, "Sistem Pendukung Keputusan Pemilihan Sekolah Terbaik Pada Kabupaten Serdang Bedagai Menggunakan Metode AHP dan SAW," *J. INFOTEK*, vol. 1, no. 2, pp. 96–105, 2016.

[4] I. Aqli, D. E. Ratnawati, and M. Data, "Sistem Rekomendasi Pemilihan Sekolah Menengah Atas Sederajat Kota Malang Menggunakan Metode AHP ELECTRE Dan TOPSIS," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 4, p. 279, 2016.

[5] T. D. Salma and Y. S. Nugroho, "Sistem Rekomendasi Pemilihan Sekolah Menengah Tingkat Atas Menggunakan Metode Naive Bayes," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 2, no. 2, p. 85, 2016.

[6] S. Dewi, G. Abdillah, and A. Komarudin, "Sistem Pendukung Keputusan Pemilihan Sekolah Menengah Atas Boarding School di Jawa Barat Menggunakan TOPSIS," in *Seminar Nasional Sistem Informasi dan Teknologi Informasi (SENSITEK) 2018*, 2018, pp. 132–136.

[7] R. Oldakowski and C. Bizer, "SemMF: A framework for calculating semantic

similarity of objects represented as RDF graphs," *4th Int. Semant. Web Conf.*, pp. 2–4, 2005.

- [8] E. Turban, J. E. Aronson, and T.-P. Liang, *Decision Support Systems and Intelligent Systems, 7th Edition*. Pearson, 2005.
- [9] D. Coomans and D. L. Massart, "Alternative k-nearest neighbour rules in supervised pattern recognition: Part 1. k-Nearest neighbour classification by using alternative voting rules," *Anal. Chim. Acta*, vol. 136, pp. 15–27, Jan. 1982.