

Proposal Penelitian : Deteksi Dini Penyakit Ginjal Kronis menggunakan *Extreme Learning Machine* berbasis *Neural Network*

Moh. Faturrahman

NPM : 1506706276

Fakultas Ilmu Komputer, Universitas Indonesia.

moh.faturrahman@ui.ac.id

Abstrak

Prevalensi penyakit Ginjal Kronis sebesar 0.2 % terjadi di Indonesia dan cenderung terjadi pada masyarakat pedesaan, tidak bersekolah, pekerja swasta, dan petani. Penyakit Ginjal Kronis dapat meningkatkan resiko terjadinya komplikasi terhadap penyakit lainnya seperti penyakit jantung, hipertensi, anemia, kematian, dan proses menuju pada stadium akhir. Deteksi dini pada penyakit ini untuk menunda proses penurunan fungsi ginjal, mencegah atau mengatur komplikasi. Penggunaan teknik Pembelajaran Mesin Extreme Learning Machine (ELM) berpotensi melakukan deteksi dini penyakit Ginjal Kronis secara cepat dan efisien untuk mendukung tujuan diatas. Keunggulan dari teknik ini adalah memiliki proses pembelajaran yang sangat cepat, tidak memiliki banyak parameter yang harus diinisialisasi, tidak terjebak pada local minima, dapat menggunakan non-differentiable activation function. Tujuan dari penelitian ini mengukur performa Accuracy, Sensitivity, Specificity dan menganalisa Algoritma Extreme Learning Machine dalam mendeteksi penyakit Ginjal Kronis.

Kata Kunci : Penyakit Ginjal Kronis, Pembelajaran Mesin, Extreme Learning Machine

I. PENDAHULUAN

Penyakit Ginjal Kronis didefinisikan sebagai adanya tanda-tanda kerusakan pada ginjal (abnormalitas dalam darah, urin, atau pada tes gambar) lebih dari 3 bulan atau menurunnya *Glomerular Filtration Rate* $< 60 \text{ mL/min/1.73 m}^2$ [1]. Prevalensi penyakit Ginjal Kronis sebesar 0.2 % terjadi di Indonesia dan cenderung terjadi pada masyarakat pedesaan, tidak bersekolah, pekerja swasta, dan petani [2]. Penyakit Ginjal

Kronis meningkatkan resiko komplikasi terhadap penyakit lainnya seperti penyakit jantung, hipertensi, anemia, kematian, dan proses menuju pada stadium akhir [3]. Deteksi Dini penyakit Ginjal dapat dilakukan pemeriksaan Laboratorium dengan estimasi nilai *Glomerular Filtration Rate* dan nilai *Albuminuria*, serta menggunakan beberapa tes pendukung seperti *Ultra Sound/CT Scan* pada saluran kemih, atau tes serologic jika penderita punya riwayat diabetes atau hipertensi. Deteksi Dini bertujuan untuk menunda proses penurunan fungsi ginjal, mencegah atau mengatur komplikasi.

Pemeriksaan laboratorium memiliki kelemahan seperti biaya yang mahal dan proses yang berulang-ulang jika hasil yang didapat perlu dikonfirmasi. Penggunaan teknik Pembelajaran Mesin berpotensi melakukan deteksi dini penyakit Ginjal Kronis untuk mendukung tujuan diatas.

Pemanfaatan teknik Pembelajaran Mesin dalam bidang medis telah banyak dilakukan untuk membantu petugas medis dalam mengambil keputusan, memilih perawatan yang sesuai dengan pasien, membantu merancang pengobatan. Guang Jing et al, menggunakan ELM untuk memprediksi mortality rate dari pasien kanker kandung kemih yang telah dilakukan *Radical Cystectomy* (Operasi pengangkatan saluran kandung kemih) dengan akurasi rata-rata 80%, teknik ini menunjukkan performa yang lebih baik dibandingkan dengan metode statistik [4]. Penelitian serupa juga dilakukan oleh Vukicevic et al, dengan menggunakan *Neural Network* dan dataset yang berbeda, hasil yang dicapai memiliki performa akurasi 92.5% [5].

Khusus pada penyakit Ginjal Kronis, penggunaan Teknik Machine Learning Naive Bayes dalam mendeteksi penyakit Ginjal Kronis sudah dilakukan oleh Chetty et al, dengan menggunakan algoritma seleksi fitur pada pra-pemrosesan data, proses deteksi mampu menghasilkan akurasi 99%[6]. Dengan menggunakan dataset yang sama, Chen Zewei et al, menggunakan *Fuzzy Classifier* untuk mendeteksi penyakit Ginjal Akut, nilai rata-rata dari akurasi yang didapat mencapai 98% [7]. Berdasarkan beberapa hasil penelitian pemanfaatan teknik Pembelajaran Mesin pada bidang medis, menunjukkan bahwa teknik Pembelajaran Mesin merupakan salah satu solusi efektif yang dapat mengatasi masalah-masalah pada bidang Medis.

