**Aplikasi Metode *K-Nearest Neighbor* dalam Deteksi Risiko Preeklamsia**

M. Thoriqul Aziz E, Alfian Muhammad Nur, Fashalli Giovi Bilhaq

Program Studi Teknik Biomedis, Fakultas Sains dan Teknologi

Universitas Airlangga Surabaya

**Abstrak**

Preeklamsia merupakan gangguan kehamilan yang ditandai dengan gejala hipertensi dan proteinuria. Hal ini diakibatkan vasopasme dan aktivasi endotel pada usia kehamilan lebih dari 20 minggu. Menurut penelitian penyakit ini dapat disebabkan oleh berbagai faktor diantaranya riwayat hipertensi, faktor keturunan, primigravida dan nulipara, kehamilan berulang kali, riwayat diabetes, ganguan ginjal, serta penambahan berat badan berlebih selama kehamilan yaitu lebih dari 1 kg/minggu. Penyebab utama kematian dari kasus ini sebesar 42% diakibatkan oleh penundaan pencarian bantuan medis, 39% disebabkan oleh pengetahuan terbatas dari pasien mengenai preeklamsia, dan 39% disebabkan oleh kondisi pasien yang tidak tertolong. Penelitian ini bertujuan untuk dapat membantu masyarakat dalam mendeteksi risiko preeklamsia lebih dini sehingga dapat dilakukan tindakan pencegahan guna menurunkan angka kematian ibu akibat preeklamsia. Sampel pada penelitian ini berasal dari RSU Haji Surabaya dengan data sejumlah 110 data yang terdiri dari 55 data normal dan 55 data preeklamsia dengan masing-masing 17 variabel input yang berupa factor risiko sebagai masukan serta 2 kondisi keluaran yaitu normal dan preeklamsia. Penelitian ini dilakukan menggunakan metode klasifikasi *K-nearest neighbor* dengan berdasarkan data yang jaraknya paling dekat dengan objek menggunakan metode pengukuran jarak *Euclidean distance*. Dari data yang dianalisis menggunakan program klasifikasi *K-nearest neighbor* dengan metode pengukuran jarak *Euclidean distance* dihasilkan bahwa nilai k atau jarak tetangga terdekat yang menghasilkan kerja maksimal berada pada jumlah 11 dengan nilai akurasi sebesar 71,43%, nilai spesifisitas sebesar 64,29%, dan nilai sensitivitas sebesar 78,57%.

Kata kunci: preeklamsia, deteksi risiko, k-nearest neighbor, Euclidean distance.

# **PENDAHULUAN**

Preeklamsia merupakan gejala klinis berupa hipertensi dan proteinuria pada kehamilan akibat vasopasme dan aktivasi endotel pada usia kehamilan lebih dari 20 minggu [1]. Di Indonesia, menurut Survey Demografi dan Kesehatan Indonesia (SDKI) pada tahun 2012 tercatat peningkatan signifikan Angka Kematian Ibu (AKI) yaitu 359 per 100.000 Kelahiran Hidup (KH), lebih tinggi dari survey sebelumnya pada tahun 2007. Di Jawa Timur menurut Dinas Kesehatan Jawa Timur pada tahun 2014 terdapat AKI 93,52 per 100.000 Kelahiran Hidup. Penyebab utamanaya adalah meningkatnya kasus preeklamsia/eklamsia dari 26,92% pada tahun 2010 menjadi 34,88% pada tahun 2012 [2].

Faktor Risiko yang menyebabkan meningkatnya pernyakit ini adalah faktor riwayat hipertensi, faktor keturunan, primigravida dan nulipara[2], kehamilan berulang kali, riwayat diabetes, ganguan ginjal, serta penambahan berat badan berlebih selama kehamilan yaitu lebih dari 1 kg/minggu [1]. Menurut penelitian terbaru, umur tidak lagi mempengaruhi preeklamsia secara signifikan jika dibandingkan dengan hipertensi dan faktor keturunan [2].

Dari kasus kematian yang ditimbulkan oleh preeklamsia, penyebab utama sebesar 42% dari kematian diakibatkan oleh penundaan pencarian bantuan medis, 39% disebabkan oleh pengetahuan terbatas dari pasien mengenai preeklamsia, dan 39% disebabkan oleh kondisi pasien yang tidak tertolong [3].

