

Penerapan Algoritme *Modified K-Nearest Neighbour* Pada Pengklasifikasian Penyakit Kejiwaan Skizofrenia

Anjelika Hutapea¹, M. Tanzil Furqon², Indriati³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹anjelika.hutapea@gmail.com, ² m.tanzil.furqon@ub.ac.id, ³indriati.tif@ub.ac.id

Abstrak

Skizofrenia adalah penyakit yang mengalami keretakan jiwa atau keretakan kepribadian (*splitting of personality*). Masalah gangguan jiwa terjadi hampir di seluruh negara di dunia. Penyakit kejiwaan Skizofrenia terdiri dari lima jenis yaitu Paranoid, Hebefrenik, Katatonik, Tak Terinci dan Simpleks. Kemiripan gejala dari masing-masing jenis menimbulkan kesulitan paramedis untuk menentukan kelas penyakit Skizofrenia. Oleh karena itu dibutuhkan sistem yang bertujuan untuk melakukan klasifikasi penyakit kejiwaan Skizofrenia dengan menerapkan salah satu metode klasifikasi yaitu *Modified K-Nearest Neighbor* (MKNN). Langkah yang akan dilakukan oleh sistem adalah proses perhitungan jarak *assymetric binary*, perhitungan nilai validitas dan perhitungan *weight voting* sehingga mendapatkan hasil akhir yang akan digunakan untuk menentukan kelas klasifikasi berdasarkan masukan nilai K yang telah ditentukan. Pengujian sistem terdiri dari pengujian pengaruh nilai K dan pengujian pengaruh nilai K -Fold. Hasil dari pengujian pengaruh nilai K menghasilkan akurasi optimum senilai 37,045% pada nilai $K=7$ dan K -Fold=10. Hasil dari pengujian pengaruh nilai K -Fold menghasilkan akurasi terbesar senilai 28,4462% pada nilai K -Fold=5.

Kata kunci: *assymetric binary, K-Fold, validitas, weight voting.*

Abstract

Schizophrenia is a disease that has a soul crack or a splitting of personality. Problems of mental disorder almost in all countries in the world. Schizophrenia has a fifth types: Paranoid, Hebephrenic, Catatonic, Unspecified and Simplex. The similarity of symptoms causes the paramedic difficulty to determine the class of Schizophrenia. Therefore, a system that aims to Schizophrenia classification by applying one of the classification method Modified K-Nearest Neighbor (MKNN). The system will perform the step by calculating the distance of assymetric binary, the calculation of the validity value and weight voting value in order to get the final result that will be used to determine the classification based on the value of K that has been determined. The testing of this system consist of testing the effect of K value and testing the influence of K -Fold value. The result of testing the effect of K value gives the greatest accuracy equal to 37,045% at $K=7$ and K -Fold=10. The result of testing the effect of K -Fold value gives the greatest accuracy of 28,4462% at K -Fold=5.

Keywords: *assymetric binary, K-Fold, validity, weight voting.*

1. PENDAHULUAN

Masalah kesehatan terdiri dari berbagai jenis, salah satunya adalah gangguan kesehatan jiwa. Salah satu bentuk gangguan jiwa yang terdapat di dunia adalah Skizofrenia. Skizofrenia berasal dari kata "Skizo" yang berarti retak atau pecah, dan "frenia" yang berarti jiwa. Sehingga dapat disimpulkan Skizofrenia adalah penyakit yang mengalami keretakan jiwa atau keretakan kepribadian (*splitting of personality*) (Hawari, 2003).

Prevalensi penderita Skizofrenia di Indonesia berdasarkan data Risesdas 2013 terdapat gangguan mental emosional yang ditunjukkan dengan gejala-gejala depresi dan kecemasan untuk usia 15 tahun ke atas mencapai sekitar 14 juta orang atau 6% dari jumlah penduduk Indonesia. Sedangkan prevalensi gangguan jiwa berat, seperti Skizofrenia mencapai sekitar 400.000 orang atau sebanyak 1,7% per 1.000 penduduk (Kementerian Kesehatan Republik Indonesia, 2016).

