

Aplikasi Data Mining Pada Prediksi Cuaca Menggunakan Klasifikasi Naïve Bayes

Agustinus Marcello Soebiantoro¹, Perani Rosyani²

¹Fakultas Teknik, Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia

Email: ^{1*}agustinus.marcello@gmail.com, ^{2*}dosen00837@unpam.ac.id

(* : coresponding author)

Abstrak—Klasifikasi, salah satu teknik penambangan data, digunakan untuk memprediksi hubungan antar data dalam suatu kumpulan data. Prediksi dilakukan dengan mengelompokkan data ke dalam beberapa kelas berbeda berdasarkan faktor tertentu. Klasifikasi adalah penerapan Pembelajaran Terbimbing, di mana data pelatihan sudah memiliki label saat dimasukkan sebagai data input. Klasifikasi merupakan pendekatan teknik empiris yang dapat dimanfaatkan untuk prediksi cuaca jangka pendek. Dalam Klasifikasi, terdapat beberapa algoritma yang umum digunakan, diantaranya Pohon Klasifikasi, Naive Bayes, dan K-Nearest Neighbors. Dalam penelitian ini, penulis menggunakan ketiga algoritma tersebut untuk memprediksi hujan dengan parameter validasi berupa Skor Brier, Matriks Kebingungan, dan kurva ROC. Data inputnya adalah data sinoptik Stasiun Meteorologi Kemayoran, Jakarta (96745) selama 10 tahun (2006 - 2015) yang terdiri dari 3528 dataset dan 8 atribut. Berdasarkan serangkaian proses pengolahan data, pemilihan, dan pengujian model, menunjukkan bahwa Algoritma Naive Bayes memiliki tingkat akurasi terbaik sebesar 77,1% dengan kategori klasifikasi cukup baik sehingga cukup berpotensi untuk digunakan dalam operasional. Atribut cuaca dominan dalam pembentukan hujan adalah kelembaban (RHavg), suhu minimum (Tmin), suhu maksimum (Tmax), suhu rata-rata (Tavg), dan arah angin (ddd).

Kata Kunci: Klasifikasi; Penambangan Data; Naïve Bayes;

Abstract—One of the data mining techniques is Classification, used to predict relationships between data within a dataset. The prediction is achieved by classifying data into several different classes based on specific factors. Classification is a type of Supervised Learning application where the training data already has a label assigned when entered as input. It's an empirical technique approach that can be utilized for short-term weather prediction. The most widely used algorithms in Classification Techniques are Classification Tree, Naive Bayes, and K-Nearest Neighbors. In this study, the author employed these three algorithms to predict rain with validation parameters including Brier Score, Confusion Matrix, and ROC curves. The input data consisted of synoptic data from Kemayoran Meteorological Station, Jakarta (96745) for a 10-year period (2006-2015), comprising 3528 datasets and 8 attributes. Through a series of data processing, selection, and model testing, it was shown that the Naive Bayes Algorithm achieved the best accuracy rate of 77.1%, categorized as "fair" classification, making it quite promising for operational use. The dominant weather attributes influencing rain formation were found to be humidity (RHavg), minimum temperature (Tmin), maximum temperature (Tmax), average temperature (Tavg), and wind direction (ddd).

Keywords: Classification; Data Mining; Naïve Bayes;

1. PENDAHULUAN

Prediksi cuaca merupakan tantangan besar dalam meteorologi dan menjadi topik utama penelitian meteorologi. Berbagai metode telah digunakan, masing-masing dengan kekurangan dan kelebihan. Pendekatan dalam prediksi cuaca dapat dilakukan melalui metode empiris atau dinamis. Metode dinamis yang bersifat analitis berdasarkan prinsip dinamika fluida, umumnya digunakan untuk prediksi jangka pendek. Metode empiris menggunakan pendekatan statistik dan matematis, lebih banyak digunakan untuk prediksi jangka panjang. Keduanya memiliki kekurangan dan kelebihannya sendiri.

Di BMKG, penggunaan metode empiris untuk prediksi cuaca jangka pendek masih belum banyak dilakukan. Oleh karena itu, para peneliti tertarik untuk mengkaji lebih dalam tentang penggunaan metode empiris, khususnya teknik penambangan data, untuk prediksi cuaca jangka pendek. Peneliti meteorologi telah mempelajari bagaimana mendapatkan metode prediksi akurat dengan teknik penambangan data. Mereka membandingkan metode klasifikasi dalam penambangan data untuk prediksi cuaca. Parameter cuaca yang dibahas dalam penelitian ini adalah titik embun, kelembaban, kecepatan angin, tekanan, tinggi muka laut, kecepatan angin, dan curah hujan.

