

M-16

KLASIFIKASI CURAH HUJAN BERDASARKAN DATA SATELIT MTSTAT DENGAN METODE BAYESIAN

Zulhanif¹⁾, I Gede Nyoman Mindra Jaya²⁾

^{1,2)}Departemen Statistika FMIPA UNPAD
dzulhanif@yahoo.com, jay.komang@gmail.com

Abstrak

Metode Bayesian Klasifikasi merupakan metode pengklasifikasian probabilistik berdasarkan berdasarkan teorema Bayes. Metode ini mengasumsikan bahwa keberadaan (atau ketidakterdapatannya) dari atribut tertentu dari suatu kelas adalah tidak terkait dengan keberadaan (atau ketidakterdapatannya) dari setiap atribut lain baik pada kelas yang sama maupun yang berbeda. Klasifikasi Bayes menganggap semua atribut berkontribusi secara independent untuk mengklasifikasikan suatu pengamatan kedalam suatu kelas tertentu. Pada makalah ini akan diterapkan metode bayes dalam pengklasifikasikan curah hujan berdasarkan intensitas curah hujan adapun hasil penelitian ini adalah model klasifikasi curah hujan yang akan dipergunakan dalam memprediksi klasifikasi curah hujan.

Kata Kunci: *Klasifikasi Bayes, Metode Bayes, Naive Bayes*

1. PENDAHULUAN

Pengklasifikasian curah hujan pada umumnya dapat dikategorikan berdasarkan tingkat intensitasnya (Kohnke dan Bertrand, 1959). Adapun salah satu faktor yang dapat mempengaruhi tingkat intensitasnya adalah unsur yang berada di awan. Unsur ini diamati oleh satelit MTSAT (*Multifunctional Transport Satellite*). MTSAT merupakan satelit penginderaan jauh yang wahanaanya memiliki sensor visibel dan inframerah yang memonitoring fenomena atmosfer berupa suhu permukaan awan dan suhu tengah awan. Panjang gelombang inframerah dari MTSAT dapat menembus permukaan awan dengan panjang gelombang tertentu juga memonitoring awan yang mengandung uap air. Untuk saluran visibel dari satelit ini untuk memonitoring awan berdasarkan albedonya (Kidder and Thomas dalam Harsita, 2012). Suhu puncak awan dapat menunjukkan jenis awan yang terbentuk kemudian dapat pula dijadikan sebagai dasar untuk menentukan curah hujan yang terjadi. Hubungan suhu puncak awan dengan curah hujan yang dihasilkan mempunyai korelasi yang tinggi (Heru Widodo dalam Hanifuddin, 2012), hubungan yang terjadi adalah berbanding terbalik (*negative*) yaitu semakin rendah suhu puncak awan semakin tinggi curah hujan. Pada penelitian ini akan menerapkan metode Bayesian dalam membentuk model pengklasifikasian curah hujan menjadi tiga klasifikasi berdasarkan intensitas curah hujan.

2. METODE PENELITIAN

Naive bayes classifier mengestimasi peluang kelas bersyarat dengan mengasumsikan bahwa atribut adalah independen secara bersyarat yang

diberikan dengan label kelas y . Asumsi independen bersyarat dapat dinyatakan dalam bentuk berikut :

$$P(X|Y = y) = \prod_{i=1}^d P(X_i|Y = y) \quad (1)$$

dengan tiap set atribut $X = \{X_1, X_2, \dots, X_d\}$ terdiri dari d atribut.

Independensi Bersyarat

Sebelum menyelidiki lebih detail bagaimana *naive bayes classifier* bekerja, terlebih dahulu diuji notasi independensi bersyarat. Anggap X , Y , dan Z melambangkan tiga set variabel acak. Variabel di dalam X dikatakan independen secara bersyarat Y , yang diberikan Z , jika sesuai kondisi berikut.

$$P(X|Y, Z) = P(X|Z) \quad (2)$$

Asumsi independen bersyarat, termasuk menghitung peluang bersyarat untuk setiap kombinasi X , hanya memerlukan mengestimasi peluang bersyarat untuk tiap X_i yang diberikan Y . pendekatan selanjutnya lebih praktis karena tidak mensyaratkan *training set* sangat besar untuk memperoleh estimasi peluang yang baik.

