

KOMPARASI ALGORITMA LR, K-NN DAN SVM UNTUK ESTIMASI AREA KEBAKARAN HUTAN

Fitriyani¹⁾, Ranga Sanjaya²⁾
Prodi. Sistem Informasi¹⁾²⁾
Universitas BSI Bandung^{1),2)}
fitriyani.fyn@bsi.ac.id¹⁾, ranga.rsj@bsi.ac.id²⁾

Abstrak

Kebakaran hutan menimbulkan berbagai permasalahan seperti asap yang dapat mengganggu sistem pernapasan, kerusakan lingkungan dan bencana lainnya. Kebakaran hutan juga dapat berdampak pada biaya yang akan dikeluarkan untuk menyelesaikan masalah yang timbul akibat kebakaran hutan, sehingga diperlukan penelitian untuk mengukur tingkat radiasi api pada area yang terbakar. Algoritma LR (Linear Regression), K-NN (K-Nearest Neighbor) dan SVM (Support Vector Machine) merupakan metode untuk regresi dan klasifikasi. Pada penelitian ini dilakukan perbandingan atau komparasi untuk mendapatkan algoritma terbaik dalam estimasi area kebakaran hutan.

Kata Kunci : Estimasi, Kebakaran Hutan, Linear Regression, K-Nearest Neighbor, Support Vector Machine.

I. PENDAHULUAN

Kebakaran hutan merupakan salah satu masalah lingkungan yang sering terjadi dan menimbulkan dampak negatif seperti nyawa manusia yang terancam karena asap dari kebakaran hutan, sumber daya yang habis terbakar dan kerusakan lingkungan. Dampak dari kebakaran hutan adalah biaya yang sangat tinggi untuk memperbaiki masalah yang ditimbulkan. Berbagai studi dilakukan untuk mengurangi dampak yang ditimbulkan akibat kebakaran hutan. Akan tetapi masih terdapat kendala pada keterbatasan pemahaman tentang dampak negatif kebakaran hutan terhadap ekosistem, serta sulitnya melakukan analisis penyebab kebakaran hutan. Disamping itu, antisipasi terhadap meluasnya area yang terbakar juga masih sulit dilakukan karena keterbatasan informasi terhadap indikator-indikator penyebab meluasnya area yang terbakar. Sehingga mengakibatkan ketidakpastian dan lambatnya penanganan terhadap kebakaran hutan^[1].

Disamping dampak negatif terhadap lingkungan dan kesehatan, asap dari kebakaran hutan dapat mengganggu aktivitas sehari-hari. Asap yang ditimbulkan oleh kebakaran dapat membatasi jarak pandang, terutama untuk transportasi penerbangan^[2]. Kebakaran hutan

membutuhkan biaya yang sangat besar terukur dari hilangnya nyawa manusia, sumber daya, kerusakan lingkungan dan biaya untuk perbaikan daerah yang terkena bencana^[3]. Area yang terbakar harus dilakukan estimasi untuk menentukan seberapa kuat radiasi api pada vegetasi disekitarnya^[4] sehingga dapat memberikan informasi mengenai kerusakan lahan yang terjadi akibat kebakaran hutan.

Penelitian terhadap estimasi area kebakaran hutan yang sudah dilakukan diantaranya dengan menggunakan metode komputasi antara lain Linear Regression^[5], K-Nearest Neighbor^[6], Support Vector Machine^[7]. Pada penelitian ini menggunakan dataset kebakaran hutan yang merupakan dataset publik dari UCI Repository. Alasan penggunaan dataset publik pada penelitian ini dikarenakan dataset publik dapat dengan mudah diperoleh oleh peneliti lain sehingga hasil penelitian dapat dibandingkan secara langsung. Disamping itu, prosentase penggunaan dataset publik untuk penelitian lebih besar dibandingkan dengan penggunaan dataset privat^[8].

Pada penelitian ini akan digunakan tiga algoritma pembelajaran, yaitu algoritma Linear Regression (LR), K-Nearest Neighbor (K-NN) dan Support Vector Machine (SVM), kemudian dari ketiga algoritma ini di cari algoritma

terbaik untuk model estimasi area kebakaran hutan. K-Nearest Neighbor merupakan algoritma regresi yang cukup populer [6]. Sedangkan SVM merupakan algoritma yang baik untuk mengatasi masalah klasifikasi dan regresi baik linier maupun nonlinier [9] [10].

