

IJCIT
(Indonesian Journal on Computer and Information Technology)
Journal Homepage: <http://ejournal.bsi.ac.id/ejurnal/index.php/ijcit>

Peningkatan Akurasi Pada Algoritma ID3 Menggunakan Operator Bagging Dalam Mendiagnosa Kesehatan Kehamilan

Hikmatulloh¹, Susilawati², Siti Nurajizah³

¹Teknik Informatika, Universitas Nusa Mandiri
Jakarta, Indonesia
e-mail: hikmatulloh.hkl@nusamandiri.ac.id

²Sistem Informasi Kampus Kota Sukabumi, Universitas Bina Sarana Informatika
Sukabumi, Indonesia
e-mail: susilawati.ssl@bsi.ac.id

³Sistem Informasi Akuntansi Kampus Kabupaten Karawang, Universitas Bina Sarana Informatika
Karawang, Indonesia
e-mail: siti.snz@bsi.ac.id

ABSTRAK

Semakin meningkatnya resiko kematian yang dialami oleh ibu hamil, salah satu penyebabnya dikarenakan faktor keterlambatan dalam mengambil sebuah keputusan. Hal tersebut menjadi dasar untuk dilakukan kembali peningkatan akurasi pembelajaran mesin (*mechine learning*) dalam mengolah data untuk mendiagnosa kesehatan kehamilan. Penelitian ini dilakukan dengan menggunakan dataset dari penelitian sebelumnya pada tahun 2019 berjudul "Penerapan Algoritma *Iterative Dichotomiser Three* (ID3) Dalam Mendiagnosa Kesehatan Kehamilan", yang menghasilkan akurasi sebesar 80.33%. Tujuan penelitian ini yaitu untuk meningkatkan akurasi dari algoritma *Iterative Dichotomiser Three* (ID3) yang telah dilakukan pada penelitian sebelumnya tersebut. Metode yang digunakan untuk meningkatkan akurasi dalam penelitian ini yaitu menggunakan *metode ensemble learning bagging*. Sampel rasio yang diterapkan pada metode *metode ensemble learning bagging* ini sebesar 0.9 dengan 5 kali iterasi. Hasil penelitian ini adalah berupa peningkatan akurasi setelah ditambahkan metode *ensemble learning bagging* pada algoritma ID3 dengan hasil peningkatan sebesar 3.19%, sehingga akurasi yang didapat dalam model data yang diolah untuk mendiagnosa kesehatan kehamilan naik menjadi 83.52%.

Katakunci: *bagging*, *ensemble*, *id3*, kehamilan

ABSTRACTS

The increasing risk of death experienced by pregnant women, one of the causes is due to the delay in making a decision. This becomes the basis for improving the accuracy of machine learning in processing data to diagnose pregnancy health. This research was conducted using a dataset from previous research in 2019 entitled "Penerapan Algoritma Iterative Dichotomiser Three (ID3) Dalam Mendiagnosa Kesehatan Kehamilan", which resulted in an accuracy of 80.33%. The purpose of this study is to improve the accuracy of the Iterative Dichotomiser Three (ID3) algorithm that has been carried out in the previous research. The method used to improve accuracy in this study is to use the ensemble learning bagging method. The sample ratio applied to this bagging method is 0.9 with 5 iterations. The amount of increase



obtained after adding the bagging method to the ID3 algorithm was 3.19%, so that the accuracy obtained in the processed data model to diagnose pregnancy health rose to 83.52%.

Keywords: bagging, ensemble, id3, pregnancy

1. PENDAHULUAN

Peningkatan akurasi pada pembelajaran mesin (*machine learning*) dalam mendiagnosa suatu masalah kesehatan ibu hamil penting menjadi perhatian. Karena, resiko kematian ibu hamil yang semakin tinggi akibat adanya faktor keterlambatan mengambil keputusan untuk dirujuk (Aji, Furqon, & Widodo, 2018). Hasil persentase akurasi yang baik atas ketepatan diagnosa kesehatan kehamilan memungkinkan praktisi terkait dapat memperoleh masukan yang berharga tentang pemeriksaan mengenai kesehatan kehamilan.