Pada tiap jenis penyakit tersebut terdapat gejala yang saling beririsan sehingga terdapat kesulitan bagi paramedis untuk menentukan jenis penyakit tersebut. Oleh sebab itu, dibutuhkan sistem yang bertujuan untuk melakukan klasifikasi penyakit kejiwaan Skizofrenia dengan menggunakan teknik yang tepat serta memilih metode yang baik. Pada penelitian tersebut menggunakan salah satu metode yang merupakan implementasi dari data mining yaitu metode *Modified K-Nearest Neighbor* (MKNN). MKNN merupakan metode yang dapat digunakan untuk mendiagnosis penyakit Skizofrenia. MKNN adalah pengembangan dari metode KNN yang dirancang untuk mengatasi kelemahan dari jarak data dengan *weight voting* pada *outlier* KNN. Algoritme MKNN menambahkan proses baru untuk melakukan klasifikasi yaitu, perhitungan nilai validitas untuk mempertimbangkan validitas antar data latih dan perhitungan *weight voting* untuk menghitung bobot dari masing-masing terdekat. Penambahan dua proses baru dalam MKNN diharapkan dapat memperbaiki setiap kesalahan pada proses K-NN (Prasetyo, 2012).

2. SKIZOFRENIA

Skizofrenia merupakan salah satu bentuk gangguan psikosis yang menunjukkan beberapa gejala psikotik, ditambah dengan cerita lain seperti jangka waktu, konsekuensi dari gangguan tersebut dan tidak tumpang tindih dengan gangguan lain yang mirip (Arif, 2006).

Terdapat lima tipe atau kelompok pada penyakit kejiwaan Skizofrenia yaitu:

1. Skizofrenia Paranoid
2. Skizofrenia Hebefrenik
3. Skizofrenia Katatonik
4. Skizofrenia Tak Terinci
5. Skizofrenia Simpleks

Pada penyakit kejiwaan Skizofrenia terdapat dua bagian gejala yaitu gejala positif berupa waham atau delusi, halusinasi, kekacauan alam pikir, gaduh, merasa dirinya “Orang Besar”, pikirannya penuh dengan kecurigaan dan menyimpan rasa permusuhan serta gejala negatif berupa alam perasaan tumpul dan mendatar, mengasingkan diri, emosi yang tidak sesuai, pasif dan apatis, sulit dalam berfikir abstrak, serta tidak ingin apa-apa dan serba malas.

3. KLASIFIKASI

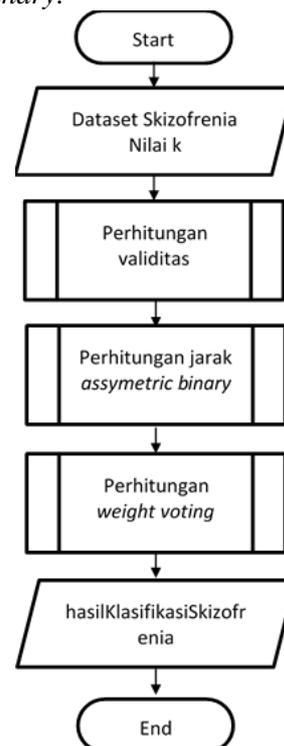
Klasifikasi adalah menentukan sebuah record data baru ke salah satu dari beberapa kategori (atau kelas) yang telah didefinisikan sebelumnya. Terdapat dua pekerjaan utama yang dilakukan dalam klasifikasi yaitu membangun model yang digunakan sebagai *prototype* untuk disimpan sebagai memori dan menggunakan model sebagai pengenalan/klasifikasi/prediksi untuk mengetahui letak kelas dari objek tersebut (Hermawati, 2013).

4. ALGORITME MODIFIED K-NEAREST NEIGHBOR (MKNN)

Metode *Modified K-Nearest Neighbor* (MKNN) merupakan sebuah metode yang memasukkan label kelas dari data berdasarkan poin pada data latih yang telah divalidasi dengan nilai *K*. Dalam metode ini, setiap sampel pelatihan harus didapatkan nilai validitas berdasarkan setiap titik sesuai dengan tetangganya (Parvin, 2010). Diagram alir dalam penyelesaian proses MKNN dapat dilihat pada Gambar 1.

