

PREDIKSI PENYAKIT MATA KATARAK DAN NON KATARAK DENGAN MENGGUNAKAN METODE K-NEAREST NEIGHBOR

Fitra Kurnia¹, Septiani Wahyu Ningsi², Siti Monalisa³, Ichsan Fahmi⁴
UIN Sultan Syarif Kasim Riau, Program Studi Teknik Informatika Fakultas Sains dan Teknologi^{1,2}
UIN Sultan Syarif Kasim Riau, Program Studi Sistem Informasi, Fakultas Sains dan Teknologi³
fitra.k@uin-suska.ac.id¹
Program Studi Pendidikan Teknik Elektro FKIP Universitas Nusa Cendana, Kupang, NTT⁴
ichsan.fahmi@staf.undana.ac.id⁴

ABSTRACT

Various complaints and types of diseases related to eye disorders are increasing. Myopia, Cataracts, Dry eye, Trauma, Astigmatism Myopia, Pterygium, Presbyopia, and Conjunctivitis are some of the eye-related diseases. Cataract is an eye disease that continues to experience improvement. This study aims to classify cataract and non-cataract eye diseases using the K Nearest Neighbor method. This method classifies objects based on learning data that is closest to the object. Preliminary learning data used are data from January to December 2016 from Pekanbaru eye hospital with 8 types of eye diseases and 6 important attributes. Overall, the K Nearest Neighbor algorithm implemented in a web application with mysql DBMS has been able to predict cataract and non-cataract eye disease. The principle of this algorithm is to compare testing data (test data) with training data (training data). Based on the analysis of data requirements and the process of selection, cleaning and data transformation generated 6709 learning data records. With the choice of $k = 7$ the highest accuracy is 82.62% in the ratio of 90% of training data and 10% of test data. The lowest accuracy value is 68.15%, obtained at a ratio of 50% of training data and 50% of test data.

Keywords: Data mining, Cataracts, K-Nearest Neighbor, Eye Disease, Prediction

ABSTRAK

Berbagai keluhan dan jenis penyakit yang terkait gangguan pada mata semakin meningkat. Miopi, Katarak, Dry eye, Trauma, Astigmatisme Myopia, Pterygium, Presbiopia, dan Konjungtivitis adalah sebagian dari penyakit terkait mata. Katarak merupakan salah satu penyakit mata yang terus mengalami peningkatan. Penelitian ini bertujuan melakukan klasifikasi penyakit mata katarak dan non katarak dengan menggunakan metode K Nearest Neighbor. Metode ini melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang paling dekat dengan objek tersebut. Data awal pembelajaran yang digunakan merupakan data periode Januari sampai Desember 2016 dari rumah sakit mata Pekanbaru dengan 8 jenis penyakit mata dan 6 atribut penting. Secara keseluruhan algoritma K Nearest Neighbor yang diimplementasikan dalam aplikasi web dengan DBMS mysql ini telah mampu untuk memprediksi penyakit mata katarak dan non katarak. Prinsip algoritma ini adalah membandingkan data testing (data uji) dengan data training (data latih). Berdasarkan analisis kebutuhan data dan proses selection, cleaning dan data transformation dihasilkan 6709 record data pembelajaran. Dengan pilihan $k=7$ didapat akurasi tertinggi adalah 82,62% pada perbandingan 90% data latih dan 10% data uji. Nilai akurasi terendah adalah 68,15%, diperoleh pada perbandingan 50% data latih dan 50% data uji.