II. KAJIAN LITERATUR

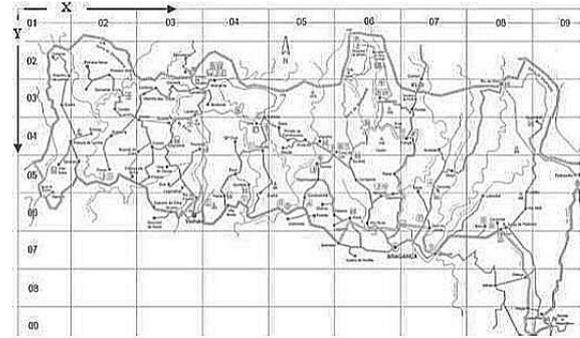
2.1. Kebakaran Hutan

Kebakaran hutan adalah peristiwa dimana wilayah yang terdapat banyak pohon, semak, paku-pakuan dan rumput mengalami perubahan bentuk yang disebabkan pembakaran yang besar-besaran. Kebakaran hutan menyebabkan hutan dilanda api sehingga membuat ekosistem hutan menjadi rusak bahkan hilang. Dampak yang disebabkan kebakaran hutan dapat berupa positif dan negatif tetapi dampak negatif melebihi dampak positif [11].

Dampak yang ditimbulkan kebakaran hutan sangat kompleks. Kebakaran hutan tidak hanya berdampak terhadap ekologi dan mengakibatkan kerusakan lingkungan, akan tetapi dampak dari kebakaran hutan ternyata mencakup bidang lainnya [12], salah satunya dampak ekonomi yang tak terhitung akibat rusak atau hilangnya sumber daya hutan [13].

2.2. Dataset

Dataset kebakaran hutan terdiri dari 517 record dan 12 atribut, dimana 12 atribut ini sebagai variabel prediktor dan 1 atribut sebagai variabel target. Atribut ini terdiri dari sumbu x, sumbu y, bulan, hari, data meteorologi (temp, relative, humadity, wind, rain) dan 4 indeks cuaca kebakaran yaitu fine fuel moisture code (FFMC), duff moisture code (DMC), drought code (DC), dan initial spread index (ISI). FFMC merupakan atribut indikator untuk bahan yang mudah terbakar. DMC merupakan atribut yang mengukur konsumsi bahan bakar pada kayu. DC merupakan indikator dari efek kemarau pada bahan bakar hutan. Sedangkan ISI merupakan indikator tingkat penyebaran api dan indikator kesulitan pengendalian api. Pada Gambar 1 dapat dilihat peta alam Taman Montesinho di Portugal yang menjadi sumber referensi dataset kebakaran hutan yang diteliti. Pada Tabel 1 dapat dilihat deskripsi atribut dari dataset kebakaran hutan.



Gambar 1. Peta Alam Taman Montesinho

Tabel 1. Deskripsi Dataset Kebakaran Hutan

Attribute	Description
X	x-axis coordinate (from 1 to 9)
Y	y-axis coordinate (from 1 to 9)
Month	Month of the year (January to December)
Day	Day of the week (Monday to Sunday)
FFMC	FFMC code
DMC	DMC code
DC	DC code
ISI	ISI index
Temp	Outside temperature (in °C)
RH	Outside relative humidity (in %)
Wind	Outside wind speed (in km/h)
Rain	Outside rain (in mm/m2)
Area	Total burned area (in ha)

2.3. Linear Regression

Linear Regression atau Regresi Linear menemukan garis terbaik dari 2 atribut atau variable, sehingga salah satu atribut tersebut dapat memprediksi atribut yang lain. Model regresi dan log-linear dapat digunakan untuk memperkirakan data yang diberikan. Sederhananya data Linier Regression dimodelkan agar sesuai dengan garis lurus, contohnya secara acak variabel y (disebut variabel respon) dapat dimodelkan sebagai fungsi linier dari yang lain variabel acak x (disebut variabel prediktor) dengan persamaan:

$$y = wx + b$$

Dimana varians y di asumsikan kontans. Pada data mining x dan y adalah atribut basis data numerik. Koefisien w dan b (disebut koefisien regresi), tentukan kemiringan garis dan irisan y, masing-masing. Koefisien ini bisa diselesaikan dengan metode kuadrat terkecil, yang meminimalkan kesalahan antara garis

sebenarnya yang memisahkan data dan perkiraan garis [14].