Kasus kematian terbesar yang disebabkan oleh preeklamsia disebabkan terutama karena terlambatnya pencarian bantuan medis dan pengetahuan terbatas pasien terkait preeklamsia. Maka dari itu, diperlukan pengetahuan terkait gejala penyakit serta pendeteksian dini untuk mencegah terjadinya perkembangan penyakit lebih lanjut. Dengan menggunakan pendekatan teknologi, diharapkan ibu hamil akan lebih memperhatikan respon kondisi tubuh terhadap kondisi bayi pasca kehamilan. Dengan menggunakan pendekatan teknologi, diharapkan ibu hamil akan lebih memperhatikan respon kondisi tubuh terhadap kondisi bayi pasca kehamilan[4].

Salah satu cara yang dapat dimanfaatkan untuk melakukan pendeteksian otomatis adalah menggunakan teknologi kecerdasan buatan. Salah satu contohnya adalah sistem pakar. Tujuan dari penggunaan sistem pakar ini adalah untuk memindahkan kepakaran dari seorang ahli ke dalam sistem komputer. Sehingga komputer memiliki kemampuan untuk mendeteksi penyakit yang diinginkan. Salah satu metode yang dapat digunakan yaitu *K-Nearest Neighbors* [5]. Metode ini mampu membantu sistem untuk menganalisis data dari kasus-kasus sebelumnya untuk selanjutnya memberikan kesimpulan berdasarkan pendekatan terhadap data yang dianalisis. Algoritma KNN memiliki arsitektur yang sederhana, dimana klasifikasi yang dilakukan oleh sistem didasarkan pada kedekatan antar data. Selain karena sederhana, beberapa kelebihan dari KNN antara lain efektif, memiliki intuitive yang baik, dapat menentukan banyak domain pada klasifikasi data yang cukup besar. Algoritma KNN menentukan jarak data testing terhadap data training, kemudian klasifikasi dilakukan dengan melihat jumlah tetangga terdekat [6]. Dalam kasus ini, Algoritma KNN digunakan sebagai klasifikasi antara data rekam medis pasien preeklamsia dan pasien normal.

Metode sebelumnya yang digunakan dalam mendeteksi penyakit yang sama adalah dengan metode backpropagation. Akan tetapi, menurut studi menyatakan bahwa metode algoritma KNN dalam melakukan identifikasi penyakit dapat memiliki akurasi yang lebih baik dalam jumlah data yang lebih besar dibandingkan dengan metode backpropagation. Selain itu, proses training dari algoritma KNN lebih cepat karena tidak melalui jumlah iterasi sebesar metode backpropagation, sehingga tidak membutuhkan spesifikasi alat yang cukup besar dalam pengolahan data[7].

# **Metode**

**Tahapan Penelitian**

Dalam tahapan penelitian ini dimulai dengan menentukan latar belakang permasalahan, tujuan dari penelitian, serta batasan masalah yang akan dibahas dalam penelitian. Selanjutnya pada tahap kedua, terkait dengan metode pengumpulan data kasus Preeklamsia. Tahap ketiga yaitu penerapan metode K-Nearest Neighbor (KNN) pada aplikasi MATLAB dengan pembagian data training dan testing KNN. Tahap keempat merupakan bagian analisis hasil prediksi KNN. Pada tahap kelima atau terakhir yaitu memberikan hasil kesimpulan dan saran.

