

# Penilaian Risiko Induksi Kehamilan Hipertensi Menggunakan Pendekatan Machine Learning

Sirinat Wanriko

Narit Hnoohom\*

Konlakom Wongpatikaseree

Informasi Gambar dan Kecerdasan  
Laboratorium, Jurusan Komputer  
Teknik, Fakultas Teknik,  
Universitas Mahidol

Gambar Laboratorium Informasi dan  
Kecerdasan, Jurusan Teknik Komputer,  
Fakultas Teknik, Mahidol University Nakhon  
Pathom, Thailand Nakhon

Departemen Teknik Komputer,  
Fakultas Teknik, Mahidol  
Universitas

Pathom, konlakorn Thailand. won@mahidol. ac. th sirinat. wan@student.mahidol. ac .th Penulis korespondensi:

narit. hno@mahidol. ac. th

Anuchit Jitpattanukul

Olarik Musigavong

Dinamis Cerdas dan Nonlinier  
Pusat Riset Inovasi,  
Jurusan M Matematika, Fakultas  
Sains Terapan, Raja Mongkut  
Universitas Teknologi Utara  
Bangkok  
Bangkok, Thailand  
anuchit.j @sci.kmutnb.ac.th

Departemen Obstetri dan Ginekologi  
(Endokrinologi Reproduksi dan  
Infertilitas)  
Rumah Sakit Chaophraya Abhaibhubejhr  
Prachin Buri, Thailand  
dr.olarik@gmail.com

pengobatan dapat mengurangi keparahan dan bahaya. Data dari  
Organisasi Kesehatan Dunia (WHO)

**Abstrak**— Penelitian ini bertujuan untuk mengembangkan model prediktif penilaian risiko hipertensi akibat kehamilan dengan menggunakan pendekatan pembelajaran mesin. Hipertensi yang diinduksi kehamilan merupakan komplikasi yang berdampak serius pada ibu hamil dan janin. Ini adalah tiga penyebab kematian tertinggi di dunia pada wanita hamil [1]. Saat ini, penyebab pasti hipertensi akibat kehamilan tidak diketahui dan oleh karena itu tidak dapat dicegah. Deteksi dini dan pengobatan yang diterima dapat mengurangi keparahan dan bahayanya. Dataset publik Logan (2020) digunakan dalam penelitian ini [2]. Kumpulan data dikumpulkan dari studi kasus-kontrol pada determinan 83 preeklamsia dan lima kasus eklamsia di antara 352 wanita hamil yang melahirkan di rumah sakit daerah di Nairobi, Kenya. Menurut dataset, 75 persen wanita hamil dan 25 persen ibu hamil yang mengalami pre-eklamsia dan eklamsia. Dengan demikian, ini akan menghasilkan masalah klasifikasi yang tidak seimbang ketika salah satu dari dua kelas memiliki lebih banyak data daripada kelas lainnya. Dengan demikian, masalah ini diselesaikan dengan Teknik Pengambilan Sampel Minoritas Sintetis (SMOTE). Penilaian risiko hipertensi akibat kehamilan dilakukan pada tujuh algoritma pembelajaran mesin, yaitu regresi logistik (LR), K-nearest neighbor (KNN), pohon keputusan (DT), hutan acak (RF), jaringan saraf perceptron multilayer (MLP), mendukung mesin vektor (SVM), dan naif Bayes (NB). Pada hasil percobaan, RF memiliki akurasi tertinggi sebesar 89,62 persen dibandingkan dengan algoritma pembelajaran mesin lainnya.

**Kata kunci** - Hipertensi yang diinduksi kehamilan, pembelajaran mesin, klasifikasi tidak seimbang, teknik oversampling minoritas sintetis

I. Sebagai pengantar io n

Hipertensi yang diinduksi kehamilan merupakan komplikasi yang berdampak serius pada ibu hamil dan janin [2]. Ini sebagian besar terjadi setelah minggu kedua puluh kehamilan. Selain kematian, ada risiko komplikasi serius; seperti, plasenta prematur, banjir paru, pembekuan darah abnormal, sementara Konferensi Internasional ke-6 tentang Seni