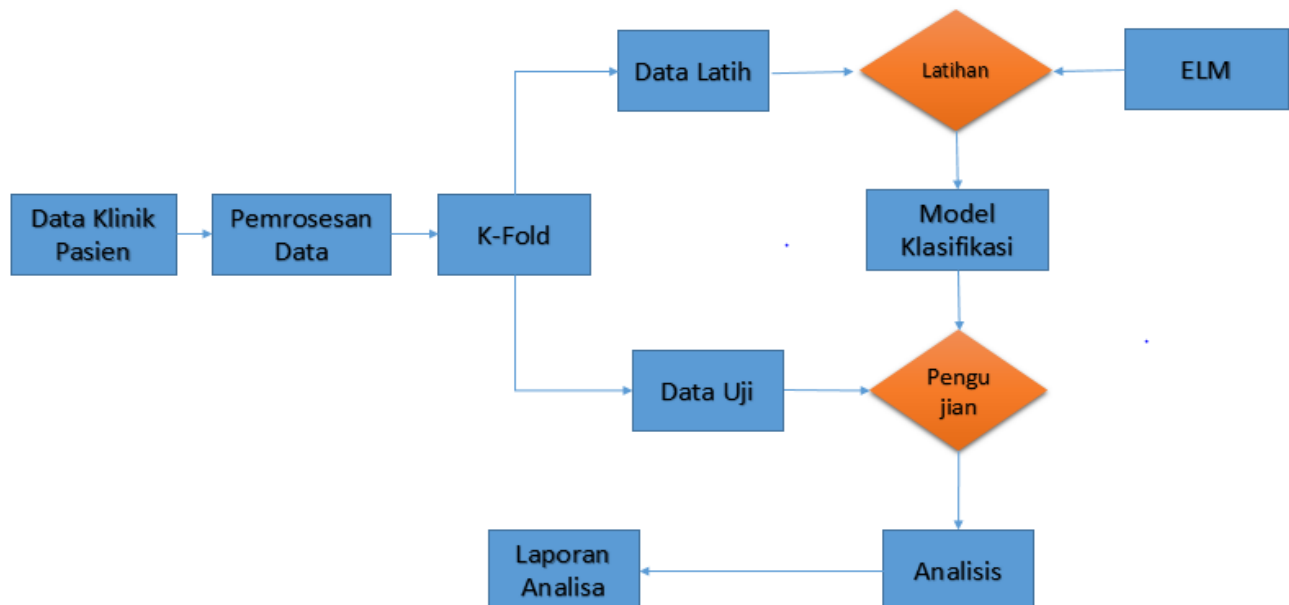
Salah satu Teknik Machine Learning yang dapat digunakan pada bidang medis adalah Extreme Learning Machine [8], [9] [10], yang merupakan variasi dari Single Layer Feedforward Network (*SLFN*). Keunggulan dari teknik ini

adalah memiliki proses pembelajaran yang sangat cepat, tidak memiliki banyak parameter yang harus diinisialisasi, tidak terjebak pada *local minima*, dapat menggunakan *non-differentiable activation function* [11].

Oleh karena itu tujuan dari penelitian ini mengukur performa dari ELM dalam mendeteksi penyakit Ginjal Kronis. Adapun penjelasan mengenai ELM akan dijelaskan pada bab selanjutnya.

Pada rancangan proposal penelitian ini ditulis melalui alur penulisan sebagai berikut, BAB II membahas skema penelitian yang akan dilakukan, mencakup tahapan-tahapan penelitian. BAB III, membahas teknik Pembelajaran Mesin yang digunakan, data klinik pasien, Perangkat Keras dan Perangkat Lunak yang digunakan. BAB IV membahas kesimpulan.

II. METODE



Gambar 1 : Alur Penelitian

Alur dari penelitian ini dapat dilihat pada gambar 1 diatas.

