

OPTIMASI ALGORITMA C4.5 MENGGUNAKAN PARTICLE SWARM OPTIMIZATION (PSO) UNTUK KLASIFIKASI PENYEBAB PERCERAIAN

¹Nafillah

¹Program Studi Teknik Informatika Universitas Yudharta Pasuruan
Pasuruan
e-mail: nafillah@gmail.com

Received : Maret, 2019

Accepted : April, 2019

Published : Juni, 2019

ABSTRACT

The number of divorces that occur in Indonesia is increasing every year, from the data that is known that most divorce cases are carried out by couples under the age of 35 years, the cause also varies from domestic violence, economic factors, infidelity, incompatibility, etc. From the existing divorce data, it is necessary to extract information to predict what factors are causing the divorce. In this study will use the C4.5 algorithm as a classification and prediction method, which will be optimized using the PSO (Particle Swarm Optimization) algorithm to improve the performance of the C4.5 algorithm. From the data mining process, the classification is produced with an accuracy of 87.4%, the accuracy level is higher when compared to the previous research which is only 57.63%.

Keywords: Divorce rate, divorce factor, C4.5 Algorithm, PSO Algorithm, Optimization.

BAB I PENDAHULUAN

Jumlah perceraian yang terjadi di Indonesia meningkat tiap tahunnya, angka perceraian mengalami kenaikan mencapai 16-20 % berdasarkan data yang diperoleh dari tahun 2009 sampai dengan tahun 2016. Dan dari data tersebut dapat diketahui bahwa kebanyakan kasus perceraian dilakukan oleh pasangan yang berusia dibawah 35 tahun, disamping itu, selama 10 tahun terakhir jumlah pernikahan di usia muda mengalami peningkatan, hal itu berbanding lurus dengan meningkatnya angka perceraian [1]. Dilaporkan dari Pengadilan Agama Kabupaten Pasuruan kepada Kantor Urusan Agama (KUA) wilayah Kecamatan Pandaan, kasus perceraian mengalami peningkatan dari tahun 2011 sampai dengan tahun 2015. Dari data perceraian yang ada perlu dilakukan penggalian informasi agar data tersebut bisa bermanfaat, yaitu dengan menggunakan teknik penggalian data yang bertujuan untuk mengklasifikasi faktor yang menjadi penyebab perceraian berdasarkan kriteria tertentu.

Dalam ilmu penggalian data (Data Mining) terdapat banyak sekali metode yang digunakan untuk klasifikasi dan prediksi data, salah satunya yaitu algoritma C4.5, dalam penelitian ini akan menggunakan algoritma C4.5 sebagai metode klasifikasi dan prediksi. Namun, kinerja algoritma C4.5 memiliki beberapa kelemahan dalam melakukan klasifikasi, menurut (Wenefrida Tulit Ina, 2013) hasil klasifikasi algoritma C4.5 kurang maksimal jika jumlah kelas tujuan atau kelas label sangat banyak, namun jika jumlah kelas tujuan sedikit dapat diperoleh klasifikasi dengan hasil baik [4]. Dari kelemahan tersebut maka perlu dilakukan optimasi untuk meningkatkan kinerja algoritma C4.5 dengan menggunakan algoritma optimasi. Algoritma optimasi yang digunakan dalam penelitian kali ini adalah *Particle Swarm Optimization* (PSO) yaitu salah satu algoritma optimasi yang dapat digunakan untuk pengambilan keputusan.

Tujuan dari penelitian ini yaitu untuk: Mengetahui jumlah peningkatan angka perceraian di wilayah Kantor Urusan Agama (KUA) Kecamatan Pandaan, Kabupaten Pasuruan serta factor apa yang dominan terhadap perceraian tersebut.

BAB II TINJAUAN PUSTAKA

2.1. Data Mining

Data mining adalah proses untuk menguraikan penemuan pengetahuan dalam suatu database. Menurut (Turban, dkk. 2005) [10] *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakrit dari berbagai database besar. Data mining memiliki beberapa nama alternatif meskipun memiliki konsep yang berbeda seperti KDD (*knowledge discovery in database*), analisis pola, arkeologi data, pemanenan informasi, dan intelegensia bisnis. Data mining diperlukan saat data yang tersedia terlalu banyak namun tidak diketahui pola apa yang bisa didapatkan.