2.4. K-Nearest Neighbor

Golongan K-Nearest Neighbor mengadaptasikan jumlah smoothing untuk kepadatan data lokal. Tingkat smoothing dikendalikan oleh k, jumlah tetangga yang diperhitungkan, yang jauh lebih kecil daripada N, ukuran sampel. Mari kita tentukan jarak antara a dan b, misalnya |a - b|, dan untuk setiap x, kita definisikan:

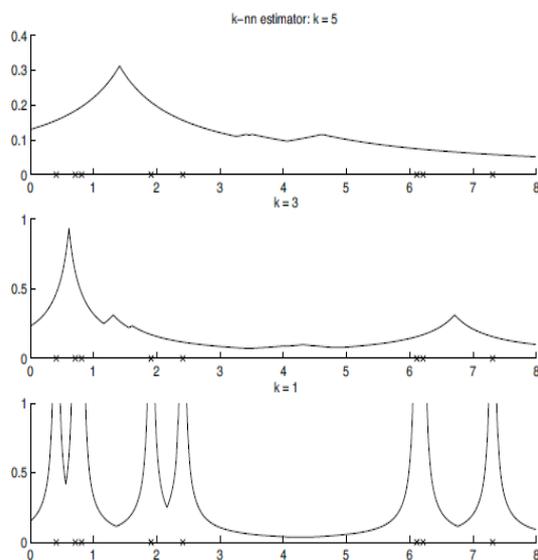
$$d_1(x) \leq d_2(x) \dots d_N(x)$$

menjadi jarak yang diatur dalam urutan menaik, dari x ke titik

dalam sampel $d_1(x)$ adalah jarak ke sampel terdekat, $d_2(x)$ adalah jarak ke terdekat berikutnya, dan seterusnya. Jika x^t adalah titik data, maka kita tentukan $d_1(x) = \min_t |x - x^t|$, dan jika i adalah indeks dari sampel terdekat, yaitu, $i = \operatorname{argmin}_t |x - x^t|$, lalu $d_2(x) = \min_{j \neq i} |x - x^j|$, dan seterusnya.

Estimasi kepadatan K-Nearest Neighbor (K-NN) adalah

$$\hat{p}(x) = \frac{k}{2Nd_k(x)}$$



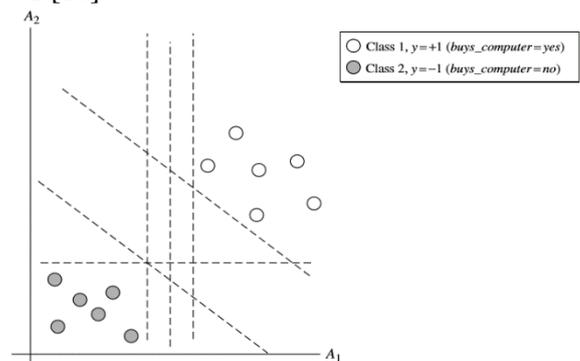
Gambar 3. Perkiraan K-Nearest Neighbor untuk Berbagai Nilai K [15]

Ini seperti penduga yang naif dengan $h = 2d_k(x)$, perbedaannya adalah untuk memperbaiki h dan memeriksa berapa banyak sampel yang jatuh kedalam bin, disini k diperbaiki, jumlah observasi yang jatuh ke dalam bin, dan hitung ukuran bin. Di mana kepadatan tinggi, bin kecil, dan di mana kepadatan rendah, bin lebih besar. Hal ini dapat dilihat pada Gambar 3 [15].