1. Tahap pertama dengan mengumpulkan Data Klinik Pasien, yang dapat diunduh pada *UCI Machine Learning Repository*
2. Tahap kedua, melakukan pra-pemrosesan data, yaitu mengisi nilai missing values

3. Tahap ketiga, melakukan pembagian data latih dan data uji menggunakan metode *K-Fold Crossvalidation* dengan menggunakan nilai rata-rata pada variabel *continuous* dan nilai mode pada variabel *discrete*, melakukan normalisasi [0,1] pada variabel yang memiliki nilai *continous*

nilai K= 10. Data Latih selanjutnya digunakan untuk melatih algoritma ELM dalam mengenali dan memprediksi data pasien pada penyakit Ginjal Akut. Data Uji digunakan untuk melakukan validasi dan mengukur performa dalam memprediksi penyakit Ginjal Akut pada pasien. Mekanisme pengujian yang digunakan akan dijelaskan pada bab selanjutnya.

4. Proses Latihan, dimana Algoritma ELM akan belajar mengenali pola data untuk memprediksi apakah suatu pasien mengidap penyakit Ginjal Akut atau tidak, proses pelatihan ini juga mencari konfigurasi parameter terbaik dalam memprediksi penyakit Ginjal Akut. Parameter yang akan diubah konfigurasinya adalah, fungsi aktivasi dan jumlah *Hidden Neuron* pada ELM. Hasil keluaran dari proses ini adalah sebuah model klasifikasi ELM yang memprediksi penyakit Ginjal Akut pada pasien.
5. Proses Pengujian, melakukan uji coba terhadap model ELM yang didapat pada proses latihan dengan memberikan data uji untuk diprediksi, data uji merupakan data yang tidak diketahui sebelumnya pada proses latihan. Algoritma prediksi Support Vector Machine (SVM), Backpropagation Neural Network (BPN), akan digunakan sebagai pembanding terhadap model ELM.
6. Analisa, Tahap analisa yaitu membandingkan hasil eksperimen pengujian model ELM, SVM dan BPN, terhadap nilai Accuracy, Sensitivity, Specifisity, melakukan analisa mendalam temuan-temuan yang terjadi saat proses Eksperimen. Menyajikan hasil Eksperimen dalam bentuk grafik agar mudah dianalisa secara visual.
7. Laporan Analisa, adalah proses penulisan laporan akhir penelitian ditulis dalam format publikasi *paper*, yang terdiri dari Abstrak, Pendahuluan, Metode, Eksperimen, Hasil Eksperiment, Kesimpulan dan Saran.

III. EKSPERIMEN

1. Algoritma

Extreme Learning Machine

Extreme Learning Machine (ELM) merupakan salah satu teknik Machine Learning untuk pengenalan pola, proses pembelajaran dalam

ELM menggunakan solusi *Moore Penrose Inverse* dan *Smallest Norm Least-Square*. Berdasarkan [11] Dengan memodifikasi proses pembelajaran pada *SLFN* (*Single Layer Feed Forward Neural Network*) dengan *L hidden neurons*, dimana $L < N$ jumlah data pada proses latih. Untuk N data (x_i, t_i) , dimana $x_i = [x_{i1}, x_{i2}, \dots, x_{im}]^T \in R^N$ dan $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in R^m$. SLFN dapat dinotasikan sebagai berikut

$$\sum_{i=1}^L \beta_i g(w_i \cdot x_i + b_i) = o_j, j = 1, \dots, N \quad (1)$$

w_i adalah bobot yang menghubungkan *Input Layer* dan *Hidden Layer*. $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ adalah bobot yang menghubungkan *Hidden Layer* dengan *Output Layer*. Sedangkan $w_i \cdot x_i$ adalah *inner product* dari $w_i \cdot x_i$ dengan asumsi, output layer adalah linear.

SLFN dengan *L Hidden Neurons* menggunakan fungsi aktivasi $g(x)$ dapat mendekati setiap data dengan meminimalkan *error* $\sum_{j=1}^N \|O_j - t_j\| = 0$, dikarenakan terdapat β_i , w_i , dan b_i yang dapat dinotasikan

$$\sum_{i=1}^L \beta_i g(w_i \cdot x_i + b_i) = t_j, j = 1, \dots, N \quad (2)$$

Pendekatan ELM diatas dapat ditulis sebagai berikut

$$H\beta = T \quad (3)$$

Dimana

$$H(w_1, \dots, w_L, b_1, \dots, b_L, x_1, \dots, x_N) = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \dots & g(w_1 \cdot x_N + b_1) \\ \dots & \dots & \dots \\ g(w_L \cdot x_1 + b_L) & \dots & g(w_L \cdot x_N + b_L) \end{bmatrix}_{N \times L} \quad (4)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \dots \\ \beta_L^T \end{bmatrix}_{L \times m} \text{ dan } T = \begin{bmatrix} t_1^T \\ \dots \\ t_N^T \end{bmatrix}_{N \times m} \quad (5)$$

H merupakan *Hidden Layer Output Matrix* dari *Neural Network*, kolom ke- i pada matrix H

merupakan hasil output vektor dari input x_1, x_2, \dots, x_N .

