

Penerapan Data Mining Untuk Akurasi Analisis Cuaca di Australia Menggunakan Algoritma J48 Decision Tree.

Freyza Fachrurrozi Kusuma

Universitas Mercu Buana

*Corresponding author email:
freyza.kusuma.8@gmail.com

Manuscript history:

Diterima 02-06-2023

Direvisi 02-06-2023

Diterima untuk terbit 02-06-2023

ABSTRAK

Cuaca merupakan suatu kondisi udara pada waktu dan wilayah tertentu. Setiap wilayah memiliki kondisi cuaca yang berbeda-beda, tergantung dari letak geografis dan iklim wilayahnya. Cuaca dapat mempengaruhi kegiatan atau aktivitas manusia. Walaupun demikian, cuaca cenderung datang dan berubah-ubah secara tiba-tiba. Hal ini dapat mengakibatkan beberapa orang tidak siap akan datangnya cuaca pada saat itu. Salah satu cara untuk menghindari hal itu adalah dengan cara memprediksi cuaca. Dengan adanya permasalahan tersebut, penulis bertujuan untuk melakukan penelitian untuk mengatasi permasalahan ini secara menyeluruh. Penelitian ini melibatkan data-data kondisi cuaca berdasarkan variabel-variabel tertentu. Data tersebut diambil dari banyak kota di Australia dengan periode tahun 2007 – 2017. Data tersebut dikumpulkan dari sumber yang terpercaya. Data tersebut merupakan data klasifikasi, dengan output variabel target “yes” dan “no”. Sehingga algoritma yang digunakan untuk penelitian ini adalah algoritma J48. Algoritma J48 merupakan versi pengembangan dari algoritma konvensional pohon keputusan atau decision tree yang dipakai di banyak penelitian, yaitu ID3. Algoritma tersebut dinilai sesuai dengan data yang tersedia dan dapat membantu menjawab permasalahan ini. Penelitian ini menggunakan aplikasi Weka.

Kata Kunci: Cuaca, Algoritma J48, Pohon Keputusan, Data Mining, Weka.

I. Pendahuluan

Terkadang cuaca tidak dapat kita perkirakan, walaupun sudah banyak platform untuk memperkirakan cuaca. Selain itu, cuaca cenderung berubah secara drastis bila ada suatu kondisi fenomena alam. Cuaca adalah suatu bentuk permulaan yang diintegrasikan dengan pengertian akan kondisi udara secara eksplisit dan sesaat pada suatu tempat dan waktu tertentu. Dengan kata lain, cuaca merupakan suatu kondisi udara pada waktu dan wilayah tertentu. Perubahan udara yang terjadi pada atmosfer dengan waktu tertentu dan dapat berubah secara tiba-tiba.

Banyaknya parameter dalam menentukan suatu cuaca menyebabkan ketepatan dan kecepatan dalam memprediksi cuaca kurang terpenuhi. (A. Joshi, 2015). Hal ini yang menyebabkan timbulkan suatu metode analisis untuk memprediksi cuaca. Metode tersebut dinamakan metode analisis data mining. Metode analisis data mining yang digunakan untuk penelitian ini adalah metode klasifikasi data. Metode tersebut sesuai dengan data penelitian ini. Dalam metode ini, terdapat beberapa algoritma. Salah satu algoritmanya adalah algoritma J48. Algoritma J48 merupakan versi lanjutan dari algoritma konvensional pohon keputusan yang terkenal di banyak penelitian, yaitu ID3.

Penelitian ini bertujuan untuk mendapatkan nilai akurasi dari dataset cuaca dan memeriksa prediksi dari cuaca, baik cuaca hari ini maupun cuaca besok. Algoritma J48 ini merupakan algoritma C4.5. Penamaan J48 ini berasal dari aplikasi WEKA. WEKA merupakan aplikasi untuk mengolah data mining. Aplikasi ini memberikan kemudahan dalam mengolah data mining yang kami harapkan.

II. Landasan Teori

Berisi tentang kajian teori yang mendukung penelitian ini dilakukan. Teori harus berasal dari buku dan atau artikel di jurnal-jurnal terakreditasi maupun jurnal internasional atau bisa juga dari *proseding*.

A. Data Mining

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan

pengetahuan yang terkait dari berbagai database besar. (Afrisawati, 2013) [1]. Selain itu, data mining adalah analisis observasional sekumpulan data untuk menemukan hubungan tidak terduga dan untuk meringkas data dengan cara baru yang dapat dipahami dan berguna bagi pemilik data (Larose, 2006) [1].

