

SIMULASI KLASIFIKASI HUJAN WILAYAH KOTA BANDUNG DENGAN METODE DECISION TREE MENGGUNAKAN ALGORITMA C4.5

Tria Farhatan Musyaffa¹, Hetti Hidayati², Veronikha Effendy³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Penambahan data cuaca yang dihasilkan oleh BMKG terus meningkat, hal tersebut menjadi timbunan data. Penelitian yang dilakukan adalah memanfaatkan data cuaca beberapa tahun ke belakang untuk memprakirakan hujan. Teknik yang dilakukan adalah dengan data mining menggunakan metode pohon keputusan algoritma C4.5. Data masukan yang digunakan adalah data klimatologi BMKG stasiun geofisika kelas 1 Bandung dari tahun 2005 - 2009 . Proses pengujian dilakukan dengan membagi dua data, untuk tahun 2005 - 2008 digunakan sebagai training data dan tahun 2009 digunakan sebagai testing data . Dari hasil proses pengujian yang dilakukan pada algoritma C4.5 tanpa menggunakan pruning didapatkan accuracy hanya sebesar 2,47%, hal ini diakibatkan anomali cuaca pada training data cuaca yang digunakan, sehingga rule yang terbentuk tidak bekerja dengan baik. Sedangkan, pada algoritma C4.5 dengan menggunakan prepruning didapatkan accuracy sebesar 74,24%, pohon keputusan pada algoritma C4.5 menggunakan prepruning dilakukan penyederhanaan dengan memprediksi kemungkinan error pada node - nodenya, sehingga rule yang terbentuk dapat menghasilkan accuracy yang lebih baik.

Kata Kunci : data mining , pohon keputusan , algoritma C4.5

Abstract

The addition of weather data generated by BMKG continues to increase, it becomes a heap of data. The study was conducted utilizing weather data for the past few years predicting rain. The technique is done by using a data mining algorithm C4.5 decision tree method. The input data used is a geophysical station climatological data BMKG 1st class Bandung from 2005 - 2009. The testing process is done by dividing the two data, for the years 2005 - 2008 are used as training data and the year 2009 is used as the testing data. From the results of the tests performed on the algorithm C4.5 without pruning obtained using only the accuracy of 2.47%, this is due to weather anomalies in the training weather data is used, so the rule does not work well formed. Meanwhile, the C4.5 algorithm using prepruning obtained accuracy of 74.24%, C4.5 decision tree algorithm using the simplified prepruning predict the likelihood of errors in nodes, so the rule is formed can produce better accuracy.

Keywords : data mining, decision tree, C4.5 algorithm

Telkom
University

1. PENDAHULUAN

1.1 Latar Belakang Masalah

Dalam proses mendapatkan data cuaca diperlukan alat khusus untuk mengukur berbagai atributnya seperti curah hujan, suhu, kecepatan angin, dan kelembaban udara. Curah hujan diukur dengan menggunakan alat penakar hujan, alat ini digunakan untuk mengukur tinggi rendahnya curah hujan. Sedangkan untuk mengukur suhu, kecepatan angin, dan kelembaban udara, didapatkan dengan menggunakan alat yang bernama AWS (*Automatic Weather Station*), alat ini bekerja untuk mengumpulkan data cuaca. Seiring dengan berjalannya waktu data cuaca tersebut akan bertambah banyak dan menjadi sebuah gunung data. Pertumbuhan pesat dari akumulasi data cuaca dapat memungkinkan kondisi yang disebut "*Rich of Data but Poor of Information*" karena data cuaca yang terkumpul tidak dapat digunakan untuk pengambilan keputusan dan akan menjadi kuburan data. Kemunculan *data mining* dilatarbelakangi oleh adanya masalah data *explosion*, dimana data yang terkumpul hanya sebagai dokumentasi saja. Pertanyaan yang muncul, apakah data cuaca yang terkumpul tersebut akan menjadi dokumentasi saja lalu terabaikan atau dari data cuaca tersebut dapat digali informasi yang berguna untuk prakiraan cuaca akan datang.

Dalam tugas akhir ini, akan melakukan penelitian menggunakan *data mining* untuk memprakiraan cuaca dengan menggunakan data cuaca beberapa tahun ke belakang. Secara garis besar *data mining* adalah untuk mendeskripsikan apa yang telah terjadi (*descriptive data mining*) dan memprediksikan apa yang akan terjadi (*predictive data mining*). Salah satu contoh teknik *descriptive data mining* dan *predictive data mining* adalah klustering dan klasifikasi. Penggunaan klustering dipakai ketika data tidak diketahui bagaimana harus dikelompokkan, pengelompokan data dilakukan dengan membagi keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan. Sedangkan, klasifikasi terdapat informasi bagaimana data dikelompokkan, kemudian dilakukan *training data* pada sistem dengan data yang sudah diberikan label. Selanjutnya, sistem akan mengklasifikasikan data yang baru ke dalam kelompok yang ada dan tidak ada pertambahan kelompok.

