



JURNAL SISTEM INFORMASI DAN TEKNOLOGI (S I N T E K)



Situs Jurnal

<https://sintek.stmikku.ac.id/index.php/home>

ANALISIS DAN PREDIKSI RISIKO KELAHIRAN BAYI MENGGUNAKAN K-MEANS DAN DEEP NEURAL NETWORK (DNN)

Muklisin Illahudin¹, Nana Suarna², Agus Bahtiar³, Mulyawan⁴, Irfan Ali⁵¹²³⁴⁵STMIK IKMI Cirebon

Jl. Perjuangan No. 10 B Majasem Kota Cirebon

Email: ¹mukhlisinilahudin25@gmail.com, ²st_nana@yahoo.com, ³agusbahtiar038@gmail.com,
⁴wm7488748@gmail.com, ⁵irfanaali0.0@gmail.com.

ABSTRAK

Risiko kelahiran bayi merupakan indikator penting dalam evaluasi kesehatan ibu dan anak sehingga diperlukan pendekatan analitis yang mampu mengidentifikasi pola risiko secara akurat. Penelitian ini bertujuan menganalisis dan memprediksi risiko kelahiran bayi dengan mengintegrasikan metode K-Means dan Deep Neural Network (DNN). Dataset yang digunakan terdiri dari 983 data rekam medis ibu hamil yang telah melalui tahap pengumpulan data, pembersihan, dan preprocessing meliputi normalisasi, encoding variabel kategorikal, penanganan outlier, serta seleksi fitur. Metode K-Means digunakan untuk mengelompokkan data berdasarkan kemiripan karakteristik klinis guna membentuk representasi pola risiko awal, yang selanjutnya digunakan sebagai fitur tambahan pada model DNN. Model DNN dirancang menggunakan beberapa hidden layer dengan fungsi aktivasi ReLU dan regularisasi dropout. Hasil pengujian menunjukkan bahwa model menghasilkan akurasi sebesar 61,93% dan nilai ROC AUC sebesar 0,6402, yang mengindikasikan performa moderat dalam memprediksi risiko kelahiran bayi. Stabilitas kurva loss dan akurasi menunjukkan proses pelatihan yang berjalan dengan baik tanpa overfitting signifikan. Secara praktis, model ini berpotensi digunakan sebagai alat bantu awal bagi tenaga kesehatan dalam mengidentifikasi ibu hamil dengan risiko kelahiran lebih tinggi sehingga dapat dilakukan pemantauan dan intervensi lebih dini.

Kata Kunci: *Machine learning, K-Means, Deep Neural Network, Risiko Kelahiran Bayi, Prediksi.*

1. PENDAHULUAN

Risiko kelahiran bayi masih menjadi tantangan dalam upaya peningkatan kualitas layanan kesehatan ibu dan anak. Identifikasi dini terhadap ibu hamil yang berpotensi mengalami komplikasi sangat penting untuk mencegah dampak yang lebih serius pada ibu maupun bayi. Namun, analisis faktor risiko kehamilan memiliki kompleksitas tinggi karena melibatkan berbagai variabel klinis dan demografis yang saling berhubungan secara nonlinier.

Perkembangan machine learning membuka peluang baru dalam analisis data medis karena kemampuannya mempelajari pola kompleks dari

data historis [2]. Metode clustering seperti K-Means dapat digunakan untuk mengidentifikasi pola laten dalam data tanpa label [3], sedangkan Deep Neural Network (DNN) mampu memodelkan hubungan nonlinier yang kompleks pada data medis [1], [4]. Integrasi kedua metode tersebut diharapkan dapat meningkatkan pemahaman terhadap pola risiko kelahiran dan menghasilkan model prediksi yang lebih informatif.

Penelitian ini bertujuan mengembangkan model prediksi risiko kelahiran bayi dengan menggabungkan K-Means sebagai metode eksplorasi pola data dan DNN sebagai model prediksi utama. Hasil penelitian diharapkan

memberikan kontribusi dalam pemanfaatan machine learning untuk mendukung pengambilan keputusan di bidang kesehatan ibu dan anak.

2. LANDASAN TEORI

2.1 Machine Learning

Machine learning merupakan pendekatan komputasional yang memungkinkan sistem belajar dari data untuk menghasilkan prediksi atau keputusan tanpa pemrograman eksplisit [2].

2.2 K-means Clustering

K-Means adalah metode clustering yang membagi data ke dalam K kelompok berdasarkan jarak minimum terhadap pusat cluster (centroid) [3].

2.3 Deep Neural Network

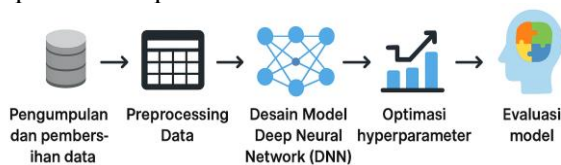
DNN merupakan jaringan saraf berlapis banyak yang mampu mempelajari representasi fitur nonlinier secara mendalam dan banyak digunakan dalam prediksi kesehatan [1], [4].

