

PERAMALAN POTENSI KEBAKARAN HUTAN DENGAN METODE DATA MINING

Ermatita, Sukemi, Yudha Pratomo

Program Studi Magister Informatika
Fakultas Ilmu Komputer Universitas Sriwijaya
ermatitaz@yahoo.com

Abstrak

Kebakaran hutan merupakan kejadian yang selalu berulang setiap tahun, terutama di Sumatera Selatan. Penanggulangan kebakaran hutan ini, terutama di lahan gambut, perlu dianalisis agar dapat diprediksi pola penyebaran kebakaran lahan gambut. Hal ini perlu dilakukan karena pemadaman kebakaran lahan gambut sulit dilakukan. Penelitian ini memproses data faktor-faktor yang memberikan potensi kebakaran hutan menerapkan data mining. Metode yang digunakan dalam penelitian ini. Hasil dari penelitian ini menunjukkan bahwa faktor-faktor penyebab dari kebakaran hutan yang diproses mempunyai potensi yang sama dalam menyebabkan kebakaran hutan.

Kata kunci : Data Mining, Kebakaran Hutan

1. PENDAHULUAN

Kerusakan lingkungan dapat disebabkan oleh kebakaran. Kebakaran ini disebabkan oleh beberapa factor antara lain: praktik umum pembakaran lahan pertanian, penggundulan hutan, dan musim hujan yang tertunda (karena kondisi iklim yang paling tidak biasa dalam 20 tahun terakhir, seperti El Nino). Pembakaran lahan adalah cara yang cepat dan mudah dalam menyiapkan tanah untuk menanam benih yang baru. Pada Tahun 2015, Sumatra Selatan menjadi provinsi penyumbang utama kabut asap kepada negara-negara tetangga akibat kebakaran yang sebagian besar terjadi di lahan gambut (gambut adalah tanah yang terdiri dari sebagian bahan tanaman yang membusuk terbentuk di daerah sawah), yang cenderung menghasilkan api di bawah tanah yang bersifat lama, berasap, dan sangat besar.

Efek dari kabut asap ini memiliki dampak terhadap kesehatan masyarakat luas, juga berkontribusi terhadap penyakit pernafasan dan kematian dini di seluruh Asia Tenggara. Di Singapura, media cetak maupun elektronik memberikan *update* hampir setiap jam mengenai bahaya terkena polusi apabila berada di luar ruangan.

Penanggulan kebakaran hutan ini telah dilakukan oleh pemerintah dalam berbagai usaha. Penanggulangan kebakaran hutan, terutama kebakaran hutan yang terjadi di lahan gambut. Kebakaran hutan di lahan gambut sulit untuk di lihat secara kasat mata. Karena kebakaran ini kadang terjadi pada bagian lapisan bawah tanah gambut. Untuk itu perlu prediksi/peramalan dan deteksi dengan menerapkan data mining.

2. METODOLOGI PENELITIAN

Pada penelitian ini akan dilakukan pencampuran antara pengalaman, nilai, informasi kontekstual dan penglihatan para ahli yang menyediakan kerangka evaluasi yang berkaitan dengan pengalaman baru dan informasi.

Pengembangan model prediksi dilakukan dengan :

- a) Menentukan kriteria dari solusi penanggulangan kebakaran hutan
- b) Mengembangkan model prediksi kebakaran hutan pada lahan gambut

2.1.1. Support Vector Machine (SVM)

Menurut Sembiring SVM menggunakan model linier untuk mencari hyperplane terbaik sebagai pemisah dua buah kelas pada vector input. *Hyperplane* terbaik dapat ditentukan dengan menghitung nilai margin *hyperplane*. Margin adalah jarak antara *hyperplane* dengan pattern terdekat dari masing-masing kelas. *Pattern* yang paling dekat dengan margin maksimum hyperplane inilah yang disebut dengan *support vector*. Jika kedua kelas -1 dan +1 dan hyperplane berdimensi d didefinisikan sebagai:

$$\vec{w} \cdot \vec{x} + b = 0 \dots\dots\dots (1)$$

Pattern \vec{x}_i pada sampel negative (-1) dan positif (+1) dapat dirumuskan:

$$\vec{w} \cdot \vec{x} + b \leq -1 \dots\dots\dots (2)$$

$$-\vec{w} \cdot \vec{x} + b \geq +1 \dots\dots\dots (3)$$

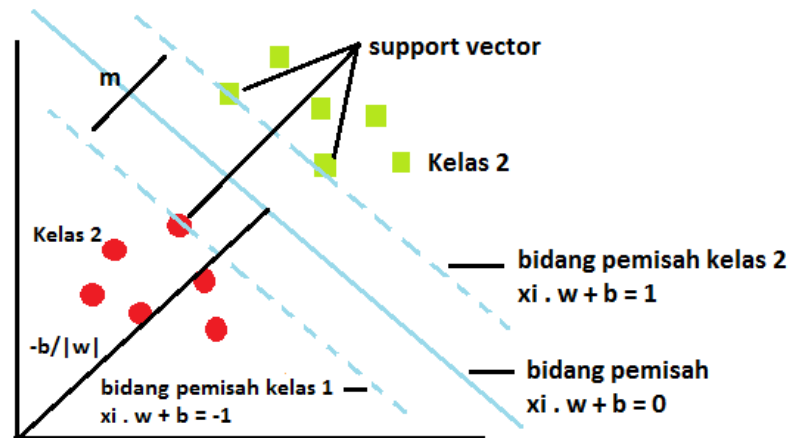
Pemrograman kuadratik digunakan untuk mencari nilai margin terbesar, yaitu $\frac{1}{\|\vec{w}\|}$ dengan mencari titik minimal.

$$\min_{\vec{w}} \tau(w) = \frac{1}{2} \|\vec{w}\|^2 \dots\dots\dots (4)$$

Dengan menggunakan pengali Lagrange maka bentuk primal pemrograman kuadratik dapat diubah menjadi bentuk dual dengan persamaan sebagai berikut:

$$L(\vec{w}, b, \alpha) = \frac{1}{2} \|\vec{w}\|^2 - \sum_{i=1}^l \alpha_i (y_i (\vec{x}_i \cdot \vec{w} + b) - 1) \dots\dots\dots (5)$$

Dimana ($i = 1, 2, \dots, l$) dan α_i adalah Lagrange multipliers yang bernilai 0 atau positif. Nilai optimal dari persamaan diatas dapat dihitung dengan meminimalkan L terhadap \vec{w} dan memaksimalkan L terhadap α_i . Data yang berkorelasi dengan α_i positif disebut dengan support vector.



Gambar 1. Konsep Support Vector Machine (SVM)

Metode pembelajaran mesin lain untuk meningkatkan akurasi prediksi dan sekaligus untuk menghindari data yang *overfitting* adalah metode Support Vector Machine (SVM). Tujuan awal dari SVM adalah untuk menyelesaikan masalah dalam mengklasifikasikan pola.

3. PEMBAHASAN

Support Vector Machine (SVM) adalah metode dalam data mining untuk mendapatkan prediksi data yang di ambil. Karakteristik SVM tidak hanya diperuntukan agar proses klasifikasi menjadi lebih baik, tetapi juga agar menghasilkan generalisasi yang baik dari data pelatihan.

Hasil penelitian telah mendapatkan data dari sumber-sumber yang ada. Ada banyak data yang didapat berkenaan dengan factor-faktor dari lahan gambut ini. Namun Model ini akan memproses data yang disimulasikan dalam matrik 3 x 3 sebanyak 10 sampel diambil dari data sebenarnya.