4.1 Perhitungan Assymmetric Binary

Jarak kedekatan antara data latih dan data uji terdiri dari beberapa jenis. Salah satunya adalah jarak kedekatan untuk atribut bilangan biner. Untuk menghitung data biner pada atribut gejala penyakit digunakan rumus *assymmetric binary*.



Gambar 1. Diagram Alir Proses MKNN

Untuk menghitung jarak tersebut, dibutuhkan tabel kemungkinan untuk atribut biner yang akan ditunjukkan pada Persamaan 1.

$$d(x, y) = \frac{r+s}{q+r+s} \tag{1}$$

Keterangan: d adalah jarak antara titik data latih (x) dan titik data uji (y), yaitu x dan y sebanyak jumlah data yang tersedia. Dengan q , r dan s adalah hasil dari Tabel 1 (Han, 2012).

4.2 Perhitungan Validitas

Pada perhitungan validasi pada data bergantung pada nilai K terdekat dengan tujuan perhitungan ini adalah untuk menghitung jumlah titik dengan label sama pada data latih. Persamaan yang digunakan untuk menghitung nilai validitas pada data latih adalah seperti Persamaan 2 (Parvin, 2010).

Tabel 1. Perhitungan Assymetric Binary

		Data y		
		1	0	sum
Data x	1	q	r	$q + r$
	0	s	t	$s + t$
sum		$q + s$	$r + t$	p

$$Validity(x) = \frac{1}{H} \sum_{i=1}^H S(lbl(x), (N_i(x))) \tag{2}$$

Keterangan:

- H : Jumlah titik terdekat
- $lbl(x)$: Kelas x
- $N_i(x)$: Label kelas titik terdekat x
- S : Menghitung kesamaan antara titik a dan titik b berdasarkan k terdekat.

Untuk menghitung kesamaan antara antara titik kelas x dan kelas titik terdekat x dibuat variable baru yaitu a dan b . Variabel a adalah titik kelas x dan b adalah kelas titik terdekat x . Dalam menyamakan antara titik a dan data ke- b , digunakan fungsi S sebagai berikut:

$$S(a, b) = \begin{cases} 1, & a = b \\ 0, & a \neq b \end{cases} \tag{3}$$

Pada persamaan 3, a dan b merupakan label kelas pada data latih. S akan bernilai 1 jika label a sama dengan label b . S akan bernilai 0 jika label a tidak sama dengan label b (Parvin, 2010).

4.3 Perhitungan Weight Voting

Pada algoritme Modified K-Nearest

Neighbor (MKNN), Weight voting adalah salah satu metode KNN yang menggunakan K tetangga. Pada setiap K tetangga terdekat dihitung nilai validitas dan dikalikan dengan hasil perhitungan jarak assymetric binary antara data latih dan data uji. Nilai α merupakan nilai smoothing regulator (pemulusan) dengan nilai 0,5. Kemudian nilai weight voting tersebut akan dijumlahkan untuk tiap kelas dan jumlah yang menghasilkan nilai terbesar akan diambil sebagai nilai sebuah keputusan. Weight voting tersebut dapat dihitung dengan Persamaan 4 (Parvin, 2010).

$$W(x, y) = Validity(x) \times \frac{1}{d(x,y)+\alpha} \tag{4}$$

Keterangan:

- $W(x, y)$: Weight Voting data ke- x dan y
- $Validity(x)$: Nilai Validitas data ke- x
- $d(x, y)$: Jarak assymetric binary antara data latih dan data uji
- α : Smoothing regulator

4.4 K-Fold Cross Validation

Menurut Rohani, Abbas., et al. (dalam Jiang, Ping., 2017) K-Fold Cross Validation adalah salah satu dari jenis pengujian cross validation yang berfungsi untuk menilai kinerja proses sebuah metode algoritme dengan membagi sampel data secara acak dan mengelompokkan data tersebut sebanyak nilai K k-fold. Kemudian salah satu kelompok k-fold tersebut akan dijadikan sebagai data uji sedangkan sisa kelompok yang lain akan dijadikan sebagai data latih.

