

PENERAPAN DATA MINING UNTUK CLUSTERING KUALITAS BATU BARA DALAM PROSES PEMBAKARAN DI PLTU SEBALANG MENGGUNAKAN METODE K-MEANS

Muhammad Bakri

*Sistem Informasi, STMIK Teknokrat
Jl. H.ZA Pagaralam, No 9-11, Labuhanratu, Bandar Lampung
Email : muhbakri1985@gmail.com*

Abstrak

Penelitian ini bertujuan untuk menjelaskan bagaimana membuat cluster kualitas batubara dari data rekapitulasi kualitas dan kuantitas batubara pada PT PLN(Persero) Sektor Pembangkitan Sebalang. Kondisi yang ada saat ini adalah belum adanya analisis data kualitas batubara terhadap kinerja perusahaan yang digunakan untuk membantu pegawai dalam mengevaluasi kinerja PLTU dan pasokan batubara, memberikan informasi pengaruh kualitas batubara terhadap produksi listrik yang dihasilkan. Salah satu solusi yang dapat diterapkan pada permasalahan ini dengan data mining untuk clustering kualitas batubara menggunakan algoritma K-Means.

Metode Pengumpulan data yang digunakan yaitu metode dokumentasi, dan observasi untuk memperoleh data yang di perlukan melalui data yang telah tersedia maka akan memberikan kemudahan kepada peneliti untuk melakukan penelitian. Tools yang digunakan dalam penerapan data mining adalah Weka.

Hasil penelitian berdasarkan hasil analisis secara keseluruhan akan menghasilkan output cluster kualitas batubara yang buruk, normal, dan baik di PT PLN (Persero) Sektor Pembangkitan Sebalang..

Kata kunci: *PLN, Data Mining, Cluster, Algoritma K-Means, Kualitas batubara.*

1. Pendahuluan

PT PLN(Persero) Sektor Pembangkitan Sebalang adalah salah satu kegiatan usaha yang dimiliki PT PLN(Persero) Pembangkitan Sumatera Bagian Selatan, mempunyai 2 unit usaha Pembangkit Listrik Tenaga Uap (PLTU) berkapasitas 2x100 MegaWatt (MW). PT PLN(Persero) Sektor Pembangkitan Sebalang ini berlokasi di Desa Sebalang, Kecamatan Katibung, Kabupaten Lampung Selatan. Dalam proses pengoperasian PLTU Sebalang menggunakan dua jenis bahan bakar, yang pertama adalah High Speed Diesel (HSD) sebagai bahan bakar untuk initial firing dan batubara sebagai bahan bakar utama. Batubara merupakan bahan bakar yang paling berperan dalam proses pembakaran sehingga diperlukan suatu analisis terhadap kualitas batubara yang selama ini diterima untuk mengoptimalkan kinerja pembangkit.

Kualitas batubara yang diterima harus mempunyai standar yang sudah ditentukan dan dianalisa langsung oleh pihak ketiga antara pembeli dan pemasok yaitu surveyor independent. Permasalahan yang terjadi pada PT PLN (Persero) Sektor Pembangkitan Sebalang dalam lingkup kinerja pembangkit yaitu belum adanya evaluasi pengaruh kualitas batubara terhadap kinerja PLTU. Selain itu belum adanya standar baku untuk menentukan kualitas batubara yang diperlukan oleh PLTU.

Tujuan dari penelitian ini untuk menerapkan algoritma K-means dalam menganalisis kualitas batubara yang dipakai di PT PLN (Persero) Sektor Pembangkitan Sebalang. Analisis data yang dilakukan nantinya dapat digunakan membantu menentukan parameter kualitas batubara yang paling tepat untuk PLTU Sebalang sehingga dapat dijadikan acuan dalam pasokan batubara agar kinerja pembangkit menjadi lebih optimal.

2. Landasan Teori

A. Batubara

Batubara adalah salah satu bahan bakar fosil. Pengertian umumnya adalah batuan sedien yang dapat terbakar, terbentuk dari endapan organik, utamanya adalah sisa-sisa tumbuhan dan terbentuk melalui proses pembatubaraan [1].