Untuk lebih meminimalkan kemungkinan resiko akibat kehamilan tersebut, perlu dipilih jenis teknik sampling yang sesuai dengan kondisi populasi sehingga hasil penelitian memiliki nilai akurasi yang tinggi terhadap nilai parameter (Sumanto, 2005). Metode yang digunakan dalam meningkatkan akurasi untuk mendiagnosa kesehatan kehamilan tersebut, menggunakan salah satu metode *ensemble learning* yaitu *bagging*. *Ensemble learning* adalah metode yang digunakan untuk meningkatkan akurasi algoritma klasifikasi dengan membangun beberapa *classifier* dari data *training* (Mirqotussa'adah, Muslim, Sugiharti, Prasetyo, & Alimah, 2017). Metode *ensemble learning* ini menggunakan *voting* dan *aggregating* dari golongan-golongan data. Sedangkan operator *bagging* merupakan suatu bentuk set pelatihan dasar belajar menggunakan prosedur pembelajaran yang tidak stabil, dan kemudian, selama pengujian, mengambil rata-rata (Arrahimi, Ihsan, Kartini, Faisal, & Indriani, 2019).

Penelitian mengenai kesehatan kehamilan pernah dilakukan sebelumnya pada tahun 2018 dengan menggunakan metode *forward chaining* yang dapat mendiagnosa 13 jenis penyakit yang sering dialami ibu hamil (Abdillah, Nurajijah, & Nawawi, 2018). Namun perbedaan dari penelitian ini bukan menggunakan jenis penyakit yang di alami melainkan mengambil gejala-gejala yang di alami oleh ibu hamil seperti diantaranya mual dan muntah, perut kembung, nyeri ulu hati, sensitif terhadap bau, darah pada vagina, sakit buang air kecil, jumlah buang

air kecil, tekanan darah, denyut nadi, suhu tubuh yang di teliti pada setiap trimesternya.

Hasil penelitian sebelumnya dalam mendiagnosa kesehatan kehamilan telah dilakukan, yakni dengan menerapkan algoritma ID3 (*Iterative Dichotomiser Three*) yang menghasilkan akurasi sebesar 80.33%. Hasil akurasi tersebut dirasa masih perlu untuk ditingkatkan kembali agar menghasilkan sebuah model pembelajaran mesin yang lebih baik. Algoritma ID3 menggunakan metrik *Entropy & Information Gain*, *Entropy* digunakan menilai seberapa penting sebuah node (Sulaiman, 2020).

Tujuan dari penelitian ini yaitu untuk menguji model *ensemble learning bagging* dalam meningkatkan akurasi algoritma ID3 untuk mendiagnosa kesehatan kehamilan agar lebih tepat sasaran. Penelitian ini menggunakan dataset dan algoritma yang sama pada penelitian sebelumnya yang berjudul "Penerapan Algoritma *Iterative Dichotomiser Three* (ID3) Dalam Mendiagnosa Kesehatan Kehamilan" (Hikmatulloh, Rahmawati, Wintana, & Ambarsari, 2019).