Secara singkat algoritma pembelajaran untuk SLFN Extreme Learning Machine dapat dilakukan dalam tiga tahap :

Diberikan Data latih $N = \{(\mathbf{x}_i, \mathbf{t}_i) | \mathbf{x}_i \in \mathbf{R}^n, \mathbf{t}_i \in \mathbf{R}^m, i = 1, \dots, N\}$, fungsi aktivasi $g(x)$, dan jumlah *hidden neuron* L .

1. Inisialisasi nilai bobot \mathbf{w}_i dan bias b_i , $i = 1, \dots, N$.
2. Menghitung keluaran pada Hidden Layer dari matrix \mathbf{H}
3. Menghitung nilai dari bobot $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} = \mathbf{H}^+ \mathbf{T} \quad (6)$$

Dimana \mathbf{H} , $\boldsymbol{\beta}$ dan \mathbf{T} , dapat dilihat pada persamaan 4 dan 5.

2. Dataset

Dataset klinis penyakit Ginjal Kronis yang digunakan pada penelitian ini diambil dari *UCI Machine Learning Repository* yang tersedia secara online. Data yang dikumpulkan sebanyak 400 data, dengan jumlah atribut 25 termasuk atribut kelas. Terdapat 2 kelas pada dataset, “ckd” dan “notckd” (pasien dengan positif penyakit Ginjal Akut, dan pasien dengan negatif penyakit Ginjal Akut) dengan total jumlah “237” ckd dan 149 “notckd”, setiap atribut terdapat *missing values*. Detail dari dataset yang digunakan dapat dilihat pada tabel 1.

Tabel 1 : Keterangan Dataset Ginjal Akut

Input	Nama Atribut	Tipe data	Nilai/Satuan	Jumlah Missing Values
1	Age	Numerical	Normalisasi [0,1]	9
2	Blood Pressure	Numerical	Normalisasi [0,1]	9
3	Specific Gravity	Nominal	(1.005,1.010, 1.015,1 .020,1.025)	45
4	Albumin	Nominal	0,1,2,3,4,5	44
5	Sugar	Nominal	0,1,2,3,4,5	46
6	Red Blood Cell	Nominal	Normal, Abnormal	143
7	Pus Cell	Nominal	Normal, Abnormal	62
8	Pus Cell Clumps	Nominal	Present, Not Present	4

9	Bacteria	Nominal	Present, Not Present	4
	Blood			
10	Glucose	Numerical	mgs/dl	42
	Random			
11	Blood Urea	Numerical	mgs/dl	18
	Serum			
12	Creatinine	Numerical	mgs/dl	16
13	Sodium	Numerical	mEq/L	81
14	Potassium	Numerical	mEq/L	82
15	Hemoglobin	Numerical	gms	48
16	Packed Cell Volume	Numerical	Normalisasi [0,1]	66
17	White Blood Cell Count	Numerical	cells/cmm	99
18	Red Blood Cell Count	Numerical	million/cmm	123
19	Hypertension	Nominal	Yes, No	2
20	Diabetes Melitus	Nominal	Yes, No	2
	Coronary			
21	Artery Disease	Nominal	Yes, No	2
22	Appetite	Nominal	Good, Poor	1
23	Pedal Edema	Nominal	Yes, No	1
24	Anemia	Nominal	Yes, No	1
	Class			
25	(Output)	Nominal	ckd,notckd	0

3. K-Fold Crossvalidation

K-Fold Crossvalidation merupakan salah satu metode dalam membagi dataset menjadi data latih dan data uji, sebanyak k data. Misalkan dengan memilih nilai $k = 10$, maka dataset akan dibagi menjadi 10 subset data uji dan data latih yang tidak beririsan. Dengan Proporsi 75 % untuk data latih, 25 % untuk data uji.