Kata kunci : Data mining, Katarak, K-Nearest Neighbor, Penyakit Mata, Prediksi

PENDAHULUAN

Mata merupakan salah satu dari lima panca indra yang dimiliki manusia. Mata memiliki fungsi untuk melihat, dengan adanya mata yang sehat kita dapat melakukan segala pekerjaan dengan baik, oleh karena itu mata merupakan salah satu organ yang penting dan harus dijaga kesehatannya. Banyak hal yang dapat mempengaruhi kerusakan kesehatan mata seperti faktor internal dan faktor eksternal. Faktor internal sendiri dipengaruhi oleh penderita yang mempunyai penyakit diabetes, hipertensi, ataupun cacat bawaan lahir, sedangkan faktor eksternal yaitu disebabkan oleh polusi udara, membaca buku terlalu dekat yang menyebabkan mata lelah, penggunaan smart phone yang berlebihan, bermain game di komputer yang terlalu lama. Penggunaan smart phone yang berlebihan dapat menyebabkan kerusakan dini pada mata.

Menurut data informasi Badan Penelitian dan Pengembangan Kesehatan(2013) yang telah menyelesaikan Riset Kesehatan Dasar (Riskesdas) mencatat setiap tahun kasus baru kebutaan selalu bertambah sebesar 0,1% dari jumlah penduduk atau kira-kira 250.000 orang/tahun. Catatan pada peringatan *World Sight Day* menginformasikan bahwa sekitar 285 juta orang diseluruh dunia hidup dengan low vision dan kebutaan. Dari jumlah tersebut 39 juta orang buta dan 246 juta memiliki gangguan penglihatan sedang dan berat. Dan 80 persen dari tunanetra dapat dihindari dengan cara diobati atau dicegah kerusakannya dari awal. Angka kebutaan di Indonesia terbilang cukup tinggi, berdasarkan hasil survey *Rapid Assesment of Avoidable Blindness* yang dilakukan oleh Kementerian Kesehatan dan direkomendasikan oleh WHO di 15 provinsi di Indonesia, dapat diketahui bahwa sebagian besar angka kebutaan berada di atas 2%. Dari hasil survey tersebut ditemukan bahwa katarak menjadi penyebab utama terjadinya kebutaan di Indonesia.

Penyakit mata katarak merupakan penyakit mata yang ditandai munculnya selaput lemak di lensa mata yang dapat membuat pandangan mata menjadi kabur. Biasanya katarak diderita oleh orang yang lanjut usia, tetapi tidak menutup kemungkinan balita dan anak-anak dapat juga mengalaminya. Penyakit mata katarak dapat dicegah apabila diketahui lebih dini, tetapi apabila sudah serius penderita penyakit mata katarak wajib melakukan operasi.

Data Mining (DM) adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Teknik data mining merupakan sebuah proses ekstraksi informasi untuk menggali pengetahuan dan menemukan pola pada tumpukan data dalam database yang biasanya berskala besar (Larose, 2005). Salah satu fungsi utama data mining adalah melakukan klasifikasi dan prediksi. Menurut Han dan Kamber (2006) klasifikasi adalah satu bentuk analisis data yang menghasilkan model untuk mendeskripsikan kelas data yang penting.

Terdapat beberapa algoritma klasifikasi yang sering digunakan oleh berbagai peneliti dalam menyelesaikan beragam problem. Algoritma Naïve Bayes, C4.5 dan *K- Nearest Neighbor*(K-NN) adalah sebagian algoritma yang sering menjadi pilihan para peneliti untuk klasifikasi. Sebagian peneliti mencoba membandingkan algoritma klasifikasi tersebut untuk mencari algoritma yang paling sesuai untuk problem yang diteliti. Widyadara dan Irawan (2019) menggunakan Naive Bayes untuk menentukan tingkat kesejahteraan keluarga, sedangkan Rahman, Hidayat dan Supianto (2018) membandingkan K-NN dengan *Naive Bayes* untuk klasifikasi kualitas air bersih. Perbandingan klasifikasi yang lebih kompleks lagi misalnya dilakukan oleh Arifin (2009). Dalam penelitiannya, Arifin (2009) memfokuskan pada perbandingan akurasi klasifikasi antara algoritma *Naive Bayes*, C4.5 dan *Oner (1R)*.

