JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 1, Nomor 2, Juni 2020, Page 45-51
ISSN xxxx-xxxx (media online)
Available Online at https://ejurnal.sisfokomtek.org/index.php/jumin
DOI xxxxx/xxxxx.xxxx

Implementasi Algoritma Modified Nearest Neighbourt (M-KNN) Untuk Klasifikasi Buku

Fricles Ariwisanto Sianturi[1], Agustina Simangunsong[2], R. Mahdalena Simanjorang[3], Petti Indrayati Sijabat[4]

[1,2,3,4]Manajemen Informatika, STMIK Pelita Nusantara, Medan, Indonesia
[1]sianturifricles@gmail.com*, [2]agustinasimangunsong93@gmail.com,
[3]relimamahdalenasimanjorang@yahoo.co.id, [4]petti.jabat@gmail.com

Email Penulis Korespondens: sianturifricles@gmail.com

Abstrak- Klasifikasi buku penting pada suatu perpustakaan dengan tujuan memudahkan para pengunjung dalam mencari buku. Terkhususnya pada perpustakaan STMIK Pelita Nusantara yang dapat membantu para mahasiswa/I dalam mendapatkan buku pada perpustakaan. Dengan metode yang ada pada data mining yaitu text mining, maka dengan permasalahan yang muncul pada perpustakaan STMIK Pelita Nusantara penelitian ini mengimplementasikan algoritma untuk klasifikasi buku pada perpustakaan STMIK Pelita Nusantara dengan Algoritma Modified Nearest Neighbour (M-KNN). Modifikasi dari metode ini bertujuan untuk mengatasi kelemahan mengenai jarak data dengan weight yang memiliki banyak masalah dalam outlier pada metode KNN. Penelitian ini dilakukan pada perpustakaan STMIK Pelita Nusantara dengan menggunakan 10 jumlah dataset. Pada masing-masing jumlah dataset akan dibagi secara acak menjadi data training dan data uji. Pengujian sistem ini dilakukan dengan nilai k yaitu 3. Dan didapatkan akurasi rata-rata nilai akurasi maksimum yang dihasilkan sistem sebesar 50% pada saat jumlah dataset 10.

Kata Kunci: Data Mining, Classification, Text Mining, Modified Nearest Neighbor Algorithm

Abstract- Book classification is important in a library intending to make it easier for visitors to find books. Especially in the STMIK Pelita Nusantara library which can help students get books from the library. With the existing method in data mining, namely text mining, with the problems that arise in the STMIK Pelita Nusantara library, this study implements an algorithm for book classification in the STMIK Pelita Nusantara library with the Modified Nearest Neighbor (M-KNN) algorithm. The modification of this method aims to overcome the weakness regarding the distance of the data with the weight which has many problems in outliers in the KNN method. This research was conducted at STMIK Pelita Nusantara library using 10 total datasets. Each number of datasets will be randomly divided into training data and test data. Testing of this system is carried out with a k value of 3.And the average accuracy of the maximum accuracy value produced by the system is 50% when the number of datasets is 10

Keywords: Schools, Teachers, Homerooms, Decision Support Systems, SAW Methods.

1. PENDAHULUAN

Jumlah koleksi buku dalam sebuah perpustakaan selalu mengalami penambahan buku-buku baru, seperti yang terjadi di STMIK Pelita Nusantara. Setiap tahun masing-masing program studi memiliki anggaran pengadaan buku. Klasifikasi buku secara manual akan menyulitkan petugas perpustakaan khususnya yang kurang berpengalaman[1].

Permasalahan yang diangkat dalam penelitian ini adalah Keterbatasan pengetahuan petugas memungkinkan terjadinya kesalahan dalam mengklasifikasi buku serta membutuhkan waktu yang lama, karena petugas tersebut minimal harus membaca resensi dan daftar isinya.[2] Untuk itu perlu ada mekanisme yang cepat dan objektif untuk klasifikasi koleksi buku perpustakaan[3]- [4].