2.5. Support Vector Machine

Metode Support Vector Machine (SVM) berakar dari teori pembelajaran statistik yang memberikan hasil lebih baik dibandingkan metode lain. SVM dapat bekerja dengan baik pada pembelajaran dataset yang memiliki dimensi tinggi [16].

Dasar algoritma SVM adalah memaksimalkan batas hyperplane. Hyperplane merupakan batas keputusan pemisah terbaik antara kedua kelas, karena dapat menemukan pengukuran margin dan mencari titik maksimalnya. Margin merupakan jarak antara hyperplane dengan data terdekat dari masing-masing kelas. Data yang paling dekat ini disebut sebagai support vector [17]. Contoh SVM pada kasus dua kelas yang terpisah secara linear. Kumpulan data D sebagai $(X_1, y_1), (X_2, y_2), \dots, (X_{|D|}, y_{|D|})$, dimana X_i adalah himpunan tupel pelatihan dengan label kelas terkait y_i . Setiap y_i dapat mengambil salah satu dari dua nilai, +1 atau -1 ($y_i \in \{+1, -1\}$). Pada Gambar 1 menunjukkan adanya kelas *buys computer = yes* dan *buys computer = no*. Pada grafik dalam Gambar 4 bahwa data terpisah secara linear dengan garis lurus yang dapat ditarik untuk memisahkan semua tupel kelas +1 dan kelas -1 [14].



Gambar 4. Data training yang terpisah secara linear dengan dua Kelas [14]

2.6. RMSE (Root Mean Square Error)

RMSE (Root Mean Square Error) adalah salah satu statistik indeks error yang umum digunakan dan merupakan akar kuadrat dari MSE (Mean Squared Error) dan menghitung perbandingan (quantifies) kinerja model untuk setiap variable ^[18]. Statistik r^2 mengukur kebaikan kesesuaian regresi dari kumpulan data. Selanjutnya, statistik, yang dikenal sebagai kesalahan standar perkiraan, adalah sebuah ukuran ketepatan estimasi yang dihasilkan oleh regresi. Detailnya, s adalah satu dari statistik yang paling penting untuk dipertimbangkan ketika melakukan analisis regresi.

Untuk menemukan nilai s , pertama-tama kita menemukan Mean Square Error:

$$MSE = \frac{SSE}{n - m - 1}$$

dimana m menunjukkan jumlah variabel prediktor, yaitu 1 untuk linear sederhana kasus regresi dan lebih besar dari 1 untuk kasus regresi berganda. Seperti SSE, MSE mewakili ukuran variabilitas dalam variabel respon yang tidak dapat dijelaskan oleh regresi.

Kemudian kesalahan standar perkiraan seperti ini:

$$s = \sqrt{MSE} = \sqrt{\frac{SSE}{n - m - 1}}$$

Nilai s memberikan perkiraan residual "khas", sama dengan nilai dari standar deviasi dalam analisis univariat memberikan perkiraan deviasi tipikal. Dengan kata lain, s adalah ukuran kesalahan khas dalam estimasi, perbedaan yang khas antara nilai respons yang diprediksi dan nilai respons aktual. Dengan cara ini, kesalahan standar perkiraan s menunjukkan ketepatan prediksi yang dihasilkan oleh persamaan regresi yang diperkirakan ^[19].

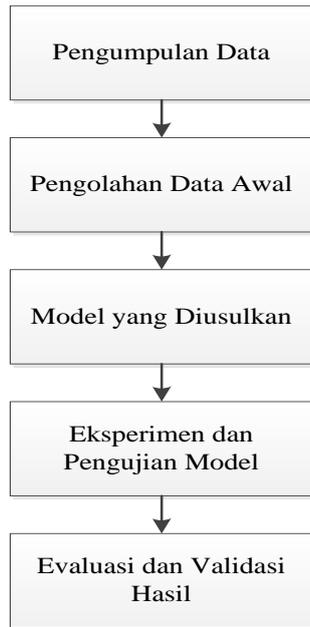
III. METODE PENELITIAN

Metode penelitian yang digunakan adalah pendekatan kuantitatif. Tujuan dari metode kuantitatif adalah untuk mengembangkan model, teori, dan hipotesis yang berkaitan dengan fenomena alam. Aspek kuantitatif

adalah untuk menekankan pengukuran yang mendasar karena memberikan hubungan antara observasi dan formalisasi model, teori dan hipotesis [20]. Penelitian merupakan kegiatan yang bertujuan untuk membuat kontribusi orisinal terhadap ilmu pengetahuan ^[21]. Terdapat empat metode penelitian yang umum digunakan, yaitu: penelitian langsung, eksperimen, studi kasus dan survey. Metode penelitian yang dilakukan pada penelitian ini dapat dilihat pada Gambar 5.

