

IMPLEMENTASI ALGORITMA DECISION TREE UNTUK MENGLASIFIKASI KONDISI KESUBURAN PRIA

Arif Amrulloh, Adityo Permana Wibowo

Program Studi Magister Teknologi Informasi, Universitas Teknologi Yogyakarta

Jl. Siliwangi (Ringroad Utara) Jombor, Sleman, Yogyakarta

¹arif.amrulloh@student.uty.ac.id, ²adityopw@staff.uty.ac.id

Abstrak

Tingkat kesuburan atau *Fertility* merupakan hal penting dalam proses melanjutkan keturunan pada pasangan suami istri. Beberapa penelitian menyatakan bahwa banyak faktor yang mempengaruhi seperti hormon, penyakit bawaan, tindakan operasi, konsumsi alkohol dan merokok. Untuk mengetahui faktor apa saja yang paling berpengaruh perlu dilakukan penelitian dalam hal tersebut. *Decision tree* merupakan salah satu metode yang dapat digunakan untuk mengklasifikasi dan memprediksi karena memiliki kemudahan dalam interpretasi hasil. Klasifikasi dapat mengetahui faktor apa saja yang paling berpengaruh terhadap tingkat kesuburan atau *fertility* sehingga dapat digunakan sebagai acuan untuk merubah pola atau kebiasaan sehari-hari. Pelatihan dan pengujian disimulasi menggunakan aplikasi *Rapidminer* dengan hasil akurasi 88% untuk pelatihan dan 92% untuk pengujian.

Kata Kunci: *Decision Tree, Fertility, Klasifikasi, Rapidminer*

1. Pendahuluan

Menurut [1] dalam [2] Kesuburan (*fertility*) pada pria merupakan salah satu faktor penting dalam proses melanjutkan keturunan pada pasangan suami istri. Beberapa dekade belakangan dibidang kesehatan di mancanegara terjadi permasalahan yaitu tingkat kesuburan pria. Beberapa penelitian tentang tingkat kesuburan menyatakan bahwa faktor-faktor yang mempengaruhi tingkat kesuburan seperti hormon, penyakit bawaan, pernah tidaknya dioperasi. Faktor lain seperti konsumsi minuman beralkohol, duduk terlalu lama, dan merokok, dsb [3].

Angka infertilitas pasangan suami istri di Indonesia yang mengalami kesulitan untuk mendapatkan anak yaitu sekitar 10%, diantara 50% nya adalah pria. Menurut WHO lebih dari 50% penyebab infertilitas bagi pria dan yang terbesar oleh faktor semen [4]. Beberapa faktor yang dapat mempengaruhi kesuburan sperma salah satunya yaitu pola gaya hidup dan lingkungan [4].

Beberapa penelitian yang membahas tentang tingkat kesuburan pria, menyatakan bahwa banyak faktor-faktor yang mempengaruhi tingkat kesuburan seperti hormon, penyakit bawaan, tindakan operasi. Faktor lain seperti konsumsi minuman beralkohol, duduk terlalu lama dan kebiasaan merokok [2].

Asap rokok yang dihirup seseorang perokok, mengandung komponen gas dan partikel tersebut sangat berpotensi untuk menimbulkan radikal bebas. Pembentukan senyawa radikal bebas yang tidak segera dinetralkan oleh system antioksidan dapat menimbulkan terjadinya stress oksidatif yang banyak. Kelebihan produksi radikal bebas atau *oxygen* yang reaktif (*ROS, reactive oxygen species*) dapat merusak sperma [5].

Penelitian terdahulu tentang kesuburan (*fertility*) sudah pernah dilakukan, diantaranya yaitu penelitian yang dilakukan oleh [2] dan [4]. Menggunakan data yang sama yaitu dari dataset UCI (University of California, Irvine) Machine Learning Repos. Penelitian yang dilakukan oleh [2] melakukan prediksi kesuburan dengan menggunakan *principal component analysis* kemudian diklasifikasikan menggunakan metode *Naïve Bayes Classifier*. Sehingga menghasilkan nilai

akurasi kebenaran sebesar 80%. Sedangkan penelitian yang dilakukan oleh [4] melakukan diagnosa kesuburan sperma menggunakan Metode K-Nearest Neighbor.