2.2. Klasifikasi

Klasifikasi merupakan sebuah teknik data mining untuk membagi sekumpulan data kedalam kategori atau kelas tertentu, setiap data akan menjadi anggota dari kelas tertentu sesuai dengan kriteria yang ditentukan. Setiap data yang ada hanya dapat menjadi anggota sebuah kelas saja tidak boleh menjadi anggota lebih dari satu kelas atau tidak menjadi anggota suatu kelas sama sekali.

2.3. Perceraian

Perceraian adalah putusnya ikatan perkawinan antara suami isteri dengan keputusan pengadilan dan ada cukup alasan bahwa diantara suami isteri tidak akan dapat hidup rukun lagi sebagai suami isteri (Soemiyati, 1982:12).

2.4. Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan, pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dipahami dengan mudah melalui bahasa alami, aturan juga dapat diekspresikan dalam bentuk bahas basis data seperti *Structured Query Language* untuk mencari *record* pada kategori tertentu.

Untuk memilih atribut sebagai akar, ditentukan berdasarkan nilai gain tertinggi dari atribut-atribut yang ada. Rumus untuk menghitung gain tertera pada persamaan 1 sebagai berikut.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (1)$$

Untuk menghitung nilai entropi dapat dilihat pada persamaan 2 sebagai berikut.

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (2)$$

2.5. Algoritma Particle Swarm Optimization (PSO)

Algoritma Particle swarm optimization adalah salah satu algoritma optimasi yang dapat digunakan untuk pengambilan keputusan. PSO adalah teknik optimasi dengan cara menghitung terus menerus calon solusi dengan menggunakan suatu acuan kualitas. Algoritma PSO mengoptimasi permasalahan dengan cara menggerakkan partikel atau calon solusi di dalam permasalahan menggunakan fungsi tertentu untuk posisi dan kecepatan dari partikel. Pergerakan partikel dipengaruhi oleh solusi terbaik dari partikel tersebut, dan solusi terbaik secara umum yang didapatkan dari partikel lain. Sekumpulan partikel ini dinamakan swarm, swarm ini akan bergerak menuju solusi terbaik.

$$v_{n+1} = v_n + c_1 rand() * (p_{best,n} - CurrentPosition_n) + c_2 rand2() * (g_{best,n} - CurrentPosition_n)$$

2.6. Classifier Subset Evaluator (CSE)

Classifier subset evaluator (CSE) merupakan sebuah metode untuk pencarian atribut secara otomatis. Metode ini digunakan pada tahap seleksi fitur. CSE mengevaluasi subset atribut pada data latih atau pada rangkaian pengujian terpisah dengan menggunakan calssifier (penggolong) untuk meng-estimasi 'merit' dari satu set atribut

2.7. Weka

Weka merupakan sebuah perangkat lunak *open source* berbasis java yang memiliki berbagai macam algoritma untuk keperluan *data mining*. Weka juga memiliki *tools* untuk pengolahan data seperti *pre-procesising*, *classification*, *regression*, *clustering*, *association rules*, dan *visualization*. Weka dikembangkan oleh Universitas Waikato di Selandia Baru, oleh karena itu nama weka merupakan singkatan dari *Waikato Environment for Knowledge Analysis*. [12]

BAB III METODE PENELITIAN

3.1 Analisa Kebutuhan

1. Kebutuhan perangkat lunak

Perangkat lunak yang dibutuhkan dalam penelitian ini yaitu:

- Sistem Operasi Komputer : Windows 8
- Software data mining : Weka 3.8.1

3.2 Kebutuhan Perangkat Keras

Perangkat keras yang dibutuhkan dalam penelitian ini yaitu:

- Processor: Intel® Core™ i3-2328M CPU @ 2.20GHz 2.20 GHz
- Installed Memory (RAM): 4,00 GB (3,36 GB usable)
- System Type : 32-bit Operating System, x64-based processor
- Harddisk: 500 GB

3.3 Bahan

Bahan yang dibutuhkan dalam penelitian ini yaitu data perceraian yang terjadi di wilayah Kantor Urusan Agama wilayah Pandaan dalam kurun waktu tahun 2015.