2.4 Evaluasi Model

Evaluasi model klasifikasi dilakukan menggunakan metrik akurasi, precision, recall, F1-score, dan ROC AUC untuk menilai performa prediksi [6].

3. METODOLOGI PENELITIAN

Penelitian ini dilakukan melalui beberapa tahapan utama, yaitu pengumpulan data, preprocessing, desain model DNN, optimasi hyperparameter, dan evaluasi model. Alur penelitian diperlihatkan pada Gambar 3.1.



Gambar 3.1 Alur Penelitian

3.1 Pengumpulan Data

Dataset penelitian terdiri dari 983 data rekam medis ibu hamil. Pembersihan data mencakup pengecekan duplikasi, penanganan missing value, dan validasi konsistensi data.

3.2 Preprocessing data

Tahap preprocessing meliputi encoding variabel kategorikal, normalisasi fitur numerik, penanganan outlier, serta pembagian data menjadi data latih dan data uji.

3.3 Desain Model Deep Neural Network

Model DNN dirancang dengan beberapa hidden layer beraktivasi ReLU, dropout untuk

mengurangi overfitting, serta lapisan output sigmoid. Proses pelatihan menggunakan optimizer Adam dan fungsi loss Binary Cross-Entropy.

3.4 Optimasi Hyperparameter

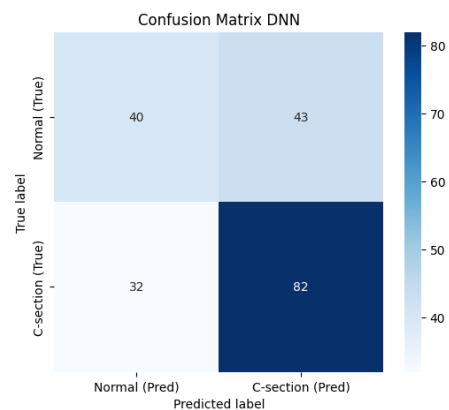
Optimasi hyperparameter dilakukan menggunakan metode Randomized Search dengan parameter seperti jumlah neuron, learning rate, batch size, dan epoch [7].

3.5 Evaluasi Model

Evaluasi model dilakukan menggunakan confusion matrix, kurva loss, kurva akurasi, serta perhitungan akurasi dan ROC AUC untuk menilai kinerja model secara menyeluruh.

a) Confusion Matrix

Confusion Matrix digunakan untuk menggambarkan jumlah prediksi benar dan salah pada setiap kelas.



Gambar 3.2 Confusion Matrix

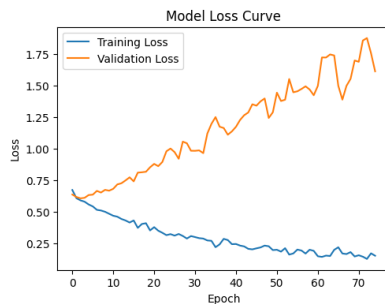
Matrik ini menampilkan distribusi jumlah TP, TN, FP, dan FN. Dari gambar, terlihat bahwa model lebih banyak menghasilkan TN dibanding TP, yang menunjukkan bahwa model lebih baik dalam mengenali kasus tidak berisiko dibanding kasus risiko tinggi. Nilai FN yang terlihat pada gambar menunjukkan model masih perlu peningkatan untuk mendeteksi kasus risiko tinggi secara lebih akurat.

b) Loss Curve

Loss Curve menunjukkan perubahan nilai fungsi kerugian (loss function) selama proses pelatihan model. Kurva ini terdiri dari:

- Training Loss: nilai loss pada data latih

- Validation Loss: nilai loss pada data uji validasi



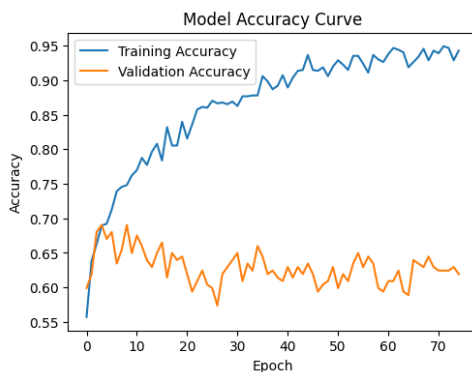
Gambar 3.3 Loss Curve

Loss Curve menunjukkan penurunan loss secara bertahap seiring bertambahnya epoch. Training loss bergerak menurun, menandakan model mampu mempelajari pola dari data latih. Validation loss juga mengalami penurunan namun menunjukkan sedikit fluktuasi. Jika validation loss mulai meningkat sementara training loss terus menurun, hal ini menunjukkan potensi overfitting. Namun, pada gambar terlihat kedua kurva tetap relatif stabil, sehingga model masih berada pada kondisi pembelajaran yang wajar.

c) Accuracy Curve

Accuracy Curve digunakan untuk menilai perkembangan performa model dalam mengklasifikasikan data secara benar pada setiap epoch. Dalam Accuracy Curve terdapat:

- Training Accuracy: akurasi pada data latih
- Validation Accuracy: akurasi pada data uji validasi



Gambar 3.4 Accuracy Curve

Accuracy Curve menunjukkan peningkatan akurasi pada data latih dan uji seiring bertambahnya epoch. Accuracy Curve umumnya bergerak naik, menandakan model mampu mengenali pola data dengan baik. Meskipun terdapat perbedaan antara akurasi latih dan uji, jaraknya tidak terlalu besar, menunjukkan bahwa model tidak mengalami overfitting yang signifikan.