4. Evaluasi

Nilai yang akan diprediksi berdasarkan data klinis pasien adalah “ckd” dan “notckd”, “ckd” merujuk pada pasien yang terdiagnosis Ginjal Akut, sedangkan “notckd” merujuk pada pasien yang tidak didiagnosis Ginjal Akut. Tujuan model prediksi ELM adalah memperkecil kesalahan dalam memprediksi apakah pasien termasuk dalam kategori “ckd” atau “notckd”. Untuk mengetahui performa dari model ELM dilakukan proses evaluasi Algoritma, yang diukur dengan menghitung nilai Akurasi, Sensitivity, Spesificity dengan menggunakan formula dibawah ini.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (8)$$

$$Spesificity = \frac{TN}{TN + FP} \quad (9)$$

TP adalah True Positif yang merujuk pada data pasien penyakit Ginjal Akut yang diprediksi secara benar. TN adalah True Negatif yang merujuk pada data pasien yang tidak mengidap Ginjal Akut yang diprediksi secara benar. FP adalah False Positive yang merujuk pada pasien yang tidak mengidap Ginjal Akut, tetapi diprediksi mengidap Ginjal Akut. FN adalah False Negative yang merujuk pada data pasien Ginjal Akut tetapi diprediksi tidak mengidap penyakit Ginjal Akut.

5. Perangkat Pendukung

Seluruh Eksperimen pada penelitian ini menggunakan satu perangkat Laptop dengan Processor Core i7 5th, RAM 4.GB, Windows Operating System 64 Bit. *Source Code* program ditulis menggunakan bahasa pemrograman R, dengan menggunakan R Studio sebagai IDE. Untuk *Source Code* ELM, SVM, dan BPN menggunakan library package caret yang tersedia *online* pada R Studio.

IV. KESIMPULAN

Penyakit Ginjal Akut mengakibatkan pasien memiliki resiko tinggi terkena komplikasi terhadap penyakit lain. Dengan melakukan deteksi dini, resiko terjadinya komplikasi dapat diperkecil, dengan memberikan perawatan untuk menghambat perkembangan penyakit Ginjal. Metode Pembelajaran Extreme Learning Machine akan digunakan untuk mendeteksi penyakit Ginjal Akut secara dini berdasarkan data klinis pasien, untuk mendukung tenaga medis dalam mengambil keputusan atau tindakan terhadap pasien. Model ELM diharapkan dapat mengatasi kelemahan-kelemahan pada pemeriksaan penyakit Ginjal Akut seperti biaya yang mahal, proses yang berulang-ulang, serta efisien.

Daftar Pustaka

- [1] "Chronic kidney disease - ClinicalKey." [Online]. Available: https://remote-lib.ui.ac.id:2194/#!/content/medical_topi c/21-s2.0-1014826. [Accessed: 25-Mar-2016].
- [2] Ministry of Health, "Riset Kesehatan Dasar (Basic Health Research)," p. 303, 2013.
- [3] J. A. Vassalotti, R. Centor, B. J. Turner, R. C. Greer, M. Choi, and T. D. Sequist, "A Practical Approach to Detection and Management of Chronic Kidney Disease for the Primary Care Clinician.," *Am. J. Med.*, vol. 129, no. 2, pp. 153–162.e7, 2015.
- [4] G. Wang, K.-M. Lam, Z. Deng, and K.-S. Choi, "Prediction of mortality after radical cystectomy for bladder cancer by machine learning techniques," *Comput. Biol. Med.*, vol. 63, pp. 124–132, 2015.
- [5] A. M. Vukicevic, G. R. Jovicic, M. M. Stojadinovic, R. I. Prelevic, and N. D. Filipovic, "Evolutionary assembled neural networks for making medical decisions with minimal regret: Application for predicting advanced bladder cancer outcome," *Expert Syst. Appl.*, vol. 41, no. 18, pp. 8092–8100, 2014.
- [6] N. Chetty, "Role of Attributes Selection in Classification of Chronic Kidney Disease Patients," pp. 0–5, 2015.
- [7] Z. Chen, Z. Zhang, R. Zhu, Y. Xiang, and P. B. Harrington, "Diagnosis of patients with chronic kidney disease by using two fuzzy classifiers," *Chemom. Intell. Lab. Syst.*, 2016.
- [8] Z. You, L. Li, Z. Ji, M. Li, and S. Guo, "Prediction of protein-protein interactions from amino acid sequences using extreme learning machine combined with auto covariance descriptor," *Memetic Comput. (MC)*, 2013 ..., pp. 80–85, 2013.
- [9] M. Fathurachman and U. Kalsum, "Heart disease diagnosis using extreme learning based neural networks," ... *Concept, Theory* ..., 2014.
- [10] S. Ismaeel, A. Miri, and D. Chourishi, "Using the Extreme Learning Machine (ELM) Technique for Heart Disease Diagnosis," no. 1, pp. 1–3, 2015.
- [11] G.-B. H. G.-B. Huang, Q.-Y. Z. Q.-Y. Zhu, and C.-K. S. C.-K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," *2004 IEEE Int. Jt. Conf. Neural Networks (IEEE Cat. No.04CH37541)*, vol. 2, pp. 25–29, 2004.