Penelitian ini menggunakan data yang sama dari penelitian sebelumnya, yaitu menggunakan sebanyak 100 data dengan 9 atribut input. Hanya saja pada penelitian ini menggunakan 3 atribut input yang diolah menggunakan Metode Decision Tree. Sebagai atribut target atau luaran dibagi menjadi 2 yaitu Normal dan Altered. Dengan mengurangi jumlah atribut input, bisa menambah nilai akurasi serta menghasilkan atribut utama untuk menentukan kesuburan (fertility).

2. Metode Penelitian

2.1. Dataset

Jumlah data sebanyak 100 data dengan 9 atribut input yang terdiri dari musim, umur, penyakit bawaan, kecelakaan atau trauma, tindakan operasi, Frekuensi konsumsi alkohol, kebiasaan merokok dan lama duduk perhari. Komposisi data yang digunakan untuk pelatihan sebanyak 75 data dan untuk pengujian digunakan sebanyak 25 data. Dari 9 atribut input yang ada, pada penelitian ini hanya digunakan 3 atribut input, seperti yang terlihat pada Tabel 1.

Tabel 2. 1. Atribut Input Klasifikasi Kondisi Kesuburan Pria

Atribut	Kriteria	Bobot Nilai
Frekuensi konsumsi alkohol	1. Beberapa kali sehari	1
	2. Setiap hari	2
	3. Beberapa kali seminggu	3
	4. Seminggu sekali	4
	5. Hampir tidak pernah atau tidak pernah sama sekali.	5
Kebiasaan merokok	1. Tidak pernah	-1
	2. Kadang-kadang	0
	3. Setiap hari	1
Lama duduk perhari	1. Kurang dari 16 Jam	0
	2. 16 jam	1

Atribut output dari data tersebut adalah melakukan pengelompokan klasifikasi kondisi kesuburan pria dengan label Normal (N) atau Altered (O). Altered adalah kondisi dimana terjadi penurunan kualitas kesuburan pada pria. Seperti terlihat pada Tabel 2.

Tabel 2.2 Atribut Output Klasifikasi Kondisi Kesuburan Pria

Atribut	Inisialisasi	Keterangan
Normal	N	Kesuburan pria mengalami kondisi Normal
Altered	O	Kesuburan pria mengalami kondisi Penurunan Kualitas

2.2. Decision Tree

Decision tree adalah struktur flowcart yang mempunyai tree (pohon), dimana setiap simpul internal menandakan suatu tes atribut, setiap cabang merepresentasikan hasil tes, dan simpul daun merepresentasikan kelas atau distribusi kelas, alur pada decision tree ditelusuri dari simpul ke akar simpul daun yang memegang prediski kelas. Decision tree adalah salah satu metode yang digunakan untuk mengklasifikasikan dan prediksi karena memiliki kemudahan dalam interpretasi hasil [6].

2.3. Klasifikasi

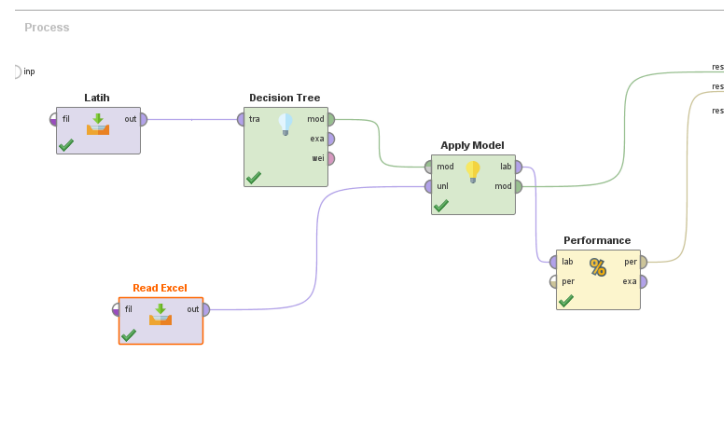
Klasifikasi dapat digambarkan sebagai berikut. Data diinput, disebut juga training set, terdiri atas banyak contoh (record), yang masing-masing memiliki beberapa atribut. Selanjutnya, tiap contoh diberi sebuah tabel class khusus. Tujuannya untuk menganalisa data input dan mengembangkan deskripsi atau model akurat untuk tiap class menggunakan fitur-fitur pada data. Deskripsi class ini digunakan untuk mengklasifikasikan data pengujian lainnya dengan lebel class tidak diketahui. Deskripsi tersebut juga dapat digunakan untuk dapat memahami tiap class dalam data [7].