4. HASIL DAN PEMBAHASAN

4.1 Hasil

Penelitian ini menghasilkan model prediksi risiko kelahiran bayi berbasis integrasi metode K-Means dan Deep Neural Network (DNN) menggunakan dataset sebanyak 983 data rekam medis ibu hamil. Proses clustering dengan K-Means mampu mengelompokkan data ke dalam beberapa cluster yang merepresentasikan karakteristik klinis berbeda, termasuk kelompok dengan indikasi risiko kelahiran lebih tinggi. Informasi cluster tersebut digunakan sebagai fitur tambahan dalam proses pelatihan model DNN. Hasil pengujian menunjukkan bahwa model DNN menghasilkan nilai akurasi sebesar 61,93% dan nilai ROC AUC sebesar 0,6402, yang menunjukkan kemampuan moderat dalam membedakan kelas risiko dan non-risiko. Confusion matrix memperlihatkan bahwa model lebih akurat dalam mengklasifikasikan kasus non-risiko dibandingkan risiko tinggi, yang dipengaruhi oleh distribusi kelas yang tidak seimbang. Selain itu, kurva loss dan kurva akurasi menunjukkan pola pelatihan yang stabil tanpa indikasi overfitting signifikan.

4.2 Pembahasan

Hasil penelitian menunjukkan bahwa integrasi metode K-Means dan DNN memberikan pendekatan yang cukup efektif dalam menganalisis pola risiko kelahiran bayi, meskipun performa prediksi belum optimal. Performa ini lebih rendah dibandingkan penelitian Miotto et al. yang menggunakan deep learning pada dataset berskala besar [4], serta pendekatan deep clustering berbasis autoencoder yang dilaporkan oleh Guo et al. [5]. Perbedaan tersebut dapat dijelaskan oleh keterbatasan ukuran dataset dan ketidakseimbangan kelas, yang menurut Sarker merupakan faktor penting dalam kinerja model machine learning pada data medis [2]. Stabilitas kurva loss dan akurasi menunjukkan bahwa arsitektur DNN yang digunakan telah sesuai dengan karakteristik data, sejalan dengan prinsip pelatihan deep learning yang dikemukakan oleh Goodfellow et al. [1]. Dengan demikian, meskipun performa masih moderat, hasil penelitian ini menunjukkan bahwa pendekatan clustering sebagai fitur pendukung DNN dapat memberikan wawasan tambahan mengenai struktur data dan pola risiko kelahiran.

5. PENUTUP

Penelitian ini membuktikan bahwa metode K-Means dan Deep Neural Network dapat diintegrasikan untuk menganalisis dan memprediksi risiko kelahiran bayi berdasarkan data rekam medis ibu hamil. Kontribusi utama penelitian ini terletak

pada pemanfaatan unsupervised learning sebagai fitur pendukung dalam model deep learning untuk memperkaya representasi data medis. Hasil penelitian menunjukkan bahwa model mampu mengenali pola risiko kelahiran secara moderat dengan proses pelatihan yang stabil. Temuan ini memberikan kontribusi ilmiah dalam pengembangan pendekatan machine learning untuk analisis risiko kesehatan ibu dan anak serta berpotensi menjadi dasar pengembangan sistem pendukung keputusan klinis.

6. SARAN

Untuk pengembangan penelitian selanjutnya, disarankan menggunakan dataset dengan jumlah dan variasi yang lebih besar agar model dapat mempelajari pola risiko secara lebih komprehensif. Selain itu, perlu diterapkan teknik penanganan ketidakseimbangan kelas seperti SMOTE atau metode cost-sensitive learning untuk meningkatkan sensitivitas model terhadap kasus risiko tinggi. Eksplorasi arsitektur deep learning lain, seperti LSTM, CNN, atau model berbasis attention, juga dapat dilakukan untuk meningkatkan performa prediksi. Penambahan variabel klinis yang lebih detail serta implementasi model dalam bentuk sistem pendukung keputusan berbasis web atau mobile juga menjadi arah pengembangan yang potensial.

DAFTAR PUSTAKA

- [1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [2] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Computer Science*, 2021.
- [3] S. Lloyd, "Least Squares Quantization in PCM," *IEEE Transactions on Information Theory*, 1982.
- [4] R. Miotto et al., "Deep Patient: An Unsupervised Representation to Predict Patient Health," *Scientific Reports*, 2016.
- [5] X. Guo et al., "Deep Clustering with Convolutional Autoencoders," *NeurIPS*, 2017.
- [6] D. M. Powers, "Evaluation Metrics," *Journal of Machine Learning Technologies*, 2011.
- [7] J. Bergstra and Y. Bengio, "Random Search for Hyperparameter Optimization," *Journal of Machine Learning Research*, 2012.