B. Klasifikasi

Klasifikasi adalah proses untuk menemukan suatu himpunan pola yang menggambarkan dan memisahkan kelas-kelas data yang satu dengan yang lainnya. Sehingga dapat menyatakan suatu objek apakah data tersebut berada di kelas yang satu atau berada di kelas lainnya. Secara umum, klasifikasi ini hanya ada dua hasil atau dua kelas, yaitu 0 dan 1 atau “Ya” dan “Tidak”. Klasifikasi merupakan proses penemuan model yang menggambarkan dan membedakan kelas data, atau cara mengklasifikasi data kedalam satu atau beberapa kelas yang sudah didefinisikan sebelumnya [2].

C. Decision Tree

Pohon keputusan adalah struktur berbentuk pohon yang hampir sama dengan diagram alur. Setiap simpul internal dalam pohon tersebut mewakili pengujian terhadap atribut, sedangkan cabang-cabangnya mewakili hasil dari pengujian tersebut. Pada setiap simpul daun, terdapat label kelas yang menunjukkan prediksi untuk suatu data. Dalam penerapannya, pohon keputusan membandingkan nilai atribut data dengan struktur pohon tersebut. Dari akar pohon hingga ke simpul daun, akan diikuti jalur yang membawa ke label kelas yang menjadi prediksi untuk data tersebut. Pohon keputusan dapat dengan mudah diubah menjadi aturan klasifikasi. Model ini sering digunakan dalam statistik, data mining, dan pembelajaran mesin untuk memperkirakan nilai target berdasarkan observasi. Ketika variabel target memiliki himpunan nilai yang terbatas, pohon keputusan tersebut disebut sebagai pohon klasifikasi, di mana simpul daunnya mewakili label kelas, dan cabang-cabangnya mewakili kombinasi fitur yang mengarah ke label-label tersebut. Keunggulan dari pohon keputusan adalah kemampuannya dalam membangun model ini dengan cepat dibandingkan dengan metode klasifikasi lainnya, serta kemampuannya menghasilkan pernyataan query untuk akses basis data yang lebih praktis. Dalam hal akurasi, pohon keputusan umumnya mencapai hasil yang setara atau bahkan lebih baik dibandingkan dengan metode klasifikasi lainnya. Dengan kata lain, pohon keputusan merupakan algoritma pembelajaran supervised non-parametrik, yang mana diutilisasikan untuk klasifikasi dan regresi. Pohon keputusan memiliki bentuk struktur hierarki dan seperti pohon, yang berisikan akar node, cabang-cabang, node internal dan node daun. Metode pohon keputusan dapat menghindari munculnya permasalahan ini dengan menggunakan kriteria yang jumlahnya lebih sedikit pada setiap node internal tanpa banyak mengurangi kualitas keputusan yang dihasilkan [3].

D. Algoritma J48

Algoritma J48 merupakan algoritma C4.5. Algoritma ini merupakan versi terbaru atau lanjutan dari versi algoritma sebelumnya, yaitu algoritma konvensional decision tree. J48 adalah salah satu jenis classifier pada metode klasifikasi dalam data mining dan bagian dari C4.5 decision tree yang sederhana [4]. Algoritma ini membuat pohon keputusan berdasarkan pada sekumpulan data training dengan cara yang sama seperti algoritma ID3 lakukan, dengan menggunakan konsep informasi entropy. Untuk mendapatkan akurasi yang tinggi, atribut yang terbaik untuk dipisah adalah algoritma yang memiliki banyak informasi yang bagus. Penamaan J48 ini berasal dari software data mining yang bernama WEKA. Setiap waktu Java Virtual Machine menjalankan J48, ia akan membuat suatu instansi dari kelas ini dengan mengalokasikan memori untuk membangun dan menyimpan suatu classifier pohon keputusan.

III. Metode Penelitian

A. Tahapan Penelitian

Adapun tahapan penelitian agar penelitian ini berjalan dengan sistematis. Pada awalnya, data diperoleh dari platform tersebut, kemudian di unduh dan diolah menjadi suatu data dengan format arff, format yang dapat dibaca dan kompatibel dengan perangkat lunak Weka. Kemudian diproses

menggunakan algoritma J48 dan melihat hasil dari algoritma tersebut. Gambar 1 menunjukkan alur tahapan penelitian.

B. Data

Data ini diperoleh dari kaggle.com, platform pusat dataset internasional. Data ini diambil dari tahun 2007 sampai tahun 2017 di kota-kota Australia. Sumber data kedua berasal dari modul perkuliahan yang berisi pengetahuan-pengetahuan tentang ilmu data mining.

C. Tabel, Grafik dan Gambar

Berikut sumber data penelitian dan alur tahapan dibawah ini.