No	Temperature	Rainfall	Groundwater Level
1	30.4	0.0	0.29
2	31.6	0.0	0.31
3	30.6	10.0	0.31
4	29.8	18.0	0.28
5	30.1	0.0	0.30
6	29.4	2.0	0.31
7	28.2	24.5	0.30
8	30.5	0.0	0.28
9	27.9	39.0	0.26
10	29.4	0.0	0.23

weights (not support vectors):

- 0.0288 * (normalized) Time=1/1/2017

- 0.191 * (normalized) Time=1/2/2017
- 0.1881 * (normalized) Time=1/3/2017
- + 0.0273 * (normalized) Time=4-Jan-17
- 0.0958 * (normalized) Time=5-Jan-17
- 0.1582 * (normalized) Time=6-Jan-17
- 0.0963 * (normalized) Time=7-Jan-17
- + 0.039 * (normalized) Time=8-Jan-17
- + 0.1734 * (normalized) Time=9-Jan-17
- + 0.4126 * (normalized) Time=10-Jan-17
- + 0.2805 * (normalized) Time=11-Jan-17
- + 0.1925 * (normalized) Time=12-Jan-17
- + 0.1001 * (normalized) Time=13-Jan-17
- + 0.5328 * (normalized) Time=14-Jan-17
- + 0.3792 * (normalized) Time=15-Jan-17
- + 0.3406 * (normalized) Time=16-Jan-17
- + 0.1969 * (normalized) Time=17-Jan-17
- + 0.0397 * (normalized) Time=18-Jan-17
- 0.044 * (normalized) Time=19-Jan-17
- 0.0161 * (normalized) Time=20-Jan-17
- 0.0337 * (normalized) Time=21-Jan-17
- 0.0394 * (normalized) Time=22-Jan-17
- 0.1402 * (normalized) Time=23-Jan-17
- + 0.1221 * (normalized) Time=24-Jan-17
- 0.0164 * (normalized) Time=25-Jan-17
- 0.1469 * (normalized) Time=26-Jan-17
- 0.2253 * (normalized) Time=27-Jan-17
- 0.256 * (normalized) Time=28-Jan-17
- 0.35 * (normalized) Time=29-Jan-17
- 0.4438 * (normalized) Time=30-Jan-17
- 0.4422 * (normalized) Time=31-Jan-17
- 0.1887 * (normalized) Time=1-Feb-17
- + 0.4358 * (normalized) Time=2-Feb-17
- + 0.176 * (normalized) Time=3-Feb-17
- + 0.0435 * (normalized) Time=4-Feb-17
- 0.0067 * (normalized) Time=5-Feb-17
- 0.0877 * (normalized) Time=6-Feb-17

- 0.013 * (normalized) Time=7-Feb-17
- 0.1694 * (normalized) Time=8-Feb-17
- 0.1139 * (normalized) Time=9-Feb-17
- + 0.0835 * (normalized) Temperature
- + 0.0771 * (normalized) Rainfall
- + 0.3945

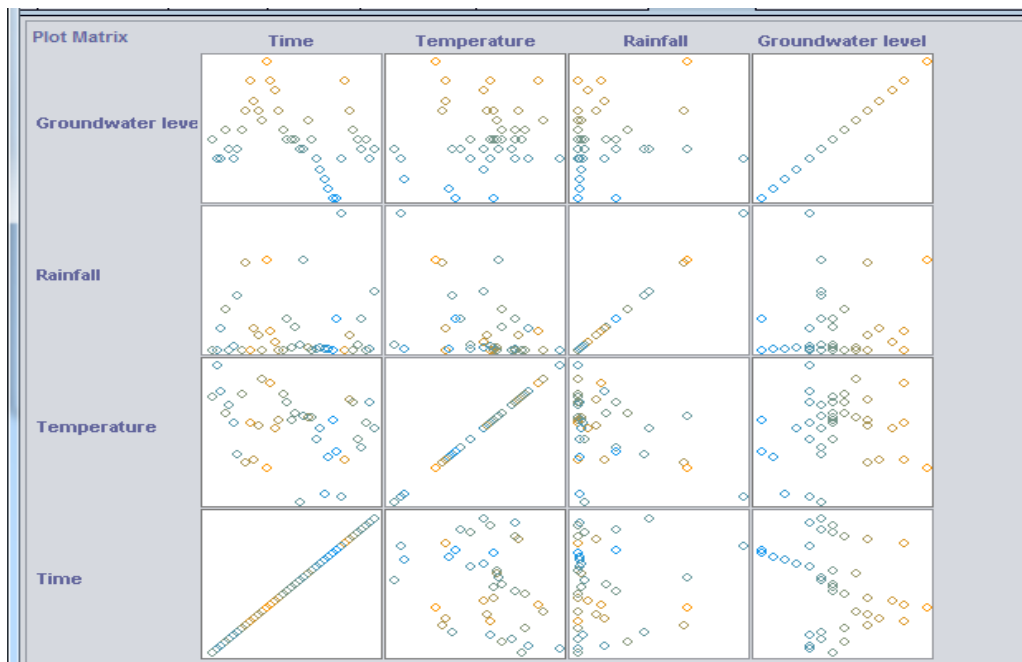
Number of kernel evaluations: 820 (90.855% cached)