Sedangkan tujuan dari penelitian ini adalah untuk mengimplementasikan algoritma M-KNN untuk klasifikasi buku pada perpustakaan di STMIK Pelita Nusantara. Dari beberapa literature mendefinikasikan bahwa Klasifikasi adalah teknik memetakan (*mengklasifikasikan*) data ke dalam satu atau beberapa kelas yang sudah didefinisikan sebelumnya.[5] Ada banyak teknik klasifikasi yang dapat digunakan, diantaranya adalah Naïve Bayes, K-Nearest Neighbor, dan Artificial Neural

Submitted: **08/05/2020**; Accepted: **30/05/2020**; Published: **14/06/2020**

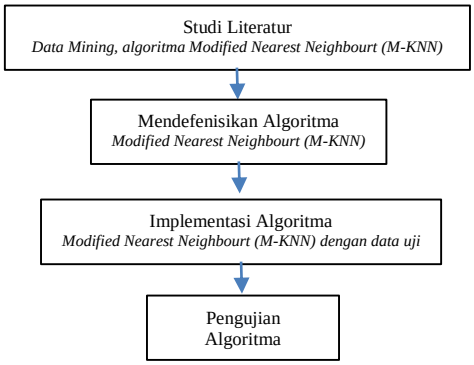
JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 1, Nomor 2, Juni 2020, Page 45-51
ISSN xxxx-xxxx (media online)
Available Online at https://ejurnal.sisfokomtek.org/index.php/jumin
DOI xxxxx/xxxxx.xxxx

Network. Penggunaan algoritma yang tepat dapat meningkatkan keakuratan keputusan yang diambil. Metode klasifikasi yang akan digunakan dalam penelitian ini adalah Modified k-Nearest Neighbour (MKNN)[6]. Dengan metode K-Nearest Neighbor (KNN) maka dikembangkan metode Modified K-Nearest Neighbor (MKNN). Gagasan utama dari metode ini adalah mengklasifikasikan sampel uji sesuai dengan tag tetangga. Metode ini adalah semacam KNN tertimbang sehingga bobot yang ditentukan dengan menggunakan prosedur yang berbeda[7]. Prosedur menghitung fraksi dari tetangga berlabel sama dengan total sejumlah tetangga [8]. Karena tingkat akurasi metode Modified K-Nearest Neighbor (MKNN) lebih baik bila dibandingkan dengan metode K-Nearest Neighbor (KNN). Seperti percobaan sebelumnya pada dataset Wine metode KNN meempunyai tingkat akurasi 83,79% sedangkan metode MKNN 85,76% dan juga pada dataset Isodata metode mempunyai tingkat akurasi K-Nearest Neighbor (KNN) 82,90% sedangkan metode Modified K-Nearest Neighbor (MKNN) 83,32%[9].
Alasan dilakukannya penelitian ini yaitu untuk memudahkan pegawai perpustakaan buku pada STMIK Pelita Nusantara dalam mengklasifikasi buku-buku yang ada pada perpustakaan sesuai dengan kebututuhan yang pada masing-masing program studi yang ada dilingkungan STMIK Pelita Nusantara[10]

2. METODOLOGI PENELITIAN

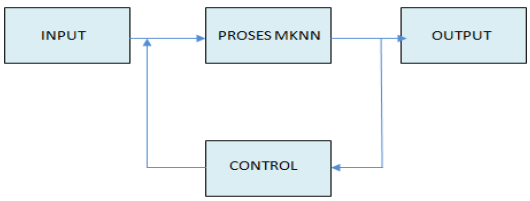
Penelitian melakukukan tahap awal yang dilakukan mempelajari tentang permasalahan-permasalahan yang muncul pada objek penelitian yaitu pada perpustakaan STMIK Pelita Nusantara tentang pengklasifikasian buku yang ada pada perpustakaan tersebut. Tahap berikutnya menerapkan algoritma untuk pengklasifikasian buku pada perpustakaan. Dalam kajian praktis, algoritma yang sesuai dengan permasalhan yang timbul adalah algoritma Modified Nearest Neighbourt (M-KNN), untuk lebih jelas dapat dilihat tahapan-tahapan penelitian pada gambar 1.

Setelah menerapkan algoritma Modified Nearest Neighbourt (M-KNN) selesai, selanjutnya menguji algoritma yang diterapkan terhadap data yang didapatkan dari perpustakaan STMIK Pelita Nusantara. Sampel diambil langsung dari perpustakaan STMIK Pelita Nusantara.



Gambar 1. Tahapan Peneltiian

Selanjutnya, penelitian ini melakuan pengklasifikasian buku dengan tahapan sebagai berikut: memasukkan label kelas dari data berdasarkan data poin pada data latih yang sudah divalidasi dengan nilai k. Dengan kata lain, pertama dilakukan perhitungan validitas data pada semua data di data latih. Selanjutnya, dilakukan perhitungan untuk mencari Weight Voting pada semua data uji menggunakan validitas data. selanjutnya adalah menjelaskan metode M-KNN secara detail mulai bagaimana menghitung validitas data sampai menentukan kelas akhir dari data sampel.