Teknik klasifikasi lain yang paling dasar dan sederhana adalah *K- Nearest Neighbor* (Parvin dkk, 2008). *K- Nearest Neighbor* (K-NN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Menurut T. Cover and P. Hart dalam ringkasan review perbandingan teknik K-NN oleh Bhatia dan Vandana (2010) diketahui bahwa K-NN memiliki beberapa keunggulan dan kelemahan. Keunggulan K-NN diantaranya adalah :1) pelatihan sangat cepat; 2) sederhana dan mudah dipelajari; 3) tahan terhadap data pelatihan yang memiliki derau; dan 4) efektif jika data pelatihan besar. Sedangkan kekurangan dari K-NN adalah: 1) nilai k menjadi bias; 2) komputasi yang kompleks; 3) keterbatasan memori; dan 4) mudah tertipu dengan atribut yang tidak relevan.

Algoritma K-NN telah banyak digunakan dalam penelitian untuk menyelesaikan beragam problem. Menurut Liu (2011) metode K-NN adalah metode klasifikasi yang menentukan kategori berdasarkan mayoritas kategori pada K-NN. Berdasarkan review yang dilakukan oleh Wu dan Kumar (2009), diketahui bahwa K-NN merupakan salah satu algoritma yang paling banyak digunakan diantara algoritma DM yang lain. Kelebihan K-NN yaitu tangguh terhadap data training yang noisy dan efektif apabila data latihnya besar. Pemilihan nilai k menjadi hal yang penting karena mempengaruhi kinerja algoritma K-NN. Nilai k yang terlalu kecil, maka

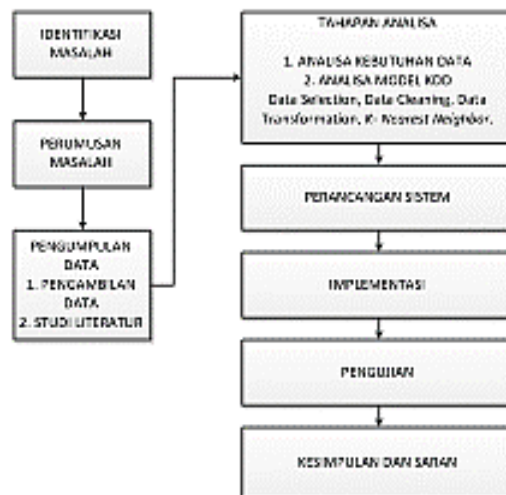
hasil klasifikasi menjadi lebih terpengaruh oleh noise. Di sisi lain, jika nilai k terlalu tinggi dapat mengurangi efek noise pada klasifikasi, tetapi membuat batasan antara setiap klasifikasi menjadi lebih kabur. Sampai saat ini masih belum diketahui nilai k -Optimal dalam algoritma K-NN untuk studi kasus yang diteliti.

METODE PENELITIAN

Metodologi penelitian adalah acuan dan tahapan yang diterapkan pada sebuah penelitian untuk mencapai tujuan penelitian. Adapun tahapan-tahapan yang ditempuh dalam penelitian terlihat pada Gambar 1.

Tahap Identifikasi Masalah

Berdasarkan observasi awal dapat diketahui bahwa penyakit mata katarak termasuk penyakit yang paling banyak diderita oleh pasien. Katarak memiliki banyak gejala kemiripan yang sama dengan penyakit mata lainnya, sehingga banyak pasien yang tidak menyadari sampai saat lensa mata telah menjadi sangat buram dan mengganggu penglihatan. Jika katarak sudah cukup menyebar maka dibutuhkan tindakan yang lebih serius lagi yang membutuhkan biaya yang sangat besar, misalnya operasi pengangkatan.