Digital, Media dan Teknologi (DAMT) dan Konferensi Bagian Utara ECTI ke-4 tentang Teknik Listrik, Elektronik, Komputer dan Telekomunikasi (NCON)

atau kebutaan permanen, dan pendarahan di otak. Saat ini, penyebab pasti hipertensi akibat kehamilan tidak diketahui dan oleh karena itu tidak dapat dicegah. Deteksi dini dan diterima

978-1-6654-1569-9/2/ 1/\$31,00 2021 IEEE

pada tahun 2010 ditemukan bahwa 99 persen ibu hamil yang meninggal tinggal di pedesaan dan miskin [3]. Selain itu, di pedesaan Thailand, terdapat kekurangan dokter dan tenaga medis. Akibatnya, beberapa instansi telah mendirikan proyek untuk mengelola masalah pendistribusian layanan medis ke pedesaan, namun masalah ini tetap saja terjadi. Menurut statistik Thailand pada tahun 2017, Dewan Medis menemukan bahwa rata-rata satu dokter harus merawat 1.143 pasien per tahun [4-5]. Dengan demikian, tujuan akhir dari penelitian ini adalah untuk mengembangkan model prediktif penilaian risiko hipertensi akibat kehamilan menggunakan pendekatan pembelajaran mesin. Tujuannya adalah untuk mendukung dokter dalam penilaian awal pasien, perencanaan perawatan, dan perawatan pemulihan.

Sisa kertas disusun sebagai berikut. Pertama, pengenalan singkat diberikan. Pekerjaan terkait dijelaskan dalam Bagian II. Bagian III berisi ikhtisar metode yang diusulkan. Hasil percobaan dimasukkan dalam Bagian IV. Akhirnya, dalam Bagian V, kesimpulan dibahas.

## II. Pekerjaan Terkait

Saat ini ada beberapa penelitian, yang menggunakan pembelajaran mesin untuk membuat model prediktif risiko hipertensi akibat kehamilan; seperti eklamsia dan preeklamsia. Tahir dkk. [6] mengusulkan prediksi risiko tingkat preeklamsia pada ibu hamil selama proses kehamilan menggunakan algoritma neural network (NN) dan deep learning (DL). Jumlah atribut dikurangi dari 17 menjadi sembilan menggunakan pengoptimalan segerombolan partikel. DL memberikan akurasi paling tinggi dengan 95,68 persen. Nikolaides et al. [7] mengembangkan indikator tahap awal untuk risiko pre-eklamsia. Dataset disusun menggunakan 6.838 kasus ibu hamil di Inggris yang melibatkan 24 variabel. Dataset hanya berisi 116 kasus

ibu hamil dengan preeklamsia. Jaringan saraf multi-slab memberikan hasil paling akurat sebesar 93,8 persen. Selain itu, Tahir et al. [8] mengusulkan deteksi pra

eklampsia dengan menggunakan neural network dibandingkan dengan algoritma lainnya. Dataset pre-eklampsia diambil dari Rumah Sakit Umum Haji Surabaya, Indonesia. Algoritma neural network dan validasi LOO memberikan akurasi paling tinggi dengan 96,66 persen. Demikian juga, Leemaqz et al. [9] menyajikan model prediktif pre-eklampsia berjenjang yang berfokus pada konvergensi beberapa model. Mereka menganalisis bahwa teorema Bayes dapat digunakan untuk mengintegrasikan banyak model. Model terintegrasi memberikan akurasi paling tinggi, di mana 81 persen diidentifikasi secara akurat pada usia kehamilan 20 minggu.

Selain penelitian di atas, pembelajaran mesin telah digunakan untuk membuat model prediksi risiko penyakit lain. Khalia et al. [10] mengusulkan Random Forest (RF) untuk memprediksi risiko delapan penyakit kronis dengan menggunakan database sampel rawat inap nasional dari sampel rumah sakit di Amerika Serikat. Pendekatan pembelajaran ansambel digunakan untuk mengatasi masalah ketidakseimbangan data. Hasil dari pengujian akurasinya adalah 88,79 persen. Selain itu, Pattekari dan Parveen [11] mengusulkan sistem prediksi penyakit jantung.