Gambar 2. Skema Kerja M-KNN

Submitted: **08/05/2020**; Accepted: **30/05/2020**; Published: **14/06/2020**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 1, Nomor 2, Juni 2020, Page 45-51
ISSN xxxx-xxxx (media online)
Available Online at https://ejurnal.sisfokomtek.org/index.php/jumin
DOI xxxxx/xxxxx.xxxx

Rumus yang digunakan untuk menghitung validitas setiap data latih adalah :

$$Validitas\ (x) = \frac{1}{H}\sum_{i=1}^{H} S\ (lbl(x), lbl(Ni(x))) \qquad (1)$$

Dalam metode MKNN, pertama weight masing-masing tetangga dihitung dengan menggunakan 1 / (de + 0.5). Kemudian, validitas dari tiap data pada data training dikalikan dengan weight berdasarkan pada jarak Euclidean. Dalam metode MKNN, weight voting tiap tetangga seperti pada persamaan berikut:

$$W(i) = Validitas(i) x \frac{1}{de+0,5} \qquad (2)$$

3. HASIL DAN PEMBAHASAN

Proses ini membahas dari tahapan algoritma Modified Nearest Neighbourt (M-KNN) dalam klasifikasi buku:

3.1. Tahapan Algoritma Modified Nearest Neighbourt (M-KNN)

1. Menghitung jarak (KNN)
2. Menghitung validitas
3. Menghitung weighted voting
4. Menentukan kelas kategori dan akurasi data.

Tabel 1. Ketentuan Bobot

Sampel	Keterangan	Bobot
SS	Sama sekali tidak	0,25
KK	Kadang-kadang	0,5
SR	Sering	0.75
SL	Selalu	1

Tabel 2. Data Set Kuisioner

No	Nama Buku	Jlh Buku
1	Sistem Pendukung Keputusan	11
2	Kriptografi	5
3	Rekayasa Perangkat Lunak	6
4	Data Mining	12
5	Algoritma Dan Pemrograman	7
6	Pengolahan Citra	6
7	Sistem basis Data	4
8	Jaringan komputer	8
9	Sistem Pakar	9
10	Multimedia	11

Tabel 3. Data TestingPerhitungan Manual

No	Nama Buku	Jlh Buku	Keterangan
1	Sistem Pendukung Keputusan	11	Betul
2	Kriptografi	5	Salah
8	Jaringan komputer	8	Betul
9	Sistem Pakar	9	Salah

Submitted: **08/05/2020**; Accepted: **30/05/2020**; Published: **14/06/2020**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 1, Nomor 2, Juni 2020, Page 45-51
ISSN xxxx-xxxx (media online)
Available Online at https://ejurnal.sisfokomtek.org/index.php/jumin
DOI xxxxx/xxxxx.xxxx

Tabel 4. Data Training Perhitungan Manual

No	Nama Buku	Jlh Buku	Keterangan
3	Rekayasa Perangkat Lunak	6	Salah
4	Data Mining	12	Salah
5	Algoritma Dan Pemrograman	7	Betul
6	Pengolahan Citra	6	Betul
7	Sistem basis Data	4	Betul
10	Multimedia	11	Salah

1. **Menghitung Jarak *Euclidean***

Data testing 1

$$d_{(1,3)=}\sqrt{\sum_{i=1}^{n}(x_{3i}-x_{1i}}$$

$$= \sqrt{(11-6)^2 + (20{,}25 - 27{,}25)^2}$$

$$= \sqrt{74} = 8{,}6023253$$

Selanjutnya melakukan perhitungan jarak *euclidean* yang sama untuk semua variabel *interval scaled*, dan hasilnya ditunjukkan pada tabel 5.

Tabel 5. Perhitungan Jarak *Euclidean* Data *Testing* 1

	Sum euclidean	Euclidean
d (1,3)	74	8,6023253
d (1,4)	56,5625	7,5208045
d (1,5)	111,0625	10,538619
d (1,6)	46,5625	6,823672
d (1,7)	117,0625	10,819543
d (1,10)	7,5625	2,75

Setelah nilai *euclidean* didapatkan, maka dilakukan proses normalisasi untuk memperoleh jarak dengan range [0,1] sesuai dengan persamaan yang ditunjukkan sebagai berikut.

$$d'(1,3) = \frac{(d_{(1,3)} - min_{euc})}{(max_{euc} - min_{euc})}$$

$$= \frac{(8{,}6023253 - 2{,}75)}{(10{,}819543 - 2{,}75)}$$

$$= 0{,}7252$$

Tabel 6. Hasil Normalisasi Jarak Euclidean Data Testing 1

	Euclidean	Normalisasi
d (1,3)	8,6023253	0,7252
d (1,4)	7,5208045	0,5912
d (1,5)	10,538619	0,9651
d (1,6)	6,823672	0,5048
d (1,7)	10,819543	1