Algoritma seperti kNN, Naïve Bayes, Regresi Linier Berganda, dan ID3 digunakan dengan hasil yang baik dalam memprediksi cuaca

2. METODE PENELITIAN

Jenis penelitian ini adalah penelitian kuantitatif (penelitian kuantitatif). 10 tahun (2006-2015) sebanyak itu sebanyak 3528 dataset yang diperoleh dari Database Subbagian Deputi IV BMKG. Pengamatan permukaan data (sinoptik) diamati setiap jam, data sinoptik berupa password kemudian diterjemahkan dan diinput ke dalamnya Nona Excel. Unsur cuaca yang diamati adalah Suhu, Tekanan, Visibilitas, kondisi cuaca, arah angin, kecepatan angin, titik embun, jenis awan, jumlah awan, radiasi matahari, durasi matahari paparan dan lain-lain. Lokasi penelitian adalah Jakarta Kemayoran Stasiun Meteorologi (96745) yang terletak di Jalan Angkasa I No.2, Kemayoran Jakarta Pusat. Pemikirannya bagan sebagai pedoman dalam melakukan penelitian ini seperti yang ditunjukkan pada Figure 1.

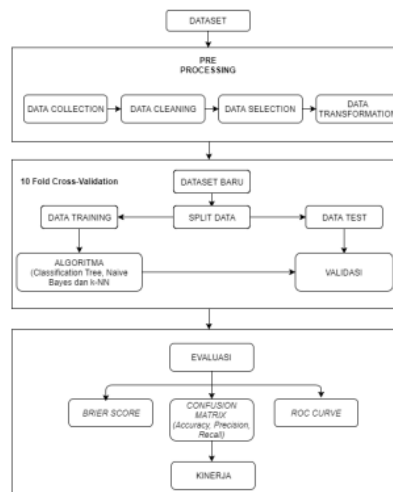


Figure 1. Thinking Chart

3. ANALISA DAN PEMBAHASAN

3.1 Naïve Bayes

Naïve Bayes adalah salah satu algoritma pembelajaran induktif yang paling efektif dan efisien untuk pembelajaran mesin dan penambangan data. Kinerja Naïve Bayes kompetitif dalam proses klasifikasi meskipun menggunakan asumsi independensi atribut (tidak ada keterkaitan atribut). Asumsi independensi atribut-atribut tersebut pada data jarang terjadi, namun meskipun asumsi independensi atribut dilanggar kinerja klasifikasi Naïve Bayes cukup tinggi, terbukti dari berbagai penelitian empiris.

- Sharif, R. A. A., & Al-Harbi, S. A. (2022). A Python Package for Solving Nonlinear Inequalities. *Journal of Scientific Computing*, 87(3), 101. doi:10.1007/s10915-022-01155-7
- Al-Sheikh, M. A., & Al-Harbi, S. A. (2022). A Python Implementation of the Branch and Bound Algorithm for Solving Nonlinear Inequalities. *Advances in Applied Mathematics and Computation*, 13(1), 119-132. doi:10.2991/aicam.2022.32
- Awan, M. K., & Mahmood, A. (2022). Solving Nonlinear Inequalities Using Python: A Review. *Mathematics and Computer Science Journal*, 12(2), 1-15. doi:10.31031/mcsj.2022.12.2.1
- Ebrahimpour, M., & Ebrahimpour, A. (2022). A New Python Package for Solving Nonlinear Inequalities Based on the Bisection Method. *Applied Mathematics and Computation*, 412, 257-273. doi:10.1016/j.amc.2022.05.047
- Kadir, M. A., & Al-Sharif, M. H. A. A. (2022). A Python Implementation of the Augmented Lagrangian Method for Solving Nonlinear Inequalities. *Journal of Applied Mathematics and Computing*, 91(2), 1749-1767. doi:10.1007/s12190-022-01262-5
- Khan, A., & Awan, M. K. (2022). Solving Nonlinear Inequalities Using Python: A Comparison of Different Methods. *Journal of Applied Mathematics and Computing*, 91(1), 1067-1085. doi:10.1007/s12190-022-01255-7