Mereka menggunakan teknik penambangan yang terdiri dari DT, Naive Bayes (NB), dan NN. Sistem ini secara cerdas dapat menjawab pertanyaan rumit dalam mendiagnosis penyakit jantung dan membantu praktisi medis. Ini lebih lanjut membantu dalam meningkatkan dan mengurangi biaya pengobatan. Selain itu, Akhil jabbar et al. [12] mengusulkan K-nearest neighbor (KNN) dan algoritma genetika untuk menganalisis penyakit jantung menggunakan enam dataset dari UCI Repository (UCI Machine Learning Repository) dan satu dataset dari berbagai rumah sakit di Andhra Pradesh, India. Untuk meningkatkan akurasi penelitian, KNN digunakan untuk mengumpulkan semua kasus dan mengklasifikasikan pasien baru menggunakan algoritme genetika untuk menghitung kesamaan.

Hasil penelitian menunjukkan bahwa penggunaan kedua metode tersebut secara bersama-sama memberikan tingkat akurasi sebesar 95,73 persen. Nayeem dkk. [13] mengusulkan multilayer perceptron neural network (MLP) untuk memprediksi penyakit jantung, penyakit hati, dan kanker paru-paru menggunakan algoritma jaringan saraf backpropagation feed-forward dan MLP untuk membedakan antara individu yang terinfeksi dan tidak terinfeksi. Mereka menggunakan dataset Arrhythmia MIT-BIH, dan hasil penelitian menunjukkan bahwa akurasi prediksi penyakit jantung adalah 82 persen, penyakit hati 82 persen, dan kanker paru-paru masing-masing 91 persen.

AKU AKU AKU Th e METODE yang Diusulkan

Pada bagian ini, proses pengembangan model prediksi risiko hipertensi akibat kehamilan seperti yang ditunjukkan pada Gambar 1. Metode yang diusulkan terdiri dari lima proses: (1) Data kehamilan, (2) data seimbang, (3) persiapan data , (4) pemilihan fitur, dan (5) mengevaluasi algoritma pembelajaran mesin.

D.Pemilihan fitur E.Evaluasi algoritma pembelajaran mesin

Gambar 1 Proses pengembangan metode yang diusulkan

A. Data kehamilan Peneliti

mengeksplorasi kumpulan data publik yang diperoleh terkait dengan risiko hipertensi akibat kehamilan. Dataset publik Logan (2020) digunakan dalam penelitian ini [2]. Kumpulan data dikumpulkan dari studi kasus-kontrol pada 83 kasus pre-eklampsia dan lima kasus eklampsia di antara 352 wanita hamil yang melahirkan di rumah sakit daerah di Nairobi, Kenya. Peneliti memilih 17 atribut yang dicocokkan dengan atribut kesehatan ibu dan anak

handbook di Thailand, dan ini digunakan untuk mencatat data kesehatan ibu dan anak [14]. 17 atribut tersebut terdiri dari usia ibu, usia kehamilan pertama, riwayat keluarga diabetes, riwayat diabetes, riwayat keluarga hipertensi, riwayat hipertensi, kunjungan antenatal care pertama, jumlah kunjungan antenatal care, jarak kehamilan, persalinan sesar, kehamilan multifetal. , graviditas, paritas, provinsi, tempat tinggal, penggunaan alkohol, dan penggunaan tembakau. Peneliti kemudian mengkonversi data untuk kesesuaian dalam proses data yang tidak seimbang.

(a) Data yang tidak seimbang.

(b) Data seimbang.

Gambar 2. Perbandingan data tidak seimbang dan data seimbang.

B. Data yang tidak seimbang

Pada data kehamilan [13], 75 persen ibu hamil dalam keadaan sehat dan hanya 25 persen yang mengalami pre-eklampsia dan eklampsia. Masalah klasifikasi yang tidak seimbang muncul ketika salah satu dari dua kelas memiliki lebih banyak data daripada kelas lainnya, dan data ini mempengaruhi klasifikasi kelas minoritas.

Klasifikasi tipikal efektif ketika setiap kelas data memiliki nomor yang sama, dan prediksi cenderung ke kelas mayoritas.

Untuk membandingkan kinerja data yang tidak seimbang dan pendekatan data yang seimbang, peneliti melakukan algoritma Synthetic Minority Over-sampling Technique (SMOTE) pada dataset [15]. SMOTE bekerja dengan memanfaatkan

the KNN algorithm to construct synthetic data. The SMOTE algorithm by selecting random data from the minority class after adding random noise to it. The SMOTE algorithm is used to generate synthetic data from the minority class. The SMOTE algorithm is used to generate synthetic data from the minority class. The SMOTE algorithm is used to generate synthetic data from the minority class.