Time taken to build model: 0.03 seconds

== Cross-validation ==

=== Summary ===

Correlation coefficient	-0.2977
Mean absolute error	0.0264
Root mean squared error	0.0338
Relative absolute error	104.004 %
Root relative squared error	104.167 %
Total Number of Instances	40



Gambar 2. Hasil Proses Suport Vector Machine

Gambar 2 di atas menunjukkan hasil dari proses dalam pemodelan dengan Support Vector Machine yang merupakan metode dalam data mining. Hasil dari proses tersebut diatas menunjukkan garis dan data dalam dua dimensi. Pada gambar diatas menunjukkan terdapat empat label, yaitu Ground level, Rainfall, Temperature, dan Time. Setiap titik berisi informasi dari ke empat label tersebut dengan rentang kelas warnanya adalah -0,35 (biru) sampai -0,21 (orange). . Garis pada bagian tengah

merupakan batas pemisah antara label atau sebagai pengelompokan yang disebut dengan *hyperplane*. Secara garis besar, yang dilakukan SVM adalah mengambil data dan menarik garis (atau hyperplane) dan membagi kumpulan data ke dalam kelompok pengamatan positif dan negatif. Yang mengindikasikan bahwa keempat factor yang dapat menyebabkan kebakaran hutan ini secara bersama-sama mempunyai kekuatan yang sama dalam potensi penyebab kebakaran hutan.

4. KESIMPULAN

Penggunaan Data mining telah banyak dilakukan oleh berbagai penelitian, hasil dari penelitian ini telah memodelkan prediksi kebakaran hutan berdasarkan factor-faktor yang ada. Hasil dari penelitian ini berupa data-data factor penyebab kebakaran hutan telah di proses dengan menggunakan data mining. Metode Support Vector Machine telah mampu memproses data factor dalam penyebab kejadian kebakaran hutan.

Referensi

- Admin. "Ini Lokasi Titik Api Penyebab Kebakaran Hutan di Sumatera." 24 Agustus 2015. (<http://blog.act.id/ini-lokasi-titik-api-penyebab-kebakaran-hutan-di-sumatera>).
- Varagur, Krithika. "Indonesia Forest Fires Intractable Problem, Despite Efforts". August 11st, 2017. (<https://www.voanews.com/a/indonesia-forest-fires-intractable-problem-despite-efforts/3982129.html>).
- Box and Jenkins. *Forecasting and Time Series Analysis Using the Statistical System*. UTS Module, May 1986. Lijuan Cao, *Support Vector Machine expert for time series forecasting*, Elsevier, *Neurocomputing*, vol.51, pp. 321-339, 2003.
- Dennis, Rona. *A Review of Fire Projects in Indonesia (1982-1998)*. 1998. prepared for CIFOR, ICRAF, UNESCO, EC JRS Ispra
- Dr. Ir. H. Dodi Supriadi dan Dr. Karl-Heinz Steinmann. *Summary Report of Results and Achievements 2003-2007*. July 2007. South Sumatra Forest Fire Management Project.