1) Tabel

Tabel 1. Sumber Data Penelitian

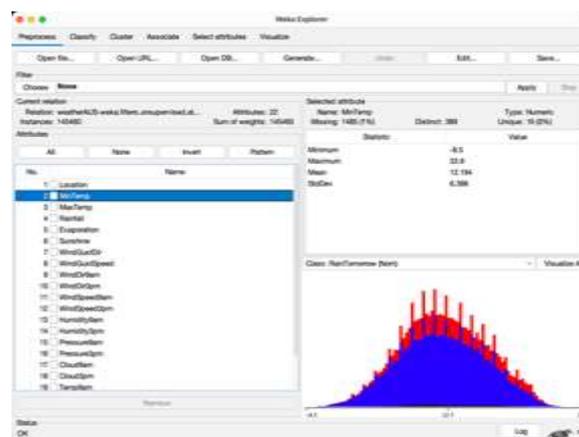
No	Nama	Sumber
1	Kaggle.com	Internet
2	Modul Perkuliahan	Buku Modul

2) Grafik



Gambar 1. Alur Tahapan Penelitian

3) Gambar



Gambar 2. Atribut-Atribut Data

IV. Hasil Dan Pembahasan

A. Hasil

Kami menggunakan dataset “weatherAUS.csv” yang dikonversi menjadi format .arff. Dataset ini memiliki 22 atribut, dimana memiliki variabel target yang berlabel “RainTomorrow”. Setelah

dilakukan pemrosesan algoritma pada aplikasi Weka, hasil didapat bahwa tingkat akurasi data adalah sebesar 84,1893% atau jika dibulatkan sebesar 84,19%, dengan total data sebanyak 58.184 data, 1.330 diantaranya tidak memiliki kelas sama sekali atau null.

B. Pembahasan

Hasil juga menunjukkan bahwa sebanyak 47.865 data diklasifikasikan benar dan 8.989 data diklasifikasikan salah. Hal ini menunjukkan bahwa dengan menggunakan model algoritma ini menghasilkan data yang cukup akurat. Untuk decision tree nya, sebanyak 2.147 node daun terbentuk dan 2.656 node untuk ukuran pohon keputusan tersebut.

V. Kesimpulan

Kesimpulan dari penelitian ini adalah dataset tersebut diambil dari kaggle.com, website untuk seluruh dataset yang ada di dunia. Data tersebut diambil dari tahun 2007 hingga tahun 2017 di kota-kota Australia. Dataset ini menggunakan algoritma J48 menghasilkan tingkat akurasi sebesar 84,19%, yang mana data tersebut cukup akurat. Untuk penelitian lebih lanjut, peneliti mungkin akan mempertimbangkan untuk memperhatikan nilai AUC yang dihasilkan.

Daftar Rujukan

- Elyas, H.A., J. Prayoga (2020), Implementasi Data Mining Pola penjualan Sparepart Motor Honda Pada PT Rotella Persada Mandiri Dengan Menggunakan Algoritma Apriori, *Journal of Information System, Computer Science and Information Technology*, Vol 1 No 1, Juni 2020.
- Saputra, A.Y., Y. Primadasa (2018), Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbour, *Techno.COM*, Vol 17 No 4 Hal 395 – 403, November 2018.
- Syamsu, S., Muhajirin, N.S. Wijaya (2019), Rules Generation Untuk Klasifikasi Data Bakat dan Minat Berdasarkan Rumpun Ilmu, *Jurnal Teknologi Informasi dan Komunikasi*, Vol 9 No 1, Hal 40 – 51, Juni 2019.
- Agustiani, S., A. Mustopa, A. Saryoko, W. Gata, S.K. Wildah (2020). Penerapan Algoritma J48 Untuk Deteksi Penyakit Tiroid, *Paradigma – Jurnal Informatika dan Komputer*, Vol 22 No 2, September 2020.
- Novandya, A. (2017). Penerapan Algoritma Klasifikasi Data Mining C4.5 Pada Dataset Cuaca Wilayah Bekasi. *Konferensi Nasional Ilmu Sosial & Teknologi (KNIST)*.
- A. Joshi, B. K. (2015). Weather Forecasting and Climate Changing Using Data Mining Application Rain Effects on Speed, *Int. J. Adv. Res. Comput. Commun. Eng.*, 19–21.
- Larose, D. T. (2006). *Data Mining Methods and Models*. New Jersey: John Wiley & Sons, inc.
- J. Han, M. K. (2012). *Data Mining Concepts and Techniques 3rd Edition*. USA: Morgan Kauffman. <https://www.cs.waikato.ac.nz/ml/weka/>.