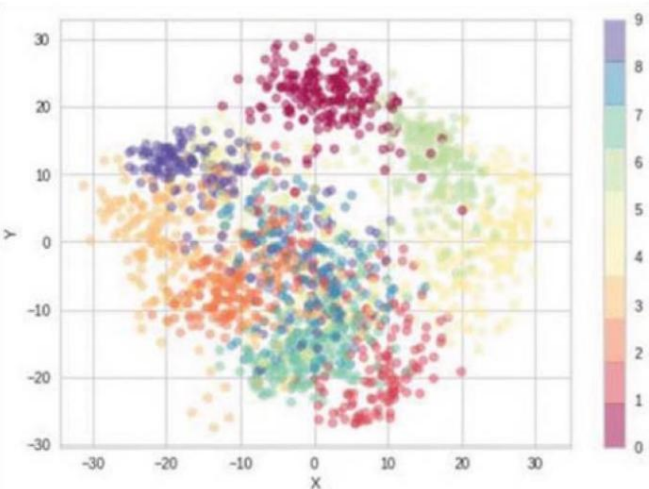
The researcher plotted the class distribution to show the imbalance data in the selected dataset, as shown in Figure 2(a). Figure 2(b) shows the balanced data is handled with the SMOTE.

C. Data preparation

In the data preparation process, the researcher used data transformation techniques to handle the imbalance data. The researcher used data transformation techniques to handle the imbalance data. The researcher used data transformation techniques to handle the imbalance data. The researcher used data transformation techniques to handle the imbalance data. The researcher used data transformation techniques to handle the imbalance data.

D. Feature selection

Feature selection for machine learning is the method of choosing features that are relevant to each other to reduce the number of features to only the features necessary in order to develop a predictive model. Principal component analysis (PCA) is a data reduction technique that is commonly used in combination with linear algebra to minimize the dimensionality of a dataset by compressing a dataset of attributes while preserving most of the information in the dataset. PCA was used to extract the three principal components on the 17 attributes, imbalanced data, and balanced data.



Gambar 3. Pasca-pemrosesan data dengan PCA.

E. Evaluate machine learning algorithms

To evaluate machine learning algorithms on the dataset, the predictive model was performed on seven machine learning algorithms which were LR, KNN, DT, RF, MLP, SVM, and NB.

To estimate the performance of the machine learning algorithm on the unseen data, the researcher first split the dataset into two parts: training and testing. For this case study, the researcher

took a 77:33 ratio keeping 77 percent of the dataset for training and the remaining 33 percent for testing. The researcher trained a machine learning algorithm on the first part then evaluated the predictions on the test set against the expected results.

IV. EXPERIMENTAL RESULTS

To obtain the predictive model of the risk assessment of pre-eclampsia, the researcher used three main steps. Firstly, the three preprocessing methods were performed on the imbalanced data and balanced data using scikit-learn with the MinMaxScaler, StandardScaler, and Normalizer. Next, PCA was used to extract the three principal components on the imbalanced data and balanced data. Finally, the three principal components were performed on seven machine learning algorithms which were logistic regression (LR), K-nearest neighbor (KNN), decision tree (DT), random forest (RF), support vector machine (SVM), and naive Bayes (NB). The evaluation results are shown in Tables I-VI.

TABEL I. HASIL HASILS OF THE IMBALANCED DATA WITH STANDARDSCALER.

Machine Learning#	Seleksi Fitur
	PCA
Logistic Regression	69.01
K-Nearest Neighbor	64.79
Decision Tree	61.97
Random Forests	70.42
Support Vector Machines	71.83
Naive Bayes	28.17
Support Vector Machines	70.42

The results can be seen in Table I. The highest accuracy rate was 71.83 percent, shown by the underlined letter, which belonged to the MLP algorithm.

TABEL II. II. HASILS OF THE IMBALANCED DATA WITH MINMAXSCALER.