Gambar 1. Tahapan Penelitian

Tahap Perumusan Masalah

Perumusan masalah diperlukan untuk memfokuskan penelitian sehingga sistem yang dikerjakan tidak melebihi batas yang sudah ditentukan sebelumnya. Pada tahap ini dilakukan pengklasifikasian dan prediksi penyakit mata dengan metode K-NN

Tahap Pengumpulan Data

Salah satu komponen penting yang menjadi bagian dari penelitian adalah proses pengumpulan data. Proses ini berpengaruh pada tahapan analisis yang dilakukan, termasuk pada hasil dan kesimpulan yang dibuat. Kesalahan dalam proses pengumpulan data dapat menyebabkan kesalahan pada analisis dan kesimpulan. Tahapan dalam pengumpulan data dapat dijelaskan sebagai berikut :

Data penyakit mata yang digunakan dalam penelitian ini didapat dari Rumah Sakit Mata SMEC Pekanbaru yang terletak di jalan Arifin Ahmad no. 92. Data yang didapat merupakan data periode bulan Januari sampai Desember tahun 2016. Berdasarkan data awal diketahui bahwa terdapat 8 jenis penyakit mata yaitu miopi, katarak, *dry eye*, *trauma*, *Astigmatisme Myopia*, *Pterygium*, *Presbiopia*, dan *Konjungtivitis*.

Studi Literatur

Pada penelitian ini literatur yang dikumpulkan antara lain jurnal dan artikel penelitian sebelumnya, buku atau *e-book* yang terkait metode K-NN dan Penyakit Katarak.

Tahapan Analisis

1. Analisis Kebutuhan Data

Pada penelitian ini data diambil dari rekam medik yang dikumpulkan di Rumah Sakit Mata SMEC Pekanbaru. Data yang dipakai pada penelitian ini yaitu data periode Januari sampai Desember tahun 2016, Data penelitian ini terdiri dari 6 atribut penting yaitu nomor rekam medik, tanggal, jenis kelamin, umur, gejala, dan penyakit.

2. Analisis *Knowledge Discovery in Database* (KDD)

a. *Data selection*

Pada data awal terdapat 6 atribut yaitu nomor rekam medik, jenis kelamin, umur, gejala, dan penyakit. Setelah dilakukan proses selection, atribut yang dipakai yaitu jenis kelamin, umur, gejala, dan penyakit. Tabel 1 berikut secara ringkas menunjukkan struktur data hasil seleksi atribut.

Tabel 1. Data Hasil Seleksi pada Atribut

No.	Atribut	Keterangan
1.	Umur	Umur dari pasien (Dibagi menjadi beberapa kategori)
2.	Jenis Kelamin	Jenis Kelamin Pasien (Laki-laki dan Perempuan)
3.	Gejala	Berisi berupa keluhan yang dialami oleh pasien penyakit mata. (Data gejala akan dikelompokkan)
4.	Penyakit	Jenis penyakit pasien (Katarak dan Non Katarak/ Miopi)

b. *Data Cleaning*

Data awal yang telah didapat dari rumah sakit mata Pekanbaru dibersihkan dengan cara menghilangkan *noise* seperti data yang tidak relevan, atribut kosong, menghilangkan data duplikat dan menggantikan data yang *missing value*. Terdapat 17 record data yang *missing value* dan data yang telah di cleaning berjumlah 2000 *record*. Tabel 2 berikut adalah contoh data *missing value*.

Tabel 2. Data *Missing Value*

No.	Jenis Kelamin	Umur	Gejala	Penyakit
1.	Perempuan		Kabur jauh, mata sakit, pusing, mata mudah lelah	Myopia
2.	Laki-laki	4 bulan	Ada benjolan di kelopak mata	
3.	Perempuan		Kabur, berasap, silau,	Katarak
4.	Laki-laki		Kabur jauh, mata sakit, pusing, berkedip	Myopia
5.	Laki-laki	9 bulan	Juling	

c. *Data Transformation*

Data awal yang berupa kalimat diinisialisasi menjadi angka atau *numeric*. Inisialisasi atribut umur menjadi kode atribut yang unik dapat dilihat pada tabel 3. Atribut lain yang penting untuk diinisialisasi adalah gejala.