Untuk mencapai tujuan dilakukan langkah-langkah sebagai berikut:

1. Menyiapkan data pelatihan dan pengujian
Data pelatihan dan pengujian dia mbil dari UCI machine learning dengan jumlah data 100, data yang digunakan untuk pelatihan sebanyak 75 sedangkan untuk pengujian sebanyak 25.
2. Penggunaan aplikasi rapidminer untuk melakukan pelatihan dan pengujian data.
3. Melatih dan menguji data menggunakan aplikasi rapidminer dengan menggunakan decision tree.

2.4. Pemodelan Klasifikasi

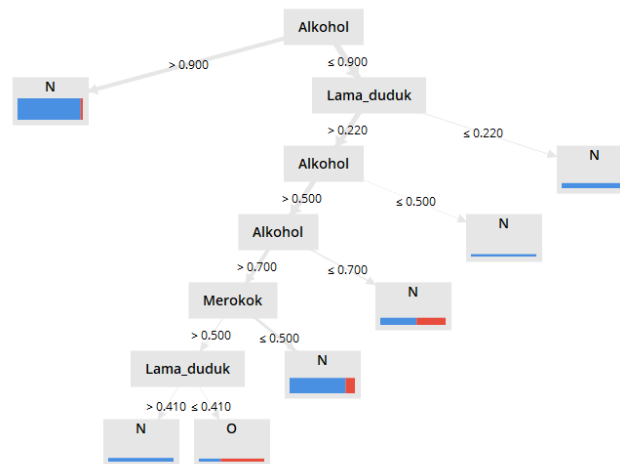
Untuk mendapatkan hasil klasifikasi yang optimal, dibutuhkan terlebih dahulu pemodelan data yang didapatkan dari proses pelatihan data. Model yang dihasilkan dari pemrosesan pelatihan data, nantinya akan digunakan untuk proses pengujian data. Penelitian ini menggunakan aplikasi bantuan Rapidminer untuk pelatihan dan pengujian data. Komponen yang digunakan untuk pemodelan data klasifikasi pada Rapidminer seperti terlihat pada Gambar 1.



Gambar 2.1 Model klasifikasi rapidminer

3. Hasil dan Analisis

Berdasarkan 75 data yang diolah sebagai data latih dengan 3 atribut input, menghasilkan model keputusan berupa pohon keputusan. Pohon keputusan tersebut mempunyai *root* adalah Frekuensi Konsumsi Alkohol, kemudian pada child nya adalah Lama Duduk Perhari. Model pohon keputusan hasil generate Rapidminer terlihat pada Gambar 2.



Gambar 2.2 Model Pohon Keputusan klasifikasi Tingkat Kesuburan Pria

Berdasarkan Gambar 2, terlihat bahwa Frekuensi Konsumsi Alkohol merupakan kriteria terpenting dari sebuah penentu tingkat kesuburan pria, terbukti Frekuensi Konsumsi Alkohol menjadi *root* dari pohon keputusan tersebut. Jika nilai Frekuensi Konsumsi Alkohol lebih dari 0.6, maka kesuburan pria tersebut adalah Normal, namun jika nilai Frekuensi Konsumsi Alkohol kurang dari atau sama dengan 0.6 maka selanjutnya terjadi proses perhitungan nilai lama duduk, begitu juga dan seterusnya.

Sehingga dengan demikian bisa diasumsikan bahwa kriteria Frekuensi Konsumsi Alkohol dan Lama Duduk Perhari merupakan kriteria yang bisa digunakan untuk menentukan tingkat kesuburan pria.