Prakiraan cuaca di sini akan menggunakan *predictive data mining*, yaitu dengan klasifikasi. Metode klasifikasi yang akan digunakan adalah *decision tree*, menurut Basuki dan Syarif, proses pada *decision tree* adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi *rule*, dan menyederhanakan *rule*[1]. Algoritma yang digunakan dalam membentuk pohon keputusan adalah C4.5. Pemilihan metode klasifikasi ini dikarenakan melihat data cuaca yang digunakan memiliki informasi bagaimana data ini dapat

dikelompokan berdasarkan curah hujan, dimana nilai curah hujan ini menentukan kategori hujan dan tidak hujan. Sedangkan, algoritma C4.5 digunakan karena kemampuannya dalam menganalisis data pengamatan unsur cuaca, serta algoritma ini juga merupakan algoritma paling berpengaruh dan terbaik dalam komunitas penelitian oleh IEEE pada Desember 2006. Hasil tugas akhir yaitu untuk mengetahui apakah model *decision tree* yang dibentuk oleh algoritma C4.5 digunakan untuk membantu memprakirakan hujan atau tidak hujan berdasarkan pengamatan data cuaca stasiun geofisika kelas 1 Bandung.

1.2 Rumusan Masalah

Dari uraian latar belakang, diperoleh rumusan masalah sebagai berikut :

1. Bagaimana mengimplementasikan metode *decision tree* menggunakan algoritma C4.5 untuk simulasi klasifikasi hujan?
2. Bagaimana tingkat keakurasian algoritma C4.5 dalam mengklasifikasikan hujan berdasarkan data klimatologi BMKG stasiun geofisika kelas 1 Bandung tahun 2009?

1.3 Batasan Masalah

Adapun permasalahan dalam tugas akhir ini dibatasi, diantaranya :

1. Dalam simulasi klasifikasi hujan ini berlaku untuk daerah Bandung.
2. Penggunaan data klimatologi BMKG stasiun geofisika kelas 1 Bandung untuk *training* data menggunakan data tahun 2005-2008, sedangkan *testing* data menggunakan tahun 2009.

1.4 Tujuan

Pembuatan Tugas Akhir ini bertujuan untuk :

1. Membuat implementasi metode *decision tree* menggunakan algoritma C4.5 pada simulasi klasifikasi hujan.
2. Melakukan pengujian akurasi algoritma C4.5 dalam mengklasifikasikan hujan berdasarkan data klimatologi BMKG stasiun geofisika kelas 1 Bandung tahun 2009.

1.5 Metodologi Penyelesaian Masalah

1. Studi Pustaka

Pencarian berupa sumber-sumber yang dapat mendukung dalam pengerjaan Tugas Akhir ini. Sumber-sumber dapat berupa buku bacaan, tugas akhir dan tesis yang berhubungan dengan topik yang diambil, maupun referensi langsung dari internet.

2. Analisis Kebutuhan dan Perancangan Sistem

Tahap ini melakukan analisis kebutuhan sistem yang akan dibangun, kemudian membuat rancangan desain sistem yang telah dianalisis.

3. Implementasi dan Pengujian Sistem

Sistem dibangun berdasarkan rancangan yang telah dibuat, dalam hal ini sistem dibangun dengan menggunakan aplikasi berbasis php dan database MySQL.

4. Analisis Hasil Pengujian

Melakukan analisis pada hasil yang didapat dari pengujian data pada system yang telah dibuat.

5. Penyusunan Laporan Tugas Akhir

Pada tahap ini, melakukan pengambilan keputusan berdasarkan analisis, kemudian membuat dokumentasi dalam bentuk laporan Tugas Akhir.

1.6 Sistematika Penulisan

Dalam penulisan Tugas Akhir in terdiri dari beberapa bagian, yaitu:

BAB I Pendahuluan

Berisi latar belakang, rumusan masalah, batasan masalah, tujuan pembahasan, metodologi penyelesaian masalah, dan sistematika penulisan.

BAB II Landasan Teori

Berisi penjelasan singkat mengenai konsep-konsep yang mendukung dikembangkannya sistem ini.