Tabel 3. Inisialisasi atribut umur

No.	Atribut	Kode Atribut	Range
1.	Umur	1	Sebelum kelahiran
		2	Bayi
		3	Anak awal
		4	Anak tengah
		5	Remaja awal
		6	Remaja akhir
		7	Dewasa awal
		8	Dewasa tengah
		9	Dewasa akhir

d. K Nearest Neighbor

Pada tahapan ini penelitian fokus pada penerapan metode K-NN pada prediksi penyakit mata. Secara ringkas langkah – langkah dalam menerapkan metode K-NN adalah sebagai berikut:

1. Menentukan nilai K.
2. Menghitung kuadrat jarak *euclidean distance* masing-masing objek terhadap data training yang diberikan.
3. Kemudian mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai *euclidean distance* terkecil.
4. Menghitung rata-rata dari nilai objek pada jangkauan K dengan menggunakan kategori K-Nearest Neighbor yang terdekat (jangkauan K) maka dapat diprediksi nilai query instance yang telah dihitung.

Perancangan Sistem

Proses perancangan sistem terdiri dari dua tahapan yaitu tahapan perancangan database dan tahapan perancangan antar muka (*interface*). Perancangan database terkait *Context Diagram*, *Data Flow Diagram* (DFD), dan *Flowchart*. Perancangan antar muka yang tepat memudahkan komunikasi antara sistem dan pengguna sistem.

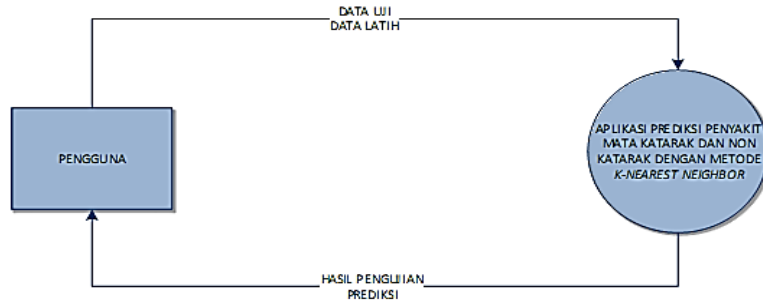
HASIL DAN PEMBAHASAN

Sesuai dengan analisis kebutuhan data dan proses KDD dihasilkan data berjumlah 6709 record. Data ini terdiri dari 6 atribut penting yakni Nomor, Nomor Rekam Medik, Jenis Kelamin, Umur, Gejala dan Penyakit. Dalam seleksi data yang telah dilakukan maka yang digunakan dalam penelitian ini hanya fokus pada dua penyakit yaitu penyakit katarak dan miopi(non katarak).

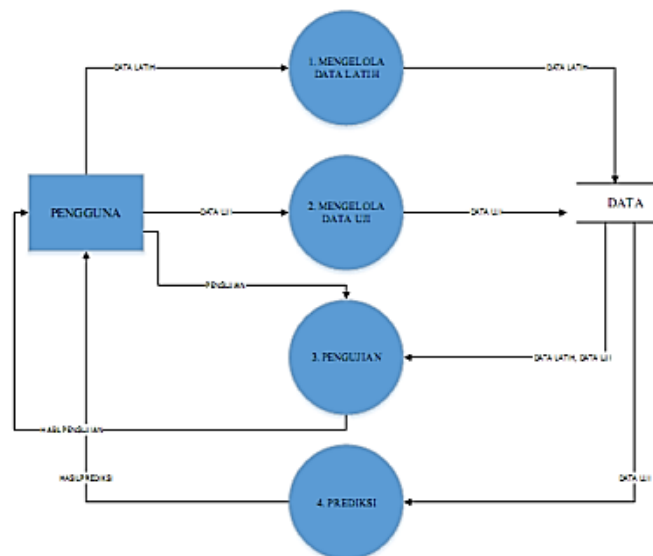
Berdasarkan analisis dan perancangan sistem maka beberapa hasil yang telah dilakukan dalam penelitian ini adalah :

1. Context Diagram dan Data Flow Diagram (DFD)

Gambar 2 tersebut memiliki satu entitas yaitu pengguna yang memiliki peran sebagai pengolah sistem yang dapat mengolah data latih dan data uji. Pengguna tersebut dapat juga melakukan pengujian terhadap sistem dan melakukan prediksi terhadap data penyakit mata katarak dan non katarak. Hasil perancangan DFD level 1 dapat dilihat pada gambar 3. Pada gambar tersebut terdapat 4 proses utama yaitu mengelolah data latih, mengelolah data uji, melakukan pengujian dan prediksi.