Submitted: **08/05/2020**; Accepted: **30/05/2020**; Published: **14/06/2020**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 1, Nomor 2, Juni 2020, Page 45-51
ISSN xxxx-xxxx (media online)
Available Online at https://ejurnal.sisfokomtek.org/index.php/jumin
DOI xxxxx/xxxxx.xxxx

d (1,10)	2,75	0

Perhitungan untuk parameter variabel kategori menggunakan persamaan dengan p=1 dan d(i,j) = 0 jika sama dan 1 jika beda. Perhitungan variabel kategori ditunjukkan pada tabel 3.6 berikut.

Tabel 7. Hasil Perhitungan Jarak Variabel Kategori Data *Testing* 1

d (1,3)	0
d (1,4)	0
d (1,5)	0
d (1,6)	1
d (1,7)	1
d (1,10)	0

Setelah perhitungan jarak masing-masing variabel telah dihasilkan, akan digunakan perhitungan jarak rata-rata dengan menjumlahkan kedua hasil perhitungan dan membaginya sesuai dengan jumlah jenis variabel seperti berikut.

$$d_{(1,3)} = \frac{d_{var.kategori(1,3)} + d'e(1,3)}{2}$$
$$= \frac{0 + 0{,}7252}{2}$$
$$= 0{,}3626$$

Selanjutnya melakukan perhitungan jarak rata-rata yang sama untuk semua data training, dan hasilnya ditunjukkan pada tabel 8. berikut.

Tabel 8. Hasil Perhitungan Jarak

	Variabel	Euclidean	Total Jarak
d (1,3)	0	0,7252	0,3626
d (1,4)	0	0,5912	0,2956
d (1,5)	0	0,9651	0,4825
d (1,6)	0	0,5048	0,2524
d (1,7)	1	1	0,5000
d (1,10)	1	0	0

2. Menghitung Validitas

Menghitung nilai validitas data training dengan persamaan. Contoh perhitungan validitas data training untuk data training pertama (x=1) adalah sebagai berikut.

$$Validitas\ (x = 1) = \frac{1}{k}\sum_{i=1}^{k} s\ (llabel(x), \left(label\left(N\text{-}i(x)\right)\right))$$
$$= \frac{1}{3}\sum_{i=1}^{k} s\ (label(x = 1), \left(label\left(Ni(x = 2)\right)\right))$$
$$= \frac{1}{3}(1 + 0 + 0)$$
$$= \frac{1}{3}$$
$$= 0{,}333333$$

Perhitungan yang sama dilakukan untuk semua data training. Hasil perhitungan validitas ini seperti yang ditunjukkan pada tabel 9.

Submitted: **08/05/2020**; Accepted: **30/05/2020**; Published: **14/06/2020**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 1, Nomor 2, Juni 2020, Page 45-51
ISSN xxxx-xxxx (media online)
Available Online at https://ejurnal.sisfokomtek.org/index.php/jumin
DOI xxxxx/xxxxx.xxxx

Tabel 9. Perhitungan Manual Validitas Data Training

)	k =1	k=2	k=3	Sum S(a,b)	Validitas
3	1	0	0	1	0,333333
4	0	0	0	0	0
5	1	1	0	2	0,666667
6	1	0	1	2	0,666667
7	0	1	1	2	0,666667
10	0	0	0	0	0

Setelah hasil perhitungan validitas selesai, tahapan selanjutnya yaitu melakukan perhitungan *weighted voting* sesuai dengan persamaan. Sebagai contoh dilakukan perhitungan *weight voting* untuk data testing 1 sebagai berikut.

$$w_{1,3} = Validitas(1)x\frac{1}{d_{(1,3)} + 0{,}5}$$
$$= 0{,}333333x\frac{1}{0{,}3626 + 0{,}5}$$
$$= 0{,}3864$$

Selanjutnya melakukan perhitungan *weighted voting* yang sama untuk semua data training. Setelah didapatkan nilai *weighted voting* dari semua data training maka dilakukan pencarian nilai *weighted voting* terbesar sesuai dengan nilai k yang telah ditentukan, yaitu k=3 dan hasilnya ditunjukkan pada tabel 10 berikut.