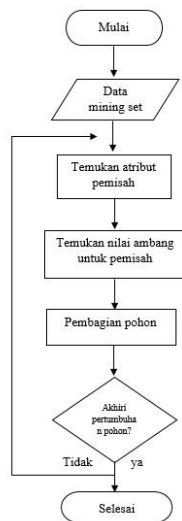
3.4. Metode *Cross-Industry Standard Process for Data Mining (CRISP-DM)*

Cross-Industry Standard Process for Data Mining (CRISP-DM) dikembangkan pada tahun 1996 oleh analis dari beberapa industri seperti DaimlerChrysler, SPSS, dan NCR. CRISP-DM p [13]

Proses data mining menurut CRISP-DM memiliki enam fase, keseluruhan fase tersebut bersifat adaptif. Urutan fase berikutnya bergantung pada keluaran dari fase sebelumnya.

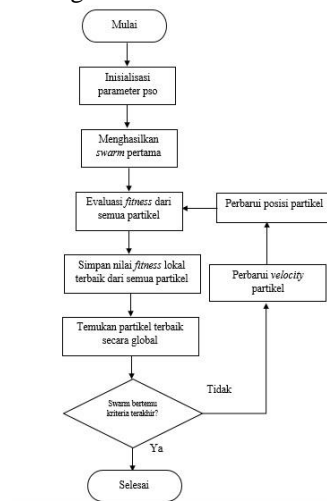
1. Memilih teknik pemodelan

- Algoritma C4.5



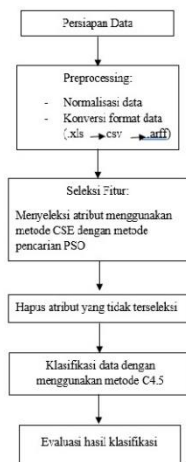
Gambar 1.1 Flowchart Algoritma C4.5

- Algoritma PSO



Gambar 1.2 Flowchart algoritma PSO

2. Membangun model



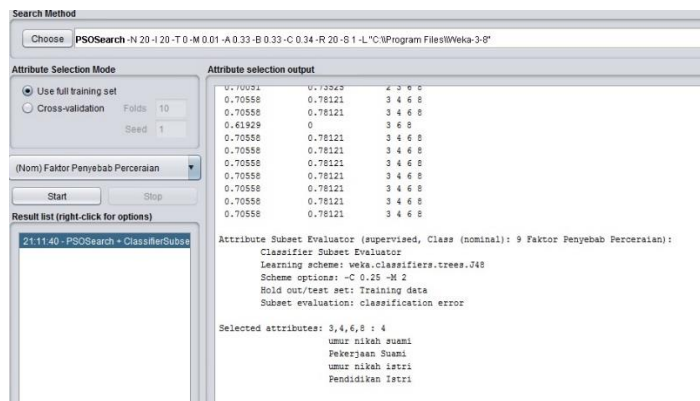
Gambar 1.3 Model Keseluruhan Sistem

BAB IV HASIL DAN PEMBAHASAN

4.1 Pembahasan

Berdasarkan model yang telah dibuat, setelah dataset telah siap digunakan maka langkah selanjutnya yang dilakukan yaitu menentukan kelas tujuan sebelum dilakukannya proses klasifikasi, beberapa kelas tersebut yaitu: (1)Perselingkuhan, (2)KDRT, (3)Ekonomi, (4)Tidak Harmonis. Beberapa kelas tersebut merupakan permasalahan yang menjadi penyebab perceraian.

Sebelum masuk ke tahap klasifikasi dilakukan tahapan seleksi fitur yang berguna untuk mendapatkan data-data dengan korelasi yang sesuai dengan kepentingannya. Dalam tahap ini digunakanlah algoritma Particle swarm Optimization (PSO) untuk menyeleksi atribut mana saja yang relevan dengan hasil yang diharapkan.