Metode penelitian ini dilakukan beberapa tahapan dalam penelitian sebagai berikut:

1. Pengumpulan data
Pada tahap ini akan dijelaskan bagaimana data tersebut dikumpulkan dan darimana data tersebut diperoleh sehingga data tersebut dapat digunakan dalam eksperimen penelitian.
2. Pengolahan data awal
Pada tahap ini akan dijelaskan bagaimana pengolahan data untuk mendapatkan data yang sudah ditransformasikan agar sesuai dengan bentuk yang diinginkan.
3. Model yang diusulkan
Setelah dilakukan tahap pengolahan data pada tahap sebelumnya, selanjutnya adalah menentukan model yang kemudian akan dilakukan pengujian terhadap data yang sudah ada.
4. Eksperimen dan pengujian model
Selanjutnya dalam tahapan ini akan dilakukan pengujian terhadap model yang diusulkan untuk mendapatkan hasil kinerja dari model yang diusulkan.
5. Evaluasi dan validasi model
Pada tahapan penelitian terakhir ini akan dilakukan evaluasi terhadap eksperimen dan pengujian pada model yang diusulkan sehingga dapat mengetahui hasil kinerja dalam penelitian ini.



Gambar 5. Metode Penelitian

3.1. Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data sekunder, yaitu data yang tidak diperoleh langsung melainkan dikumpulkan oleh pihak lain. Data sekunder yang digunakan adalah dataset publik kebakaran hutan dari bulan Januari tahun 2000 sampai bulan Desember tahun 2003.

Data tersebut dapat di unduh melalui situs

<https://archive.ics.uci.edu/ml/datasets/Forest+Fires>. Data ini berjumlah 517 record dan 12 atribut sebagai variabel prediktor dan 1 atribut sebagai variabel target seperti atribut area. 12 Atribut atau fitur variabel prediktor pada dataset ini terdiri dari sumbu x, sumbu y, bulan, hari, data meteorologi (temp, relative, humidity, wind, rain) dan 4 indeks cuaca kebakaran, yaitu FFMC, DMC, DC, dan ISI. Fine Fuel Moisture Code (FFMC) merupakan indikator untuk bahan yang mudah terbakar dan mudah membentuk pengapian di dalam hutan, Duff Moisture Code (DMC) merupakan indikasi dari konsumsi bahan bakar pada kayu. Drought Code (DC) merupakan indikator dari efek kemarau pada bahan bakar hutan. Sedangkan Initial Spread Index (ISI) merupakan indikator tingkat penyebaran api dan indikator kesulitan pengendalian api.

3.2. Pengolahan Data Awal

3.2.1. Penghapusan Atribut yang tidak Relevan

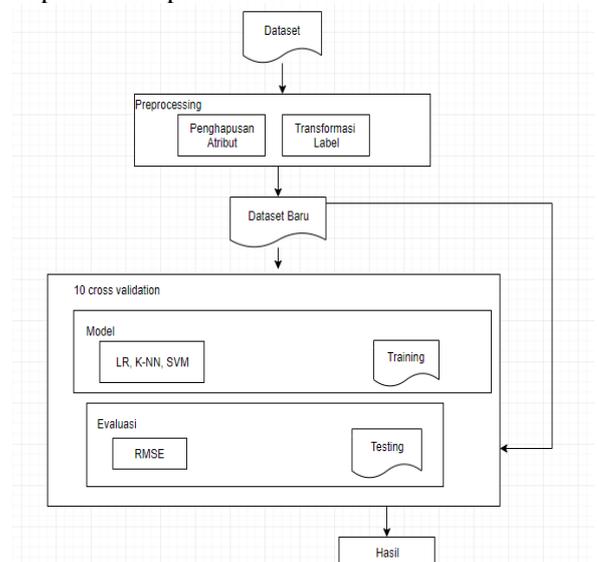
Pada dataset kebakaran hutan terdapat beberapa atribut yang tidak relevan seperti atribut temporal yaitu bulan dan hari (month and day), atribut spatial yaitu sumbu x dan y^[7] sehingga atribut tersebut dihapus.