Latar Belakang dan Perumusan Masalah

Tujuan Penelitian dan Batasan

Studi Literatur

Pengumpulan Data

Pembagian Data :  
Data Training dan Data Test

Penerapan KNN pada MATLAB

Analisis Hasil Diagnosis

**Persiapan Penelitian**

Kesimpulan dan Saran

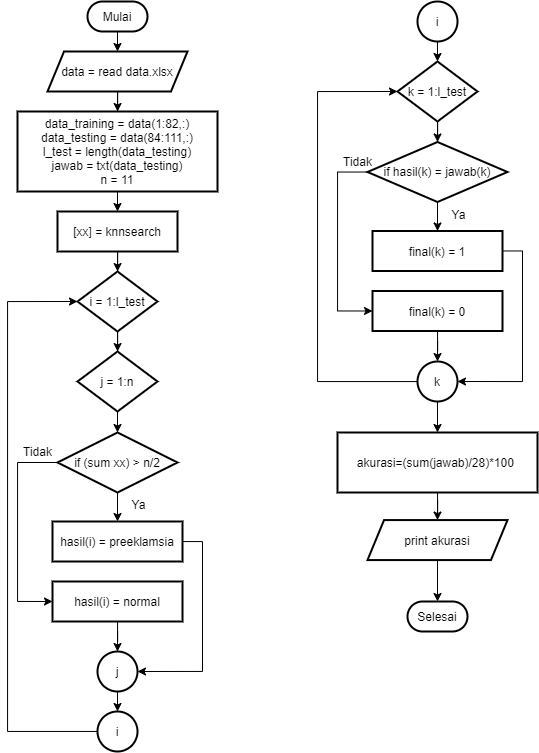
Gambar 1. Tahapan Penelitian

**Metode Pengumpulan Data**

Data Preeklamsia yang digunakan dalam penelitian ini adalah merupakan data rekam medis dari RSU Haji Surabaya. Sehingga, untuk pengumpulan data tidak bersinggungan dengan pasien secara langsung. Data ini terdiri dari ibu hamil termuda berusia 18 tahun hingga 42 tahun. Data yang didapatkan berjumlah 110 data dengan klasifikasi data pasien yang positif Preeklamsia sebanyak 55 data dan data pasien negatif Preeklamsia atau normal sebanyak 55 data. Dari 110 data tersebut kemudian akan dibagi menjadi data *training* dan data *testing.*

**Metode Diagnosis dengan KNN**

Metode yang digunakan dalam penentuan kondisi pasien adalah metode *K-Nearest Neighbor*S. *K-Nearest Neighbor*s (KNN) merupakan metode untuk menentukan hasil klasifikasi dari sampel uji berdasarkan mayoritas kedekatan kategori dengan menggunakan algoritma supervised [8]. Konsep dari KNN adalah menghitung jarak terdekat dari data yang diuji dengan K-tetangga terdekatnya dari data latih penelitian.



Gambar 2. *Flowchart* Program KNN

Langkah 1 : Pembagian Data

Data keseluruhan sejumlah 110 data dengan klasifikasi 50:50 diagnosis preklamsia dan normal. Kemudian, data dibagi dalam dua kategori yaitu data *training* dan data *testing*. Data *training* dibutuhkan untuk mengembangkan kemampuan program KNN, sedangkan data *testing* digunakan untuk mengevaluasi kinerja dari program KNN sebagai alat diagnosis preeklamsia. Komposisi pembagian data *training* dan data *testing* adalah sebagai berikut :

1. Data *Training* diambil sebanyak 82 dari 110 data. Persentase data *training* sebesar 74,54% dengan perbandingan 50:50 preeklamsia dan normal.
2. Data *Testing* diambil sebanyak 28 dari 110 data. Persentase data *testing* sebesar 25,46% dengan perbandingan 50:50 preeklamsia dan normal.

Penentuan data *training* sejumlah 75% dan data *testing* sejumlah 25% didasarkan pada aturan umum yang disebut *“rule of thumb”*[9].

Langkah 2 : Pengolahan data *training* pada program

Data *training* adalah hal mutlak yang dibutuhkan program KNN untuk memiliki kemampuan diagnosis kondisi pasien. Data *training* bisa disebut juga sebagai data pembelajaran. Kemampuan diagnosis KNN didasarkan pada variabel-variabel input yang mempengaruhi kondisi pasien. Penelitian ini menggunakan 17 variabel input yaitu usia, tinggi, BMI, paritas, jumlah kehamilan pada pasangan sekarang, riwayat preeklamsia, riwayat keguguran, riwayat hipertensi, riwayat diabetes mellitus, riwayat penyakit medis (SLE, hipotiroid), riwayat keluarga hipertensi, kehamilan kembar, lama menikah, riwayat antenatal care, riwayat infeksi saluran kencing selama kehamilan, anemia, dan jarak kehamilan.

Dari variabel-variabel input tersebut, algoritma *K-nearest neighbors* melakukan klasifikasi terhadap objek berdasarkan data yang jaraknya paling dekat dengan objek tersebut. Ada beragam cara untuk mengukur jarak antara data baru dengan data lama (data *training)* diantaranya yang paling sering digunakan adalah *Euclidean distance* [10]*.* Rumus untuk *Euclidean distance* adalah :

Persamaan diatas untuk menghitung kedekatan antara dua buah variabel. Untuk perhitungan dengan lebih dari dua variabel yang dalam penelitian ini berjumlah 17 variabel, dapat memodifikasi persamaan menjadi sebagai berikut :

Langkah 3 : *Testing* KNN

Setelah program memperoleh kemampuan untuk mendiagnosis kondisi pasien, selanjutnya dilakukan evaluasi terhadap kinerja dari program dan mencari jumlah k atau tetangga yang menghasilkan nilai akurasi paling tinggi.