4.5 Akurasi Sistem

Perhitungan ini dilakukan untuk mengetahui tingkat akurasi metode MKNN dalam hasil klasifikasi. Akurasi dapat diperoleh dari persentase kebenaran, yaitu perbandingan antara jumlah data uji dengan jumlah data keseluruhan dikalikan 100%. Akurasi bisa didapat melalui persamaan berikut: (Han, 2012).

$$Akurasi = \frac{Jumlah\ data\ diprediksi\ benar}{Jumlah\ prediksi\ yang\ dilakukan} \times 100\% \tag{5}$$

5. PENGUJIAN

Pengujian dilakukan untuk mengetahui pengaruh nilai K dan pengaruh nilai k-fold terhadap akurasi sistem.



Gambar 2. Grafik pengujian pengaruh nilai K

a) Hasil pengujian pengaruh nilai K

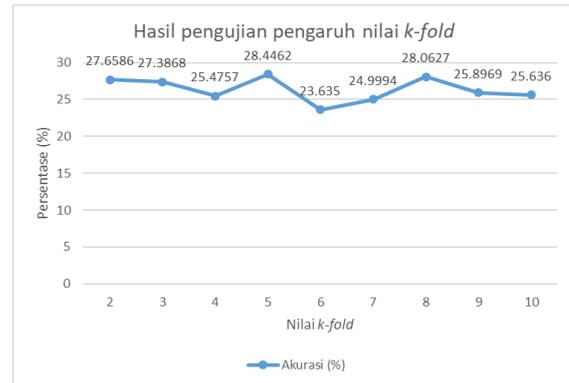
Pengujian pengaruh nilai K dilakukan menggunakan nilai $K=1$ hingga $K=10$ dengan nilai $k-fold=10$. Hasil pengujian pengaruh nilai K ini memperoleh nilai akurasi tertinggi yang terletak pada nilai $K=7$ dengan akurasi 37,045%. Hasil tersebut digambarkan pada grafik dan tabel yang ditunjukkan pada Gambar 2.

Hasil pengujian menunjukkan bahwa jumlah data latih memengaruhi akurasi sistem dalam melakukan pengklasifikasian penyakit kejiwaan Skizofrenia. Pada proses klasifikasi ini, algoritme MKNN mempelajari data latih berdasarkan pola yang ada. Akurasi yang naik turun disebabkan oleh beberapa hal antara lain kesesuaian data. Data latih yang digunakan mempunyai gejala yang memiliki sedikit kemiripan, sementara data berasal dari kelas yang sama. Sebaliknya, gejala yang memiliki banyak kemiripan terdapat pada data kelas yang berbeda. Hal ini menyebabkan sistem kurang menemukan pola dari data latih untuk melakukan klasifikasi.

b) Hasil pengujian pengaruh nilai k-fold

Pengujian pengaruh nilai $k-fold$ dilakukan menggunakan nilai $k-fold=2$ hingga $k-fold=10$. Pada tiap $k-fold$ dilakukan inputan nilai $K=1$ hingga $K=10$. Hasil pengujian pengaruh nilai $k-fold$ ini menghasilkan nilai akurasi yang berbeda sehingga dilakukan perhitungan nilai rata-rata untuk mengetahui akurasi terbaik.

Kesimpulan dari analisis hasil pengujian pengaruh nilai $k-fold$ adalah persentase terbaik terdapat pada nilai $k-fold=5$ dengan nilai akurasi 28,4462%. Pada pengujian tersebut mempunyai hasil nilai $k-fold$ terbaik tidak berada di satu nilai tertentu karena dari hasil grafik menunjukkan persentase akurasi terdapat nilai yang naik turun pada masing-masing $k-fold$. Hasil rata-rata pengujian pengaruh nilai $k-fold$ dapat dilihat pada Gambar 3.