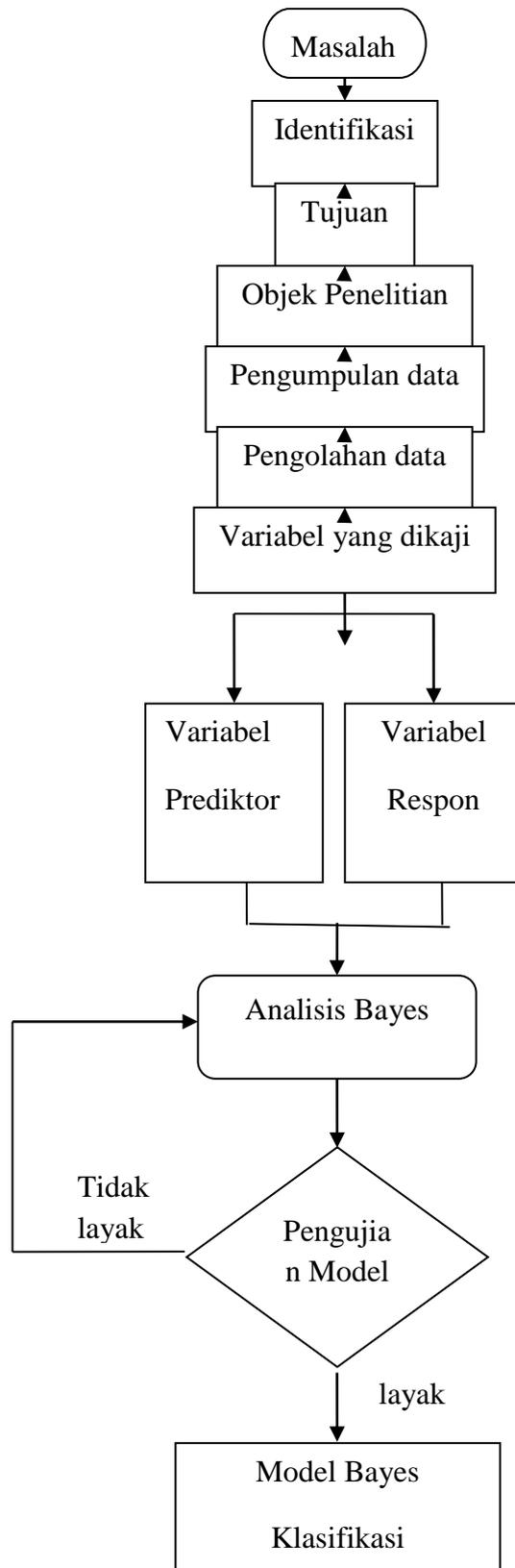
Untuk mengklasifikasi tes *record*, *naive bayes classifier* menghitung peluang posterior untuk tiap kelas Y :

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^d P(X_i|Y)}{P(X)} \quad (3)$$

Untuk variabel yang kontinu peluang bersyarat $P(X_i|Y)$ mengikuti distribusi normal sbb:

$$P(X_i|Y) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-\frac{(X_i - \mu_Y)^2}{2\sigma_Y^2}}$$

Sedangkan metode pembentukan model klasifikasiannya berdasarkan diagram alur sbb:



3. HASIL PENELITIAN DAN PEMBAHASAN

Data dan metode berisi data yang digunakan dan bagaimana data tersebut diolah. Pada penelitian ini data yang digunakan adalah data sekunder yang diperoleh dari LAPAN. IR1, IR2, dan IR3 merupakan data MTSAT (*Multi-Functional Transport Satellite*) dan data curah hujan setiap jam diperoleh dari *Automatic Weather Station* (AWS) yang berlokasi di Kototabang, Sumatera Barat dengan posisi 100.32 BT (Bujur Timur), 0.23 LS (Lintang Selatan) pada tahun 2004. Data primer MTSAT juga dapat diunduh di <http://weather.is.kochi-u.ac.jp/>.