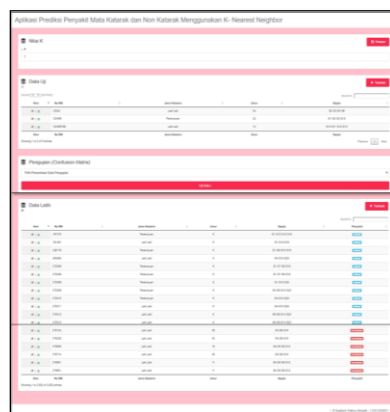
Gambar 2. Context Diagram



Gambar 3. Data Flow Diagram level 1

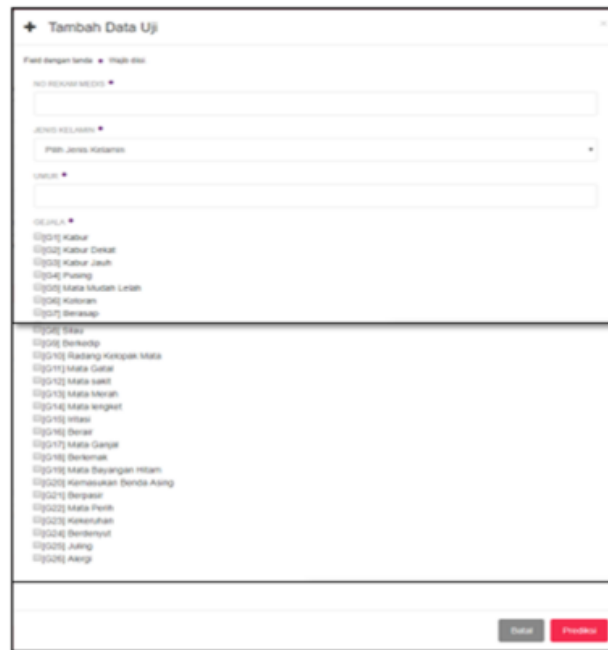
2. Halaman antar muka

Salah satu hasil akhir perancangan antar muka tersaji pada gambar 4. Gambar tersebut merupakan menu tampilan halaman utama yang menampilkan menu data latih dan data uji. Pada sistem ini pengguna bisa menambahkan nilai k nya.



Gambar 4. Tampilan Halaman Utama

Pengguna dapat juga menambahkan data uji dan data latih. Selain itu, pengguna dapat juga melihat pengujian *Cofusion Matrix* untuk melihat nilai akurasi yang didapat. Gambar 5 merupakan halaman tambah data uji.



Gambar 5. Halaman Tambah Data Uji

Tahap evaluasi dilakukan untuk pengujian akurasi sistem dalam memprediksi penyakit mata katarak dan non katarak dengan menggunakan metode K-NN. Adapun tahapan pengujian yang dilakukan yaitu dengan membagi data latih dan data uji berdasarkan pembagian kelas dari 3502 data. Untuk menghitung akurasi *confusion matrix* digunakan rumus sebagai berikut:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots(1)$$

Adapun contoh hasil perhitungan cara pembagian data latih dan data uji adalah: 90% data latih : 10% data uji

Tabel 4. Pengujian 90% (data latih) dan 10% (data uji)

No	Actual Class	Predicted Class	
		Katarak	Non Katarak
1.	Katarak	177	46
2.	Non Katarak	15	113

Dalam tabel 4. tersebut, perbandingan yang digunakan adalah 90% data latih dengan 10% data uji. Data latih yang digunakan berjumlah 3151 data dan data ujinya adalah 351 data. Perhitungan akurasi dilakukan sebagai berikut:

$$Akurasi = \frac{177 + 113}{177 + 15 + 113 + 46} \times 100\% = 82,62\%$$

Contoh lain untuk 50% data latih : 50% data uji adalah tampak pada tabel 5.