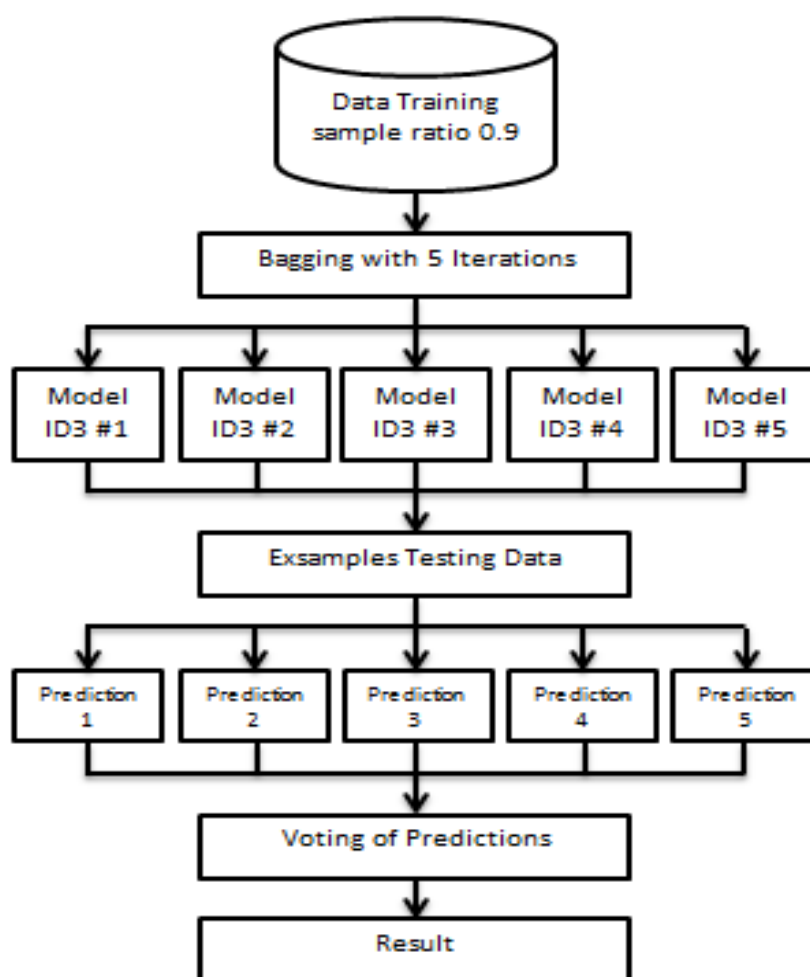
2. METODE PENELITIAN

Metode dalam penelitian ini merupakan proses pengolahan suatu model *ensemble learning* yaitu *bagging*, metode *bagging* atau *bootstrap aggregating* merupakan salah satu metode untuk meningkatkan simulasi (Foroozand & Weijs, 2017). Merupakan sebuah pembelajaran mesin untuk meningkatkan kualitas model klasifikasi maupun regresi dalam hal peningkatan akurasi dan stabilitas klasifikasi data. *Bagging* dikenal sangat efektif bila pengklasifikasi tidak stabil, yaitu ketika penturbing set pembelajaran dapat menyebabkan perubahan yang signifikan dalam perilaku klasifikasi, karena *bagging* meningkatkan kinerja generalisasi dengan cara mengurangi varians (*noise*) dengan tetap menjaga atau hanya sedikit meningkatkan bias (Setiyorini & Wahono, 2015).

Seperti terlihat pada gambar 1, data training dengan sample ratio 0,9 dalam penelitian ini digunakan untuk melatih algoritma

dengan menggunakan operator bagging, pada tahapan pelatihan dengan operator bagging tersebut digunakan sebanyak 5 iterasi yang menghasilkan 5 model ID3 (Model ID3 #1, Model ID3 #2, Model ID3 #3, Model ID3 #4, Model ID3 #5). Selanjutnya exsample testing data digunakan untuk mendapatkan performa algoritma yang sudah dilatih sebelumnya, dan

menghasilkan 5 predikti dari data testing yang di ujikan (prediction 1, prediction 2, prediction 3, prediction 4, dan prediction 5). Hasil akhir dari metode penelitian ini yaitu menghasilkan satu prediksi yang diambil dari proses voting dari kelima prediksi tersebut, dimana untuk hasil prediksinya ini merupakan hasil prediksi dari pohon model yang paling banyak muncul.



Gambar 1. Metode Penelitian

Prosedur penelitian yang dilakukan diantaranya:

1. Data yang digunakan dalam set pelatihan diambil dengan rasio data 0.9 sampel.
2. Pelatihan data yang dilakukan dalam peneltian ini akan dilakukan 5 kali iterasi.
3. Hasil dari pelatihan *bagging* dengan iterasi sebanyak 5 kali, menghasilkan 5 model pohon keputusan yang nantinya akan digunakan untuk melakukan *testing* pada data *testing* untuk memprediksi sebuah

kejadian berdasarkan atribut yang di inputkan.

4. Setelah didapatkan 5 pohon keputusan, data *testing* bisa di inputkan pada 5 model pohon keputusan yang telah didapatkan.
5. Untuk menghasilkan prediksi akhir, maka dilakukan pendugaan gabungan berdasarkan 5 pohon yang telah terbentuk, pemilihan satu dugaan dapat menggunakan *majority vote* (Agwil, Fransiska, & Hidayati, 2019).

3. HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan sebuah peningkatan akurasi dari algoritma ID3 (*Iterative Dichotomiser Three*) dalam mendiagnosa kesehatan kehamilan, dalam peningkatan akurasi tersebut menggunakan salasatu operator yang terdapat pada perangkat *Rapid*

Miner yaitu operator *bagging* (*bootstrap aggregating*).