Gambar 2.1 Seleksi Fitur Classifier Subset Evaluator dengan metode Pencarian PSO

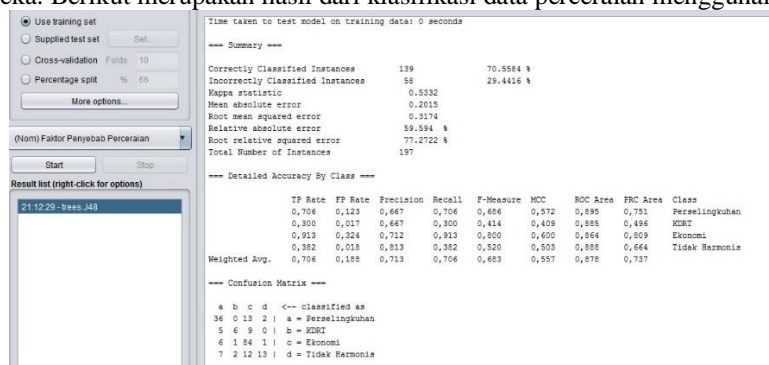
Gambar 2.1 menunjukkan metode seleksi fitur classifier subset evaluator dengan PSO, dimana jumlah fitur yang terseleksi adalah 4 fitur dari 8 fitur yang ada. Itu artinya ada 4 atribut yang tidak terpakai pada dataset perceraian. Berikut tabel 1.1 yang menunjukkan hasil seleksi fitur classifier subset eval dengan PSO.

| Seleksi Fitur | Fitur yang terseleksi dengan CSE dan PSOSearch |
|---------------|--|
| CSE + PSO | Umur nikah suami, pekerjaan suami, umur nikah istri, pendidikan istri. |

Tabel 1.1 hasil seleksi fitur menggunakan CSE+PSO

Langkah kedua yakni menghapus beberapa atribut yang tidak terseleksi pada tahap seleksi fitur, maka kita kembali ke tab preproses untuk menghapus beberapa atribut tersebut.

Langkah ketiga masuk ke tahap klasifikasi menggunakan algoritma C4.5 atau disebut algoritma J48 pada pemrograman Weka. Berikut merupakan hasil dari klasifikasi data perceraian menggunakan algoritma C4.5.



Gambar 2.2 hasil klasifikasi menggunakan algoritma C4.5

4.2 Hasil

4.2.1 Hasil Klasifikasi

| | | |
|----------|-------|-------|
| ROC Area | 0,797 | 0,878 |
| PRC Area | 0,639 | 0,737 |

Tabel 2.2 Perbandingan tingkat akurasi masing-masing model

4.2.3 Perbandingan dengan penelitian sebelumnya

Berikut ini perbandingan hasil penelitian saat ini dengan penelitian sebelumnya yang berjudul **“Implementasi Data Mining Untuk Mengetahui Faktor Penyebab Perceraian Di Wilayah Kecamatan Pandaan Dengan Menggunakan Algoritma C4.5** (Jannatul Firdausi Nuzulia, 2016)” :

| Penelitian Sebelumnya | Penelitian Saat ini |
|--|---|
| Tingkat akurasi sebesar 57,63 % | Tingkat akurasi sebesar 87,4% |
| Atribut yang memiliki nilai Gain tertinggi yang dijadikan akar dari pohon keputusan adalah atribut “Umur Nikah Suami” | Atribut yang memiliki nilai Gain tertinggi yang dijadikan akar dari pohon keputusan adalah atribut “Pekerjaan Suami” |

Tabel 2.3 Perbandingan Penelitian Sekarang Dengan Penelitian Sebelumnya

BAB V KESIMPULAN DAN SARAN

5. Kesimpulan

Dari hasil dari penelitian, pembahasan, dan pengujian model yang telah dilakukan, maka dapat disimpulkan bahwa:

1. Hasil klasifikasi dari 197 data perceraian yang terjadi di kecamatan Pandaan menggunakan algoritma C4.5 berbasis Particle Swarm Optimization menghasilkan akurasi yang lebih baik jika dibandingkan dengan hanya menggunakan algoritma C4.5 saja
2. Penyebab perceraian yang terjadi di kecamatan Pandaan yang paling tinggi disebabkan masalah Ekonomi, Perselingkuhan berada pada urutan nomor 2, kemudian masalah Tidak Harmonis dan yang paling rendah adalah masalah KDRT.
3. Seleksi fitur dengan menggunakan algoritma PSO dapat meningkatkan akurasi dari algoritma C4.5 dalam melakukan klasifikasi dalam hal ini menggunakan data perceraian.
4. Algoritma C4.5 kurang tepat jika digunakan untuk klasifikasi *multiclass* karena menghasilkan akurasi kurang dari 0.9 atau 90%.