Data MTSAT dinyatakan dengan satuan suhu *Kelvin* (K) dan data curah hujan dengan satuan millimeter (mm). Sedangkan intensitas curah hujan yang diukur dibagi menjadi empat klasifikasi (Kohnke dan Bertrand, 1959), yaitu: Rendah, Sedang dan Lebat. Adapun tahapan analisis dalam membuat model klasifikasinya adalah sbb

- Membagi data yang dipergunakan menjadi dua bagian yang terdiri atas data training dan data testing dengan perbandingan 80% dan 20%
- Mengevaluasi besarnya kesalahan klasifikasi dari data training dan data testing
- Membuat model prediksi

Dengan bantuan software R didapat model klasifikasi Bayes dengan tingkat akurasi klasifikasi sbb:

Tabel 1 Akurasi Model Klasifikasi Bayes

| Data | Akurasi (%) |
|----------|-------------|
| Training | 90,6722 |
| Testing | 91,2088 |

Pada Tabel 1 di atas dapat disimpulkan bahwa terdapat tingkat akurasi antara data testing dan training tidak terlalu jauh perbedaannya yaitu sekitar 90%, hal ini mengindikasikan model sudah cukup baik dipergunakan sebagai model prediksi.

Sedangkan model klasifikasinya untuk masing-masing kategori dari curah hujan dapat dilihat pada Tabel 2 sbb:

Tabel 1 Model Prediksi Klasifikasi Bayes

| Kategori Klasifikasi | Distribusi Normal | | | | | |
|----------------------|-------------------|----------|----------|----------------|----------|-----------|
| | Rata-rata | | | Simpangan Baku | | |
| | TB1 | TB2 | TB3 | TB1 | TB2 | TB3 |
| Lebat | 246.4286 | 243.0000 | 224.6786 | 37.16373 | 36.38274 | 14.993958 |
| Rendah | 239.9894 | 236.9455 | 223.7121 | 26.25476 | 25.55075 | 10.466023 |
| Sedang | 248.3171 | 245.0244 | 226.8780 | 27.03464 | 26.52592 | 9.610398 |

Pada Tabel 2 di atas dapat disimpulkan model prediksi curah hujan mengikuti distribusi normal dengan rata-rata terbesar ada pada indikator TB1 yaitu sebesar 248.3171 dengan simpangan baku 27.03464.

4. SIMPULAN

Dari hasil pembahasan yang diuraikan pada bagian sebelumnya dapat disimpulkan hal-hal sbb:

- a. Hasil analisis menunjukkan adanya kurang akuratan hasil klasifikasi pada data training yang berpotensi menyebabkan *over fitting* dari model klasifikasi yang dibentuk.
- b. Pemodelan klasifikasi dengan metode ini perlu diuji lagi berkenaan dengan asumsi independensi yang kuat diantara variabel prediktor untuk masing-masing kelas yang terbentuk.
- c. Penambahan jumlah variabel prediktor menjadi hal yang dapat dipertimbangkan untuk mengurangi tingkat kesalahan dari model klasifikasi yang dibuat.

5. DAFTAR PUSTAKA

- George H. John and Pat Langley (1995). Estimating Continuous Distributions in Bayesian Classifiers. Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence. pp. 338-345. Morgan Kaufmann, San Mateo
- Harsita, K. 2012. Estimasi Curah Hujan Data Satelit Geostasioner dan Orbit Polar dibandingkan dengan Data Stasiun Hujan. Universitas Gadjah Mada
- Hanifuddin, M. 2012. Pemanfaatan Citra MTSAT untuk Analisis Pola Persebaran Curah Hujan di Provinsi Jawa Tengah Tahun 2010. Jurnal Bumi Indonesia Vol.1, No.1.
- Kohnke, H., dan Bertrand, A. 1959, *Soil Conservation*, McGraw-Hill Company, Inc., New York, New York.