Atribut yang digunakan dalam pengolahan data pada penelitian ini sebanyak 14 atribut. Sedangkan untuk kelas yang dihasilkan yaitu menghasilkan kelas untuk kehamilan yang sehat dan kehamilah yang tidak sehat. Adapun rincian informasi dataset dalam penelitian ini dapat dilihat pada tabel 1.

Tabel 1. Informasi Dataset

Nama Data	Keterangan
Tahapan Kehamilan	atribut data dengan type data <i>polynomial</i> dan nilai (trimester 1, trimester 2, trimester 3)
Mual & Muntah	atribut data dengan type data <i>polynomial</i> dan nilai (meningkat, sedang, tidak ada)
Perut Kembang	atribut data dengan type data <i>polynomial</i> dan nilai (iya, tidak)
Nyeri Ulu Hati	atribut data dengan type data <i>polynomial</i> dan nilai (iya, meningkat, tidak)
Sensitif Terhadap Bau	atribut data dengan type data <i>polynomial</i> dan nilai (iya, tidak)
Darah pada Vagina	atribut data dengan type data <i>polynomial</i> dan nilai (meningkat, sedang, tidak ada)
Sakit BAK	atribut data dengan type data <i>polynomial</i> dan nilai (meningkat, sedan, tidak)
Jumlah BAK	atribut data dengan type data <i>polynomial</i> dan nilai (meningkat, menurun, sedang, stabil)
Tekanan Darah	atribut data dengan type data <i>polynomial</i> dan nilai (menurun, stabil)
Denyut Nadi	atribut data dengan type data <i>polynomial</i> dan nilai (meningkat, stabil)
Suhu Tubuh & Dehidrasi	atribut data dengan type data <i>polynomial</i> dan nilai (meningkat, stabil)
Kencang pada Rahim	atribut data dengan type data <i>polynomial</i> dan nilai (ada, sedang, tidak ada)
Bau Cairan dari Vagina	atribut data dengan type data <i>polynomial</i> dan nilai (sangat berbau, sedikit berbau, tidak berbau)
Warna Cairan dari Vagina	atribut data dengan type data <i>polynomial</i> dan nilai (berwarna, tidak berwarna)
Hasil	kelas data dengan type data <i>polynomial</i> dan nilai (sehat, tidak sehat)

Sumber: (Hikmatulloh et al., 2019)

Operator *bagging* yang digunakan untuk meningkatkan akurasi pada algoritma ID3 ini, menjalankan pengulangan sebanyak 5 *iterations*, sehingga proses ini menghasilkan 5 pohon keputusan yang nantinya akan ditarik kesimpulan berupa jumlah hasil rata-rata yang didapatkan dari 5 hasil keputusan tersebut. Dan pada proses peningkatan akurasi algoritma ID3 menggunakan operator *bagging* ini menghasilkan akurasi sebesar 83.52%. Berikut ini disajikan tabel *confusion matrix* yang dihasilkan pada algoritma *bagging* ID3 untuk melihat seberapa baik performa kualitas sebuah model dalam mendiagnosa kesehatan kehamilan.

Tabel 2. *Confusion Matrix*

#	True. Sehat	True. Tidak Sehat	Class Precision
Pred. Sehat	44	12	78.57%
Pred. Tidak Sehat	3	32	91.43%
Class Recall	93.62%	72.73%	-

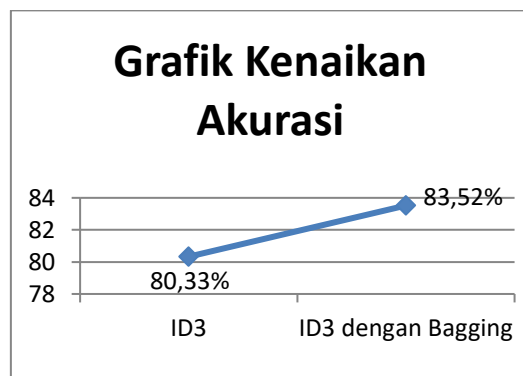
Accuracy:

$$= \frac{\sum TP}{Total Data} \times 100\%$$

$$\begin{aligned}
 &= \frac{TP(\text{sehat}) + TP(\text{tidak sehat})}{\text{Total Data}} \times 100\% \\
 &= \frac{44+32}{91} \times 100\% \\
 &= \frac{76}{91} \times 100\% \\
 &= 83,52\%
 \end{aligned}$$