Gambar 3. Grafik pengujian pengaruh nilai k-fold

Berdasarkan pengujian pengaruh nilai K dan pengaruh nilai $k-fold$ tidak terlihat perbedaan yang signifikan dikarenakan data yang berupa bilangan biner. Selain itu, gejala pada data latih tidak menunjukkan persamaan gejala yang dominan meskipun data latih terdapat pada kelas yang sama. Sebaliknya, data latih tersebut terdapat banyak kemiripan gejala sementara data latih tersebut ada pada kelas yang berbeda.

6. KESIMPULAN

Kesimpulan yang dihasilkan pada sistem penerapan algoritme *Modified K-Nearest Neighbor* (MKNN) pada pengklasifikasian penyakit kejiwaan Skizofrenia terbagi atas dua, yaitu:

1. Algoritme *Modified K-Nearest Neighbor* (MKNN) dapat diterapkan untuk pengklasifikasian penyakit kejiwaan Skizofrenia dengan proses menentukan nilai K dan nilai $k-fold$, mencari nilai validitas, jarak *assymetric binary* pada data latih dan data uji, nilai *weight voting* dan mendapatkan nilai akurasi sistem.
2. Hasil pengujian pengaruh nilai K, memperoleh nilai persentase optimum dengan akurasi 37,045% pada pengujian nilai $K=7$. Pada pengujian tersebut menggunakan nilai $K=1$ hingga $K=10$ karena pada nilai K setelah 10, persentase akurasi tetap berada di rentan angka yang berdekatan serta hasil pengujian menunjukkan bahwa apabila semakin banyak ketetangaan terdekat maka semakin banyak pula ketetangaan yang semestinya tidak dipertimbangkan menjadi pertimbangan dan masuk dalam proses perhitungan. Pada grafik terlihat bahwa nilai K berpengaruh pada akurasi namun

pada bentuk grafik tidak terdapat perbedaan yang signifikan. Sedangkan hasil pengujian pengaruh nilai *k-fold*, diperoleh nilai persentase optimum dengan akurasi 28,4462% dengan nilai *k-fold*=7. Pengujian pengaruh nilai *k-fold* ini mempunyai kesimpulan yaitu pembagian data menggunakan *k-fold* tidak menunjukkan perbedaan yang signifikan pada akurasi karena dalam kelas yang sama mempunyai sedikit kemiripan gejala sehingga sistem kesulitan dalam menentukan kelas yang tepat.

Discrimination on fNIRS by Using PCA and SVM. *International Conference on Bioinformatics and Biomedicine (BIBM)*. 389-394.

7. DAFTAR PUSTAKA

- Arif, 2006. *Skizofrenia - Memahami Dinamika Keluarga Pasien*. Bandung: Refika Aditama.
- Han Jiawei, Kamber Micheline, Pei Jian, 2012. *Data Mining Concepts and Techniques 3rd Edition*. San Fransisco: Morgan Kaufmann Publishers.
- Hawari, 2003. *Pendekatan Holistik Pada Gangguan Jiwa*. Jakarta: FKUI.
- Hermawati, Fajar Astuti., 2013. *Data Mining*. Yogyakarta: Andi.
- Kemertian Kesehatan Republik Indonesia, 2016. *Peran keluarga dukung kesehatan jiwa masyarakat*. Diakses pada tanggal 17 Desember 2017 dari <http://www.depkes.go.id/article/print/16100700005/peran-keluarga-dukung-kesehatan-jiwa-masyarakat.html>
- Kring,A.M., et al., 2014. *Abnormality of psychology 12th edition DSM V Update*. United States of America: John Wiley & Sons, Inc.
- Parvin, H., Alizadeh, H. & Minati, B., 2010. A Modification on K-Nearest Neighbor Classifier. *Global Journal of Computer Science and Technology*, 10 (14), 37-41.
- Prasetyo, Eko., 2012. *DATA MINING – Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: Andi.
- Rohani, Abbas., et al., 2017. A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I). *Renewable Energy*, 115, 411-422.
- Song, H., et al., 2016. Automatic Schizophrenia