Machine Learning#	Seleksi Fitur
	PCA
Logistic Regression	71.83
K-Nearest Neighbor	67.61
Decision Tree	60.56
Random Forests	71.83
Support Vector Machines	71.83
Naive Bayes	28.17
Support Vector Machines	71.83

As shown in Table II, the highest accuracy rate was 71.83 percent, shown by the underlined letter, which belonged to the LR, RF, MLP and SVM algorithms.

The results can be seen in Table III. The highest accuracy rate was 73.24 percent, shown by the underlined letter, which belonged to the KNN algorithm.

As shown in Table IV, the highest accuracy rate was 89.62 percent, shown by the underlined letter, which belonged to the RF algorithm.



TABLE III. RESULTS OF THE IMBALANCED DATA WITH NORMALIZER.

Machine Learning	Feature Selection
	PCA
Logistic Regression	71.83
K-Nearest Neighbor	<u>73.24</u>
Decision Tree	59.15
Random Forests	67.61
Multilayer Perceptron	71.83
Naive Bayes	28.17
Support Vector Machines	71.83

TABLE IV. RESULTS OF THE BALANCED DATA WITH STANDARDSCALER.

Machine Learning	Feature Selection
	PCA
Logistic Regression	69.81
K-Nearest Neighbor	72.64
Decision Tree	82.08
Random Forests	<u>89.62</u>
Multilayer Perceptron	45.28
Naive Bayes	48.11
Support Vector Machines	78.30

TABLE V. RESULTS OF THE BALANCED DATA WITH MINMAXSCALER.

Machine Learning	Feature Selection
	PCA
Logistic Regression	66.04
K-Nearest Neighbor	68.87
Decision Tree	79.25
Random Forests	<u>85.85</u>
Multilayer Perceptron	45.28
Naive Bayes	50.00
Support Vector Machines	59.43

The results can be seen in Table V. The highest accuracy rate was 85.85 percent, shown by the underlined letter, which belonged to the RF algorithm.

TABLE VI. RESULTS OF THE BALANCED DATA WITH NORMALIZER.

Machine Learning	Feature Selection
	PCA
Logistic Regression	47.17
K-Nearest Neighbor	69.81
Decision Tree	73.58
Random Forests	<u>85.85</u>
Multilayer Perceptron	45.28
Naive Bayes	48.11
Support Vector Machines	45.28

As shown in Table VI, the highest accuracy rate was 85.85 percent, shown by the underlined letter, which belonged to the RF algorithm.

According to the experiments conducted with the dataset, the RF algorithm yielded the best result based on the balanced data. The highest accuracy rate was 89.62 percent.

## V. CONCLUSION

This paper presented the risk assessment of pregnancy-induced hypertension using a machine learning approach. A public dataset of Logan (2020) was used in this research. In dealing with imbalanced data, the researcher performed the SMOTE algorithm on the dataset. In the data preparation process, the researcher used data transformation, which consisted of MinMaxScaler, StandardScaler and Normalizer, in order to convert data to better reveal the structure of the prediction problem. PCA was used to extract the three principal components on the imbalanced data and balanced data. In the experimental results, the RF algorithm achieved the best performance when compared to other machine learning algorithms. The prediction model yielded an accuracy rate of up to 89.62 percent, which was based on the balanced data. In future work, the researcher should apply the predictive model on the pregnancy-induced hypertension dataset from the Chaophraya Abhaibhubejhr Hospital, Prachin Buri.

## ACKNOWLEDGMENT

This research was supported by the Department of Computer Engineering, Faculty of Engineering, Mahidol University. This research was also supported the National Research Council of Thailand for the project "Mom's buddy: AI chatbot for pregnancy health information". Furthermore, this research was supported by Chaophraya Abhaibhubejhr Hospital on medical knowledge.