BAB III Perancangan Sistem

Berisi rincian mengenai perancangan sistem serta implementasi sistem yang dibuat.

BAB IV Pengujian dan Analisis Sistem

Berisi mengenai pengujian yang dilakukan terhadap sistem yang dikembangkan, serta analisis terhadap hasil pengujian.

BAB V Kesimpulan dan Saran

Berisi kesimpulan yang diambil berkaitan dnegan sistem yang dikembangkan serta saran-saran untuk pengembangan lebih lanjut.

5. PENUTUP

5.1 Kesimpulan

Dari pengujian yang dilakukan pada algoritma C4.5 tanpa menggunakan pruning didapatkan *accuracy* hanya sebesar 2,47%, hal ini diakibatkan anomali cuaca pada *training data* cuaca yang digunakan, sehingga *rule* yang terbentuk tidak bekerja dengan baik. Sedangkan, pada algoritma C4.5 dengan menggunakan *prepruning* didapatkan *accuracy* sebesar 74,24%, pohon keputusan pada algoritma C4.5 menggunakan *prepruning* dilakukan penyederhanaan dengan memprediksi kemungkinan *error* pada node-nodenya, sehingga *rule* yang terbentuk dapat menghasilkan *accuracy* yang lebih baik.

5.2 Saran

Saran yang dapat disampaikan adalah sebagai berikut :

1. Pada proses *pruning* dapat menggunakan metode *pruning* lainnya yang memungkinkan meningkatkan *accuracy*.
2. Simulasi prediksi hujan dapat dilakukan dengan metode prediksi lainnya yang dapat menangani nilai kontinu, sehingga dapat menghasilkan prediksi yang lebih akurat dibandingkan dengan menggunakan metode *decision tree*.

DAFTAR PUSTAKA

- [1] Kusrini., Luthfi, Emha Taufiq. 2009. *Algoritma Data Mining*. Penerbit Andi. Yogyakarta.
- [2] <http://repository.gunadarma.ac.id/bitstream/123456789/2240/1/01-02-013.pdf>.
- [3] <http://repository.usu.ac.id/bitstream/123456789/30698/.../Chapter%20II.pdf>.
- [4] Mujiasih, Subekti. 2011. *Pemanfaatan Data Mining Untuk Prakiraan Cuaca*. Pusat Meteorologi dan Maritim BMKG, Jakarta.
- [5] Suyanto. 2007. *Artificial Intelligence : Searching, Reasoning, Planning and Learning*. Penerbit Informatika. Bandung.
- [6] Harsani, Prihastuti., Mulyana, Iyan., Hidayat, Ade Ofik. 2011. *Prediksi Curah Hujan Bulanan Menggunakan Time Series (Single Exponential Smoothing) dan KNN (Studi Kasus : Kabupaten Padang Pariaman)*.
- [7] Oktafia, Dian., Pardede, D. L. Crispina. *Perbandingan Kinerja Algoritma Decision Tree dan Naïve Bayes dalam Prediksi Kebangkrutan*. Universitas Gunadarma Depok, Indonesia.
- [8] Han, Jiawei, dan Micheline Kamber. (2009). *Data Mining: Concepts and Techniques*. www.cs.uiuc.edu/~hanj.
- [9] Quinlan, J. Ross. (1993). *C4.5: Programs for Machine Learning*. California: Morgan Kauffman Publisher.
- [10] Zheng, Jianjun. (2004). *Study on Relationship of Training Data Size to Error Rate and the Performance Comparison for Two Decision Tree Algorithm*. Etd.lib.ttu.edu/these/available/etd-0627200831295019380293/unrestricted/31295019380293.pdf.
- [11] Liu, Bing. *Data Preprocessing*. www.cs.uic.edu/~liub/teach/cs583-fall-05/CS583-data-prep.ppt. Tanggal akses: 23 Oktober 2013.

- [12] Indriyani, Novi. 2009. Penerapan Metode Pohon Keputusan dengan Algoritma C4.5 pada Sistem Penunjang Keputusan dalam Memprakirakan Cuaca Jangka Pendek. Universitas Indonesia Depok, Indonesia.
- [13] Larose D, T. 2005. *Discovering Knowledge in Data : an Introduction to Data Mining*. Jhon Wiley & Sons Inc.
- [14] Defianti, Sofi., dan Pardede, D. L. C. (2008). Perbandingan Kinerja Algoritma ID3 dan C4.5 dalam Klasifikasi *Spam-Mail*. Universitas Gunadarma Depok, Indonesia.