3.2.2. Transformasi Label

Atribut area yang terbakar pada dataset memiliki kecenderungan positif^[7] oleh karena itu untuk mengurangi kecenderungan dan meningkatkan kesimetrisan maka fungsi logaritma $y = \ln(x + 1)$ diterapkan pada atribut area sebagai transformasi umum yang cenderung meningkatkan hasil regresi untuk target condong kanan. Kemudian area yang telah ditransformasi akan dijadikan target prediksi atau label.

3.3. Model yang diusulkan

Model yang diusulkan pada penelitian ini diawali dengan menghapus atribut yang tidak relevan dan mentransformasi label, kemudian akan terbentuk dataset yang baru. Dataset baru dibagi menjadi beberapa 10 bagian yang terdiri dari 10% data testing dan 90% data training yang diolah menggunakan 10-fold cross validation. Setelah dibentuk menjadi beberapa bagian selanjutnya data diolah menggunakan algoritma LR, K-NN dan SVM Setelah melakukan training dan testing, algoritma tersebut akan menghasilkan RSME atau tingkat error dimana semakin nilainya kecil maka semakin mendekati nilai nol. Langkah-langkah dapat dilihat pada Gambar 6.



Gambar 6. Metodologi Penelitian

IV. HASIL DAN PEMBAHASAN

Pengukuran model dilakukan dengan penghapusan atribut yang tidak relevan, dimana atribut tersebut adalah atribut temporal seperti atribut bulan (month), hari (day) dan atribut spatial yaitu koordinat X dan Y, dimana atribut tersebut dianggap tidak relevan. Pada dataset kebakaran hutan juga memiliki kecenderungan positif pada area yang terbakar [7], dimana kecenderungan tersebut dapat meningkatkan kesimetrisan sehingga digunakan fungsi logaritma $y = \ln(x + 1)$. Fungsi ini di implementasikan pada atribut area sebagai transformasi umum yang dapat meningkatkan hasil regresi untuk target condong kanan. Kemudian area yang sudah di hitung dengan fungsi logaritma menjadi atribut target atau label. Pengolahan data ini menggunakan aplikasi Rapidminer dengan menggunakan 3 algoritma, yaitu LR, K-NN dan SVM. Hasil pengukurannya menggunakan RMSE. Rule algoritma yang muncul pada pengolahan data di Rapidminer di implementasikan pada aplikasi berbasis web yang di rancang dan dibangun oleh peneliti. Pada Tabel 3 merupakan hasil RMSE pada 3 algoritma, dimana hasil terbaik ada pada algoritma LR atau Linear Regression sebesar 1.496. Nilai terbaik kedua terdapat pada algoritma K-NN atau K-Nearest Neighbor sebesar 1.981 dan algoritma SVM atau Support Vector Machine mendapat nilai RMSE terbaik ketiga sebesar 3.088.

Tabel 3 merupakan rule dari algoritma Linear Regression

Algoritma	RMSE
LR	1.496
K-NN	1.981
SVM	3.088

Pada Tabel 4 berisi mengenai rule algoritma Linear Regression yang di dapat pada saat melakukan pengolahan data di Rapidminer, sedangkan Tabel 5 merupakan rule dari algoritma K-Nearest Neighbor dan Tabel 6 adalah rule dari algoritma Support Vector Machine.