Untuk setiap diagnosis pasien preeklamsia ditandai dengan angka 1 dan untuk diagnosis pasien normal diberikan angka 0. Dalam mengevaluasi program dibagi dalam 4 kategori penilaian yaitu :

1. *True Positive* (TP) adalah penilaian untuk kategori pasien dengan diagnosis awal preeklamsia dan saat evaluasi menggunakan program dihasilkan diagnosis preeklamsia.
2. *False Positive* (FP) adalah penilaian untuk kategori pasien dengan diagnosis awal preeklamsia dan saat evaluasi menggunakan program dihasilkan diagnosis normal.
3. *True Negative* (TN) adalah penilaian untuk kategori pasien dengan diagnosis awal normal dan saat evaluasi menggunakan program dihasilkan diagnosis preeklamsia.
4. *False Negative* (FN) adalah penilaian untuk kategori pasien dengan diagnosis awal preeklamsia dan saat evaluasi menggunakan program dihasilkan diagnosis preeklamsia.

Selanjutnya untuk menghitung nilai akurasi, sensitivitas, dan spesifisitas adalah sebagai berikut :

1. Pengukuran akurasi bertujuan untuk mengetahui jumlah orang yang mendapatkan hasil tes positif.
2. Pengukuran sensitivitas bertujuan untuk mengetahui jumlah orang yang benar dinyatakan terkena penyakit dengan hasil tes positif.
3. Pengukuran spesifisitas bertujuan untuk mengetahui jumlah orang yang benar dinyatakan tidak terkena penyakit dengan hasil tes negatif.

# **Hasil dan Pembahasan**

Pada penelitian ini, digunakan metode *K-Nearest Neighbor* untuk melakukan penarikan kesimpulan deteksi preeklamsia ke dalam 2 kategori, yaitu memiliki risiko preeklamsia dan normal. Digunakan 17 variabel untuk mempertimbangkan keputusan yang diambil. Pada penelitian ini, digunakan 110 data yang terdiri dari 55 pasien non preeklamsia dan 55 pasien preeklamsia. Sebanyak 75% data pasien yang digunakan yaitu sebesar 82 data digunakan sebagai data training, sementara 28 data digunakan sebagai data testing.

Pada proses perhitungan menggunakan perangkat lunak MATLAB, digunakan metode perhitungan jarak antar data dengan metode *Euclidean Distance*. Dihitung 11 buah tetangga terdekat (Nearest Neighbor) dari tiap-tiap data testing yang diuji. Data yang telah diuji kemudian didata dari 11 buah tetangga terdekat diberi nilai kategori yaitu risiko preeklamsia dan normal. Dari 11 buah data tetangga terdekat, dihitung mayoritas dari kategori yang dimiliki dan mayoritas dari data tersebut dijadikan acuan untuk mengambil keputusan kategorisasi data testing yaitu memiliki risiko preeklamsia atau normal.

Data testing yang telah diuji kemudian ditentukan memiliki kategori penilaian berupa True Positive, True Negative, False Positive, dan False Negative. Dari pemberian kategori penilaian tersebut, kemudian dihitung nilai akurasi, sensitivitas, dan spesifisitas. Dari penelitian yang dilakukan, didapatkan nilai akurasi sebesar 71.43%, nilai spesifisitas sebesar 64.29%, dan nilai sensitivitas sebesar 78.57%

# **Kesimpulan**

Hasil dari penelitian ini menunjukkan bahwa dengan menggunakan metode klasifikasi ­*K-nearest neighbor* berdasarkan pada pengambilan tetangga terdekat metode *Euclidean distance* memmberikan hasil yang cukup baik. Dengan mengambil jumlah tetangga terdekat atau k bernilai 11, didapatkan nilai akurasi sebesar 71,43%, nilai spresifisitas 64,29%, dan nilai sensivitas sebesar 78,57%. Secara keseluruhan nilai k yang diambil bisa beragam, namun pada penelitian ini dibatasi hanya mengambil satu nilai yang memberikan hasil paling maksimal dibanding nilai lainnya.