Selanjutnya, untuk mengetahui tingkat akurasi Algoritma *Decision Tree* dalam mengklasifikasi tingkat kesuburan pria, maka dilakukan pengujian menggunakan 25 data. Berdasarkan hasil pengujian didapatkan nilai akurasi kebenaran sebesar 92% dengan nilai N=23 dan nilai O=2. Seperti terlihat pada Gambar 3.

accuracy: 92.00%

	true N	true O	class precision
pred. N	23	2	92.00%
pred. O	0	0	0.00%
class recall	100.00%	0.00%	

Gambar 2.3 : Akurasi data uji

4. Kesimpulan

Algoritma *Decision tree C4.5* dapat digunakan untuk proses klasifikasi tingkat kesuburan pria dengan nilai akurasi yang didapatkan sebesar 92.00% dengan jumlah data uji sebanyak 25 data.

Masih diperlukan variabel input tambahan yang mendukung agar tingkat akurasinya bisa mencapai angka 100%.

Berdasarkan tiga variabel input yang di uji, didapatkan hasil bahwa merokok tidak berpengaruh terhadap tingkat kesuburan.

References

- [1] D. S. Irvine, "Male reproductive health: Cause for concern?," 2000, pp. 195–208.
- [2] G. A. Irawan and A. Muliantara, "Prediksi Kesuburan (Fertility) Dengan Menggunakan Principal Component Analysis Dan Klasifikasi Naive Bayes," *J. Ilm. Ilmu Komput.*, vol. X, no.

- 2, pp. 10–14, 2017.
- [3] M. D. Ana Carolina Martini, Ph.D., Rosa Ine´s Molina, B.S., Daniel Estofa´n and M. . Daniel Senestrari, M.D., Marta Fiol de Cuneo, M.D. and Ruben Daniel Ruiz, “Effects of alcohol and cigarette consumption on human seminal quality,” in *American Society for Reproductive Medicine*, 2004, vol. 82, no. 2, p. 374.
- [4] R. Yepriyanto, Kustanto, and Y. R. W. Utami, “Sistem Diagnosa Kesuburan Sperma Dengan Metode K-Nearest Neighbor (K-NN),” *J. Ilm. SINUS*, pp. 33–44.
- [5] P. D. Sari, “EFFECT OF CIGARETTE SMOKE IN QUALITY AND QUANTITY SPERMATOZOA,” *Med. J. Lampung Univ.*, vol. 3, no. 7, pp. 102–106, 2014.
- [6] Rismayanti, “Decision Tree Penentuan Masa Studi Mahasiswa Prodi Teknik Informatika (Studi Kasus : Fakultas Teknik dan Komputer Universitas Harapan Medan),” *Query J. Sist. Inf.*, vol. 2, no. 1, pp. 16–24, 2018.
- [7] G. Indrawan, P. Sarjana, U. Pendidikan, P. Studi, and I. Komputer, “Penerapan Metode Decision Tree (Data Mining) Untuk Memprediksi Tingkat Kelulusan Siswa Smpn1,” in *Seminar Nasional Vokasi dan Teknologi (SEMNASVOKTEK)*, 2016, pp. 35–44.

Daftar Riwayat Hidup



Arif Amrulloh, lahir di Purbalingga pada tahun 1987. Lulus dan mendapatkan gelar sarjana tahun 2013 dari Prodi Sistem Informasi STMIK Indonesia Jakarta pada konsentrasi Sistem Informasi Akuntansi. Kemudian melanjutkan program Magister dari Prodi Magister Teknologi Informasi Universitas Teknologi Yogyakarta dan sekarang masih tercatat sebagai mahasiswa aktif. Bidang penelitian saat ini adalah Data Mining, *Classification*, *Decission tree* dan *machine learning*.

Alamat Email: arif.amrulloh@student.uty.ac.id



Adityo Permana Wibowo, lahir di Jakarta pada tahun 1987. Lulus dan mendapatkan gelar sarjana tahun 2009 dari Prodi Informatika Universitas Teknologi Yogyakarta pada konsentrasi Sistem Basis Data. Kemudian melanjutkan program Magister dari Prodi Ilmu Komputer Universitas Gadjah Mada, lulus tahun 2015. Bidang penelitian saat ini adalah Data Mining, *Classification*, *neural network*, dan *machine learning*. Beberapa penelitian pernah dilakukan dan dipublikasikan di beberapa jurnal nasional maupun internasional, serta prosiding.

Alamat email : adityopw@staff.uty.ac.id