5.2 Saran

Dari hasil penelitian yang telah dilakukan terdapat beberapa kekurangan, diharapkan bisa diperbaiki pada penelitian selanjutnya, maka saran yang diajukan adalah sebagai berikut:

1. Menggunakan dataset yang lebih banyak dan lebih detail untuk menghasilkan klasifikasi yang lebih akurat.
2. Menggunakan metode lain yang lebih sesuai untuk klasifikasi *multiclass*.
3. Menggunakan metode yang lebih spesifik untuk mengukur dan memvalidasi akurasi yang dihasilkan, seperti menggunakan kurva AUC atau ROC.

DAFTAR PUSTAKA

- [1] PT. Kabar Berita Semesta, "tingkat perceraian di indonesia termasuk yang tertinggi di dunia," PT. Kabar Berita Semesta, [Online]. Available: <http://www.gulalives.co>. [Accessed 12 Maret 2018].
- [2] U. I. Bandung, "Penanggulangan Perceraian dan kesiapan menikahi orang yang berbeda," Universitas Islam Bandung, 03 agustus 2016. [Online]. Available: <http://www.unisba.ac.id>. [Accessed 12 maret 2018].
- [3] E. t. l. Kusriani, "Algoritma C4.5," in *Algoritma Data Mining*, Yogyakarta, Penerbit Andi, 2009, p. 13.

- [4] W. T. Ina, "Klasifikasi Data Rekam Medis Berdasarkan Kode Penyakit Internasional Menggunakan Algoritma C4.5," *Jurnal Media Elektro*, vol. 1, no. Data Mining, p. 6, 2013.
- [5] D. C. P. Sofi Defiyanti, "Perbandingan Kinerja Algoritma ID3 dan C4.5 Dalam Klasifikasi Spam E-mail," no. Data Mining, p. 5, 2014.
- [6] J. f. & L. h. nuzulia, "Implementasi data mining untuk mengetahui faktor penyebab perceraian di wilayah kecamatan pandaan dengan menggunakan algoritma C4.5," no. Data Mining, pp. 1-8, 2016.
- [7] S. Mujab, "Pencarian model terbaik antara algoritma C4.5 dan C4.5 berbasis Particle swarm optimization untuk prediksi promosi deposito," no. Data Mining, pp. 1-7, 2013.
- [8] M. Y. Kurniawan, "Optimasi Decision tree menggunakan Particle swarm optimization pada data siswa putus sekolah," *JTIULM*, vol. 2, no. Data Mining, pp. 15-22, 2017.
- [9] K. Ma'ruf, "Komparasi pemodelan data menggunakan C4.5 dan C4.5 berbasis Particle swarm optimization untuk memprediksi kelulusan mahasiswa," no. Data mining, pp. 1-8, 2013.
- [10] E. d. Turban, *Decision Support System and Intelligent Systems*, Yogyakarta: Andi Offset, 2005.
- [11] E. Prasetyo, *Data Mining - Konsep dan Aplikasi Menggunakan Matlab*, Yogyakarta: Andi Offset, 2012.
- [12] S. A. & Y. A. Sari, "Pengenalan Weka," in *Implementasi Data Mining Menggunakan Weka*, Malang, UB Press, 2018, pp. 8-9.
- [13] K. & E. T. Lutfi, "Definisi data mining," in *Algoritma Data /mining*, Yogyakarta, Penerbit Andi, 2009, p. 7.
- [14] L. & D. T., *Discovering Knowledge in Data: an Introduction to Data Mining*, Jhon willwy & Sons, Inc., 2005.
- [15] J. f. nuzulia, "Implementasi data mining untuk mengetahui faktor penyebab perceraian di wilayah kecamatan pandaan dengan menggunakan algoritma C4.5," no. Data Mining, 2016.