**DAFTAR PUSTAKA**

[1] O. Denantika, J. Serudji, and G. Revilla, “Hubungan Status Gravida dan Usia Ibu terhadap Kejadian Preeklampsi di RSUP Dr. M. Djamil Padang,” *J. Kesehat. Masy.*, vol. 4, no. 1, pp. 212–217, 2015.

[2] L. Fatmawati, A. Sulistyono, and H. B. Notobroto, “Pengaruh Status Kesehatan Ibu Terhadap Derajat Preeklampsia / Eklampsia Di Kabupaten Gresik,” *Bul. Penelit. Sist. Kesehat.*, vol. 20, no. 2, pp. 52–58, 2017.

[3] “Maternal Mortality Largely Preventable, Large Study Shows.” [Online]. Available: https://www.medscape.com/viewarticle/841357. [Accessed: 24-Feb-2020].

[4] Purwanti, Endah, Ichroom Septa Preswari, and Ernawati Ernawati. "Early Risk Detection of Pre-eclampsia for Pregnant women using Artificial Neural Network." *International Journal of Online and Biomedical Engineering (iJOE)* 15.02 (2019): 71-80.

[5] P. S. Ramadhan, Mengenal Metode Sistem Pakar, 1st ed. Medan: Penerbit Uwais, 2018.

[6] Imandoust, Sadegh Bafandeh, and Mohammad Bolandraftar. "Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background." *International Journal of Engineering Research and Applications* 3.5 (2013): 605-610.

[7] Redjeki, Sri. "Perbandingan Algoritma Backpropagation dan K-Nearest Neighbor (K-NN) untuk Identifikasi Penyakit." *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*. Vol. 1. No. 1. 2013.

[8] J. Han, M. Kamber, "Data Mining: Concepts and Techniques, *Morgan Kaufmann Publisher,* Microsoft Research, 2006.

[9] "Training dan test set." [Online]. Available : https://www.megabagus.id/training-set-test-set. [Accessed:24-Feb-2020]

[10] Bramer, M., 2013. Principles of data mining. London.: Springer

[11] Ryadi, ALS. 2016. Ilmu Kesehatan Masyarakat. Yogyakarta; Penerbit Andi

**Lampiran**

List Code MATLAB

clc; clear; close all;

%-----------------------Input Data-------------------------------%

data = xlsread('Data.xlsx','Gab');

[num,txt,raw] = xlsread('Data.xlsx','Gab');

training = data(1:82,:); %data training

testing = data(84:111,:); %data testing

answer = txt(85:112,19); %data jawaban

[m1,n1] = size(training);

[m2,n2] = size(testing);

%----------------------Komputasi dengan KNN----------------------%

neighbor = 11; %jumlah tetangga

Mdl = createns(training,'Distance','euclidean');

KNN = knnsearch(Mdl,testing,'K',neighbor);

%-----------------Labelling/Kategorisasi Data--------------------%

% Pemberian Kode hasil perhitungan (n-hasil terdekat)

result=[];

for i=1:m2

for j=1:neighbor

if (KNN(i,j)>=1)&&(KNN(i,j)<=42)

result(i,j)=1; %PE=1

else

result(i,j)=0; %normal=0

end

end

end

% Penentuan Kategori berdasarkan hasil terbanyak

banding = int16(neighbor/2);

for i=1:m2

if (sum(result(i,:))>=banding) %klasifikasi

final(i)="preeklamsia";

else

final(i)="normal";

end

end

final=final';

% Pencocokan Hasil perhitungan dengan Realita

for i=1:m2

if (final(i)==answer(i))

jawab(i)=1;

else

jawab(i)=0;

end

end

jawab=jawab';

% Penentuan TP, FP, FN, TN

for i=1:m2

if (final(i) == answer(i))

switch final(i)

case 'preeklamsia'

status(i) = "TP";

case 'normal'

status(i) = "TN";

end

else

switch final(i)

case 'preeklamsia'

status(i) = "FP";

case 'normal'

status(i) = "FN";

end

end

end

% Penentuan Nilai

TP = nnz(strcmp(status,'TP'));

TN = nnz(strcmp(status,'TN'));

FP = nnz(strcmp(status,'FP'));

FN = nnz(strcmp(status,'FN'));

akurasi=(sum(jawab)/m2)\*100

sensitivitas = (TP/(TP+FN))\*100

spesifitas = (TN/(TN+FP))\*100