Tabel 4 Rule Algoritma Linear Regression

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
FFMC	0.008	0.014	0.030	0.956	0.528	0.598	
DMC	0.001	0.001	0.058	0.913	0.904	0.366	
DC	0.000	0.000	0.051	0.911	0.822	0.411	
ISI	-0.023	0.016	-0.076	0.890	-1.433	0.152	
RH	-0.006	0.004	-0.066	1.000	-1.376	0.169	
wind	0.075	0.036	0.096	0.960	2.093	0.037	**
rain	0.101	0.210	0.021	1.000	0.480	0.631	
(intercept)	0.280	1.297	?	?	0.216	0.829	

Tabel 5 Rule Algoritma K-Nearest Neighbor

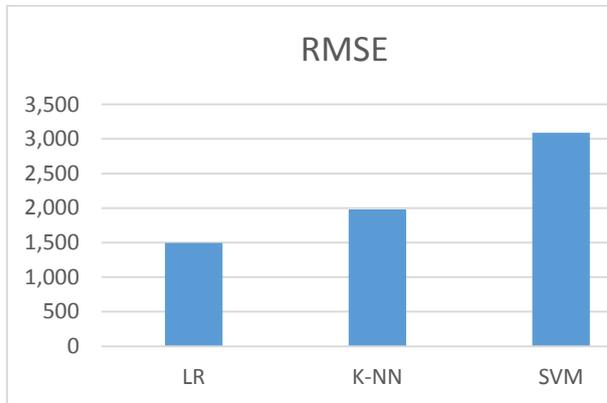
1-Nearest Neighbour model for regression. The model contains 517 examples with 8 dimensions.

Tabel 6 Rule Algoritma Support Vector Machine

Total number of Support Vectors: 517
Bias (offset): 0.505
w[FFMC] = 1.251
w[DMC] = 0.642
w[DC] = 0.767
w[ISI] = -1.365
w[temp] = 0.175
w[RH] = -1.669
w[wind] = 0.487
w[rain] = 1.470

Pada Gambar 7 dapat dilihat grafik perbandingan RMSE dari ketiga algoritma yang digunakan, yaitu algoritma Linear Regression (LR), K-Nearest Neighbor (K-NN) dan Support Vector Machine (SVM). Nilai tertinggi pada

perbandingan ketiga algoritma yang digunakan ada pada algoritma SVM, akan tetapi untuk pengukuran nilai RMSE, nilai terkecil merupakan nilai terbaik dalam penelitian estimasi. Sehingga algoritma terbaik pada penelitian estimasi kebakaran hutan adalah algoritma Linear Regression (LR).



Gambar 7. Grafik Perbandingan Nilai RMSE

V. PENUTUP

Estimasi tingkat radiasi api pada kebakaran hutan dapat membantu untuk mengukur seberapa tinggi atau besar area yang terbakar, sehingga diharapkan penanganan kebakaran hutan dapat dengan cepat dan tepat. Pada penelitian ini dilakukan estimasi dengan menggunakan tiga algoritma yaitu Linear Regression (LR), K-Nearest Neighbor (K-NN) dan Support Vector Machine (SVM). Dataset yang digunakan dalam penelitian ini adalah dataset publik kebakaran hutan yang diperoleh dari UCI Repository. Hasil pengolahan terhadap dataset dengan menggunakan ketiga algoritma, menunjukkan bahwa nilai RMSE terbaik terdapat pada algoritma Linear Regression (LR) sebesar 1.496, nilai RMSE terbaik kedua ada pada algoritma K-Nearest Neighbor (K-NN) sebesar 1.981 dan terbaik ketiga untuk nilai RMSE adalah algoritma Support Vector Machine (SVM) dengan nilai 3.088. Untuk penelitian selanjutnya dapat menggunakan beberapa metode lain seperti Neural Network dan metode ensemble seperti Bagging untuk meningkatkan nilai RMSE.