## REFERENCES

- [1] Thanomrat Prasith-thimet, and Kasem Wetsutthanon, "Causes of maternal deaths in Regional Health 4 during Fiscal Year 2014-2016," Journal of Health Science, vol.26, pp.5, 2017.
- [2] Logan Gorbee, "Replication Data for: Determinants of preeclampsia and eclampsia among women delivering in county hospitals in Nairobi, Kenya", Harvard Dataverse, version 1.0, 2020. [online] Available: <http://www.doi.org/10.7910/DVN/BYFL3J>. [Accessed: Nov.18, 2020].
- [3] WHO, UNICEF, UNFPA, "The World Bank Trends in maternal mortality: 1990 to 2010", 2010. [online] Available: <https://www.who.int/reproductivehealth/publications/monitoring/9789241503631/en/>. [Accessed: Jan.1, 2019].
- [4] The Medical Council of Thailand, "Medical statistics", 2017. [online] Available: [http://www.tmc.or.th/pdf/01\\_stat\\_med2560.pdf](http://www.tmc.or.th/pdf/01_stat_med2560.pdf). [Accessed: Jan.1, 2019].
- [5] Official Statistics Registration Systems, "The population in each age, Nationwide", 2017. [online] Available: [http://stat.dopa.go.th/stat/statnew/upstat\\_age\\_disp.php](http://stat.dopa.go.th/stat/statnew/upstat_age_disp.php). [Accessed: Jan.1, 2019].
- [6] Muhlis Tahir, Tessa Badriyah, Iwan Syarif, "Classification Algorithms of Maternal Risk Detection For Preeclampsia With Hypertension During Pregnancy Using Particle Swarm Optimization," EMITTER International Journal of Engineering Technology, vol.6, pp.236-250, 2018.
- [7] Costas K. Neocleous, Panagiotis Anastasopoulos, Kypros H. Nikolaides, Christos N. Schizas, Kleanthis C. Neokleous, "Neural networks to estimate the risk for preeclampsia occurrence," in Proc. IEEE International Joint Conference on Neural Networks, 2009, pp.2221-2224.

- [8] Muhlis Tahir, Tessy Badriyah, dan Iwan Syarif, "Algoritma Neural Networks untuk Menanyakan Faktor Preeklampsia Sebelumnya pada Wanita Hipertensi Kronis Selama Kehamilan dalam Proses Persalinan," dalam Proc. Simposium Elektronik Internasional IEEE tentang Penciptaan Pengetahuan dan Komputasi Cerdas (IES-KCIC), 2018, hlm.51-55.
- [9] SY Leemaqz, GA Dekker dan CT Roberts " Sistem Prediksi Berjenjang untuk Preeklampsia: aplikasi integratif dari berbagai model," dalam Proc. Kongres Internasional ke-20 tentang Pemodelan dan Simulasi (MODSIM), 2013, hlm.2041-2045.
- [10] Mohammed Khalilia, Sounak Chakraborty, dan Mihail Popescu, "Memprediksi risiko penyakit dari data yang sangat tidak seimbang menggunakan hutan acak," BMC Medical Informatics and Decision Making, vol.11, pp.1-13, 2011.
- [11] Shadab Adam Pattekari dan Asma Parveen, "Sistem prediksi penyakit jantung menggunakan naif Bayes," Penelitian Biomedis, vol.29, pp.2646-2649, 2018.
- [12] M.Akhil jabbar, BL Deekshatulu, dan Priti Chandra, "Klasifikasi Penyakit Jantung Menggunakan K- Nearest Neighbor dan Genetic Algorithm," Konferensi Internasional tentang Kecerdasan Komputasi: Pemodelan Teknik dan Aplikasi, CIMTA, Kalyani, Kolkata, India, 27 September 2013, hlm.85-94.
- [13] Md Osman Goni Nayeem, Maung Ning Wan, dan Md Kamrul Hasan, "Prediksi Tingkat Penyakit Menggunakan Multilayer Perceptron Jaringan Syaraf Tiruan untuk Pemantauan Pasien," International Journal of Soft Computing and Engineering (IJSCE), vol.5 , hlm. 17-23, September 2015.
- [14] Departemen Kesehatan, dan Kantor Keamanan Kesehatan Nasional (nhso) thailand, "catatan kesehatan ibu dan anak atau buku catatan merah muda umum Thailand," rumah sakit [online] Tersedia: <http://www.oic.go.th/FILEWEB/CABINF0CENTER17/DRAWER002/GENERAL/DATA0001/00001375.PDF>. [Diakses: 1 Januari 2019].
- [15] NV Chawla, KW Bowyer, LO Hall, WP Kegelmeyer, "SMOTE: Teknik Pengambilan Sampel Minoritas Sintetis," jurnal Penelitian Kecerdasan Buatan, vol.16, hlm. 321-357, 2002.