Tabel 5. Pengujian 50% (data latih): 50% (data uji)

No.	Actual Class	Predicted Class	
		Katarak	Non Katarak
1.	Katarak	629	543
2.	Non Katarak	15	565

Dalam tabel 5 tersebut, perbandingan data yang digunakan adalah 50% data latih dan 50% data uji. Data latih berjumlah 1750 data dan data uji berjumlah 1752 data. Perhitungan akurasi dilakukan sebagai berikut:

$$Akurasi = \frac{629 + 565}{629 + 15 + 565 + 543} \times 100\% = 68,15\%$$

Masing masing rasio memiliki nilai k=7. Setiap rasio dilakukan pengujian dan menghasilkan tingkat akurasi dari sistem. Pengujian akurasi beberapa perbandingan data latih dan data uji secara ringkas adalah sebagai berikut:

1. Perbandingan 50% : 50% di dapat akurasi sebesar 68,15 %.
2. Perbandingan 60% : 40% di dapat akurasi sebesar 69,10%.
3. Perbandingan 70% : 30% di dapat akurasi sebesar 70,60%.
4. Perbandingan 80% : 20% di dapat akurasi sebesar 73,61%.
5. Perbandingan 90% : 10% di dapat akurasi sebesar 82,62%.

KESIMPULAN

Metode *K- Nearest Neighbor* yang dirancang dalam bentuk web dapat digunakan untuk memprediksi penyakit mata katarak dan non katarak. Pada penelitian ini akurasi tertinggi adalah 82,62% yang diperoleh pada k=7 dengan perbandingan 90% data latih dan 10% data uji, sedangkan akurasi paling rendah adalah 68,15% didapat pada perbandingan 50% data latih : 50% data uji.

DAFTAR PUSTAKA

- Arifin.2009, *Perbandingan Akurasi Klasifikasi Dari Algoritma Naïve Bayes, C4.5, dan Oner (IR)*. TEKNO, Vol : 12, ISSN: 1693-8739
- Badan Penelitan dan Pengembangan Kesehatan Kementerian Kesehatan RI.2013.RISET KESEHATAN DASAR RISKESDAS
- Bhatia dan Vandana.2010. *Survey of Nearest Neighbor Techniques. International Journal of Computer Science and Information Security.*
- Han, J.,& Kamber, M. 2006. *Data mining Concept and Tehniques.* San Fransisco : Morgan Kauffman.
- Larose, D. T. 2005. *Discovering Knowledge in Data.* New Jersey : John Willey & Sons, Inc
- Liu, Bing. 2011.*Web Data Mining, 2nd Edition: Exploring Hyperlinks, Contents, and Usage Data.*
- Parvin, H., Alizadeh, H., & Minati, B. 2008.*MKNN : Modified K-Nearest Neighbor. Proceedings of the World Congress on Engineering and Computer Science*
- Rahman,Maulana Aditya.Hidayat,Nurul.Supianto,Ahmad Afif.2018.*Komparasi Metode Data Mining K-Nearest Neighbor Dengan Naïve Bayes Untuk Klasifikasi Kualitas Air Bersih*,Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer Vol. 2, No. 12
- Widyadara,Made Ayu Dusea dan Irawan,Roni Heri.2019.*Implementasi Metode Naïve Bayes Dalam Penentuan Tingkat Kesejahteraan Keluarga*.Journal of Computer, information system, & technology management Vol. 2, No. 1.
- Wu X, Kumar V. 2009. *The Top Ten Algorithms in Data mining.* New York: CRC Press.