Tabel 10. Hasil Perhitungan Weight Voting Data Testing 1

	Weight Voting	Kategori
d (1,6)	0,8859	Betul
d (1,5)	0,8375	Betul
d (1,7)	0,6666	Salah
d (1,3)	0,3864	Betul
d (1,4)	0	Betul
d (1,10)	0	Betul

Pada tabel 10. didapat 3 nilai weighted voting sebesar yaitu d(1,6) sebesar 0,8859 dengan kelas Impulsif, d(1,5) sebesar 0,8375 dengan kelas Impulsif, dan d(1,7) sebesar 0,6666 dengan nilai Hiperactif. Dari ketiga nilai *wighted voting* tersebut, setiap kelas kategori dijumlahkan dan hasil nilai terbesar akan dipilih menjadi keputusan. Dari data *testing* 1, maka kelas kategori ADHD yang ditentukan adalah "Impulsif" karena nilainya sebesar 1,7234 lebih besar dari jumlah *weighted voting* "Hiperactif" dengan jumlah sebesar 0,6666.

Perhitungan yang sama dilakukan untuk data testing selanjutnya yaitu data 2, 8, dan 9. Dari 4 data *testing* yang digunakan, dilakukan akurasi sebesar 50% dengan 2 nilai prediksi benar dan 2 prediksi salah yaitu pada data testing 2 dan 9. Hasil perhitungan dari 4 data testing ditunjukkan dalam tabel 11.

Tabel 11. Hasil Perhitungan Akurasi 4 Data Testing

Nilai k	Data testing	Hasil	
		Data Sebenarnya	Prediksi Sistem
3	1	Betul	Betul
	2	Salah	Betul
	8	Betul	Betul
	9	Salah	Betul
			50%

Submitted: **08/05/2020**; Accepted: **30/05/2020**; Published: **14/06/2020**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 1, Nomor 2, Juni 2020, Page 45-51
ISSN xxxx-xxxx (media online)
Available Online at https://ejurnal.sisfokomtek.org/index.php/jumin
DOI xxxxx/xxxxx.xxxx

4. KESIMPULAN

Kesimpulan yang didapat dalam penelitian ini dengan adanya pengclasifikasian buku pada perpustakaan di STMIK Pelita Nusantara dapat membantu para pengunjung untuk mencari buku sesuai dengan kebutuhan para pengunjung, hasil pengklasifikasian yang didapat dengan alagoritma Modified Nearest Neighbourt (M-KNN) idapatkan akurasi rata-rata nilai akurasi maksimum yang dihasilkan sistem sebesar 50% pada saat jumlah dataset 10.

REFERENCES

[1] A. Firman, H. F. Wowor, X. Najoan, J. Teknik, E. Fakultas, and T. Unsrat, "Sistem Informasi Perpustakaan Online Berbasis Web," E-Journal Tek. Elektro Dan Komput., 2016.
[2] M. S. Fricles Ariwisanto Sianturi, "KOMBINASI METODESIMPLEADDITIVEWEIGHTING (SAW)DENGANALGORITMA NEAREST NEIGHBOR UNTUK REKRUITMEN KARYAWAN," Mantik Penusa, vol. 3, no. 2, pp. 1–9, 2019, doi: .1037//0033-2909.I26.1.78.
[3] R. M. Anggraeni, "Perbandingan Algoritma Apriori dan Algoritma FP-Growth untuk Perekomendasi Pada Transaksi Peminjaman Buku di Perpustakaan Universitas Dian Nuswantoro," Tek. Inform., 2014.
[4] Sunjana, "Aplikasi Mining Data Mahasiswa dengan Metode Klasifikasi Decision Tree," Seminar Nasional Aplikasi Tknologi Informasi. 2010.
[5] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," J. Edik Inform., 2017.
[6] Suyanto, "Data mining Untuk Klasifikasi dan Klasterisasi Data," SpringerReference, 2017.
[7] P. Meilina, "Penerapan Data Mining dengan Metode Klasifikasi," J. Teknol. Univ. Muhammadiyah Jakarta, 2015.
[8] S. Yang, H. Jian, Z. Ding, Z. Hongyuan, and C. L. Giles, "IKNN: Informative K-nearest neighbor pattern classification," 2007, doi: 10.1007/978-3-540-74976-9_25.
[9] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2003, doi: 10.1007/978-3-540-39964-3_62.
[10] B. Kurniawan, S. Effendi, and O. S. Sitompul, "Klasifikasi Konten Berita Dengan Metode Text Mining," J. Dunia Teknol. Inf., 2012.

Submitted: **08/05/2020**; Accepted: **30/05/2020**; Published: **14/06/2020**