Penerapan operator *bagging* pada algoritma ID3 (*Iterative Dichotomiser Three*) untuk pengolahan data dalam mendiagnosa kesehatan kehamilan dapat memberikan peningkatan akurasi. Adapun grafik hasil perbandingan pada penelitian ini dengan penelitian sebelumnya oleh (Hikmatulloh et al., 2019) yang menggunakan algoritma ID3 tanpa operator *bagging* dapat dilihat pada gambar 2.

Hasil pengolahan data untuk mendiagnosa kesehatan kehamilan menggunakan algoritma ID3 dengan operator *bagging* terbukti dapat meningkatkan akurasi sebesar 3.19%.



Gambar 2. Grafik Kenaikan Akurasi

4. KESIMPULAN

Operator *bagging* dengan rasio sampel sebesar 0.9 dan 5 kali iterasi yang digunakan dalam algoritma ID3 (*Iterative Dichotomiser Three*) dapat meningkatkan akurasi dalam mendiagnosa kesehatan kehamilan. Besaran peningkatan sebanyak 3.19% sehingga akurasi yang didapat dalam model data yang dioah untuk mendiagnosa kesehatan kehamilan naik menjadi 83.52% dari penelitian sebelumnya sebesar 80.33%.

5. REFERENSI

Abdillah, A., Nurajijah, N., & Nawawi, I. (2018). Perancangan Sistem Pakar Diagnosa

Penyakit Kehamilan Berbasis Web. *Jurnal Techno Nusa Mandiri*, 15(2), 115. <https://doi.org/10.33480/techno.v15i2.910>

Agwil, W., Fransiska, H., & Hidayati, N. (2019). Analisis ketepatan waktu lulus mahasiswa dengan menggunakan bagging cart, 155–166.

Aji, A. H., Furqon, M. T., & Widodo, A. W. (2018). Sistem Pakar Diagnosa Penyakit Ibu Hamil Menggunakan Metode Certainty Factor (CF), 2(5), 2127–2134.

Arrahimi, A. R., Ihsan, M. K., Kartini, D., Faisal, M. R., & Indriani, F. (2019). Teknik Bagging Dan Boosting Pada Algoritma CART Untuk Klasifikasi Masa Studi Mahasiswa. *Jurnal Sains Dan Informatika*, 5(1), 21–30. <https://doi.org/10.34128/jsi.v5i1.171>

Foroozand, H., & Weijs, S. V. (2017). Entropy ensemble filter: A modified bootstrap aggregating (Bagging) procedure to improve efficiency in ensemble model simulation. *Entropy*, 19(10). <https://doi.org/10.3390/e19100520>

Hikmatulloh, H., Rahmawati, A., Wintana, D., & Ambarsari, D. A. (2019). Penerapan Algoritma Iterative Dichotomiser Three (Id3) Dalam Mendiagnosa Kesehatan Kehamilan. *Klik - Kumpulan Jurnal Ilmu Komputer*, 6(2), 116. <https://doi.org/10.20527/klik.v6i2.189>

Mirqotussa'adah, M., Muslim, M. A., Sugiharti, E., Prasetyo, B., & Alimah, S. (2017). Penerapan Dizcretization dan Teknik Bagging Untuk Meningkatkan Akurasi Klasifikasi Berbasis Ensemble pada Algoritma C4.5 dalam Mendiagnosa Diabetes. *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, 8(2), 135. <https://doi.org/10.24843/lkjiti.2017.v08.i02.p07>

Setiyorini, T., & Wahono, R. S. (2015). Penerapan Metode Bagging untuk Mengurangi Data Noise pada Neural Network untuk Estimasi Kuat Tekan Beton, 1(1), 37–42.

Sulaiman, O. K. (2020). *Merdeka Kreatif di Era Pandemi Covid-19*.

Sumanto, D. (2005). Presisi dan Akurasi Hasil Penelitian Kuantitatif Berdasarkan Pengambilan Sample Secara Acak. *Jurnal Litbang Universitas Muhammadiyah Semarang*, 2(2), 45–53. Retrieved from <http://jurnal.unimus.ac.id>