DAFTAR PUSTAKA

- [1] L. Tacconi, "Kebakaran Hutan di Indonesia: Penyebab, Biaya dan Implikasi Kebijakan," *Center for International*
- [2] R. K. Dewi, "Kebakaran Hutan dan Lahan, Apa Dampak dan Upaya Pencegahannya," Kompas, 25 Agustus 2018. [Online]. Available: <https://nasional.kompas.com/read/2018/08/25/14340331/kebakaran-hutan-dan-lahan-apa-dampak-dan-upaya-pencegahannya>. [Accessed 21 Oktober 2018].
- [3] A. Alonso-Betanzos, O. Fontela-Romero, B. Guijarro-Berdinas, e. Hernandez-Pereira, M. I. P. Andrade, E. Jimenez, J. L. L. Soto and T. Carballas, "An intelligent system for forest fire risk prediction and fire fighting management in Galicia," *Expert Systems with Applications*, pp. 545-554, 2003.
- [4] C. Quintano, A. Fernandez-Manso, A. Stein and W. Bijker, "Estimation of area burned by forest fires in Mediterranean countries: A remote sensing data mining perspective," *Forest Ecology and Management*, pp. 1597-1607, 2011.
- [5] E. Chuvieco, D. Cocero, D. Riano, P. Martin, J. Martinez-Vega, J. d. l. Riva and F. Perez, "Combining NDVI and Surface Temperature for The Estimation of Live Fuel Moisture Content in Forest Fire Danger Rating," *Remote Sensing of Environment*, pp. 322-331, 2004.
- [6] K. Buza, A. Nanopoulos and G. Nagy, "Nearest Neighbor Regression in the Presence of Bad Hubs," *Knowledge-Based Systems*, vol. 86, pp. 250-260, 2015.
- [7] P. Cortez and A. Morais, "A Data Mining Approach to Predict Forest Fires using Meteorological Data," 2007.
- [8] R. S. Wahono, "A Systematic Literature Review of Software Defect Prediction: Research Trends, Datasets, Methods and Frameworks," *Journal of Software Engineering*, pp. 1-16, 2015.
- [9] O. Maimon and L. Rokach, *Data Mining and Knowledge Discovery Handbook*, Neew York: Springer, 2010.
- [10] X. Wang, J. Wen, Y. Zhang and Y. Wang, "Real estate price forecasting based on SVM optimized by PSO," *Optik*, vol. 215, no. 3, pp. 1439-1443, 2014.
- [11] M. Yetti, "Kebakaran Hutan," Kompas, 3 September 2017. [Online]. Available:

- <https://www.kompasiana.com/yetti/59ab65cfcbb34c06d3769d02/kebakaran-hutan>.
[Accessed 21 Oktober 2018].
- [12] L. Vonny, "Dampak Kebakaran Hutan terhadap Lingkungan," Kompasiana, 23 November 2017. [Online]. Available: <https://www.kompasiana.com/lidiaponii/5a166ec305f1cd3b6a346482/dampak-kebakaran-hutan-terhadap-lingkungan>. [Accessed 21 Oktober 2018].
- [13] N. Romdlon, "Lumpuhnya Kota Pekanbaru dan kota besar lain di Sumatera pasti berdampak pada aktivitas ekonomi masyarakat.," Brilio, 21 September 2015. [Online]. Available: <https://www.brilio.net/news/5-dampak-kebakaran-hutan-dan-kabut-asap-yang-berkepanjangan-150921q.html>. [Accessed 21 Oktober 2018].
- [14] J. Han, M. Kamber and J. Pei, Data Mining: Concepts and Techniques, Amsterdam: Elsevier, 2012.
- [15] E. Alpaydin, Introduction to Machine Learning, London: The MIT Press, 2010.
- [16] E. Prasetyo, DATA MINING - Konsep dan Aplikasi Menggunakan MATLAB, Yogyakarta: Penerbit ANDI, 2012.
- [17] E. Prasetyo, Data Mining-Mengolah Data Menjadi Informasi Menggunakan Matlab, Yogyakarta: Penerbit ANDI, 2014.
- [18] F. Salerno, G. Viviano, E. Carraro, E. C. Manfredi, A. Lami, S. Musazzi, A. Marchetto, N. Guyennon, G. Tartari and D. Copetti, "Total phosphorus reference condition for subalpine lakes: A comparison among traditional methods and a new process-based watershed approach," *Journal of Environmental Management*, vol. 145, pp. 94-105, 2014.
- [19] D. T. Larose, Data Mining Methods and Models, Canada: Wiley-Interscience, 2006.
- [20] M. Berndtsson, J. Hansson, B. Olsson and B. Lundell, Thesis Projects, London: Springer, 2008.
- [21] C. W. Dawson, Projects in Computing and Information Systems, London: Addison Wesley, 2009.