B. Unsur – unsur dalam batubara

Beberapa unsur – unsur dalam batubara adalah sebagai berikut:

- Moisture : Menurunkan Nilai Kalori batubara
- Volatile : Lidah api burner kurang stabil
- Sulfur : Terjadinya korosi pada laluan gas panas
- Abu : Berdampak terhadap Bottom ash dan fly ash handling system
- AFT : Terbentuknya slagging dan fouling
- Kalori : Menghasilkan energi panas
- HGI : Beban kinerja crusher

C. American Society for Testing and Material (ASTM)

ASTM Internasional merupakan organisasi internasional sukarela yang mengembangkan standardisasi teknik untuk

material, produk, sistem dan jasa. ASTM Internasional yang berpusat di Amerika Serikat.

ASTM merupakan singkatan dari American Society for Testing and Material, dibentuk pertama kali pada tahun 1898 oleh sekelompok insinyur dan ilmuwan untuk mengatasi bahan baku besi pada rel kereta api yang selalu bermasalah. Sekarang ini, ASTM mempunyai lebih dari 12.000 buah standar. Standar ASTM banyak digunakan pada negara-negara maju maupun berkembang dalam penelitian akademisi maupun industri Standar yang digunakan dalam proses pengujian kualitas batubara antara lain :

- ASTM 6883 tentang sampling manual dari truck dan stock pile
- ASTM D2013 tentang preparasi sampel batubara
- ASTM D3176 tentang ultimate analysis

D. Data Mining

Data Mining merupakan proses pengekstraksian informasi dari sekumpulan data yang sangat besar melalui penggunaan algoritma dan teknik penarikan dalam bidang statistik, pembelajaran mesin dan sistem manajemen basis data. Data mining adalah proses menganalisa data dari perspektif yang berbeda dan menyimpulkannya menjadi informasi-informasi penting yang dapat dipakai untuk meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya.

E. Teknik Clustering

Clustering adalah proses pengelompokan data ke dalam kelas atau kelompok , sehingga objek dalam sebuah cluster memiliki kesamaan yang tinggi dibandingkan dengan satu sama lain tetapi sangat berbeda dengan objek dalam cluster lainnya. Ketidakmiripan dinilai berdasarkan nilai atribut yang menggambarkan objek. Seringkali, langkah-langkah jarak yang digunakan. Clustering berakar di banyak daerah, termasuk data mining, statistik, biologi dan mesin pembelajaran.

F. Algoritma K-Means

Algoritma K-Means mengambil parameter masukan, k, lalu membagi dan membuat n objek ke dalam cluster k sehingga kesamaan yang dihasilkan intra cluster tinggi tetapi kesamaan intercluster rendah. Kesamaan cluster diukur dalam kaitannya dengan nilai rata-rata dari objek dalam sebuah cluster, yang dapat dilihat sebagai centroid cluster atau pusat gravitasi².

Secara umum algoritma dasar dari K-Means Clustering adalah sebagai berikut

1. Tentukan jumlah cluster
2. Alokasikan data ke dalam cluster secara random
3. Hitung centroid/rata-rata dari data yang ada di masing-masing cluster
4. Alokasikan masing-masing data ke centroid/rata- rata terdekat
5. Kembali ke Step 3, apabila masih ada data yang

berpindah cluster atau apabila perubahan nilai centroid, ada yang di atas nilai treshold yang ditentukan atau apabila perubahan nilai pada objective function yang digunakan di atas nilai threshold yang ditentukan

Distance space digunakan untuk menghitung jarak antara data dan centroid. Adapun persamaan yang dapat digunakan salah satunya yaitu Euclidean Distance Space. Euclidean distance space sering digunakan dalam perhitungan jarak, hal ini dikarenakan hasil yang diperoleh merupakan jarak terpendek antara dua titik yang diperhitungkan. Adapun persamaannya adalah sebagai berikut:

$$d_{ij} = \sqrt{\sum_{k=1}^p \{x_{ik} - x_{jk}\}^2}$$

Dimana

d_{ij} = Jarak objek antara objek i dan j
 P = Dimensi data

X_{ik} = Koordinat dari obyek i pada dimensi k

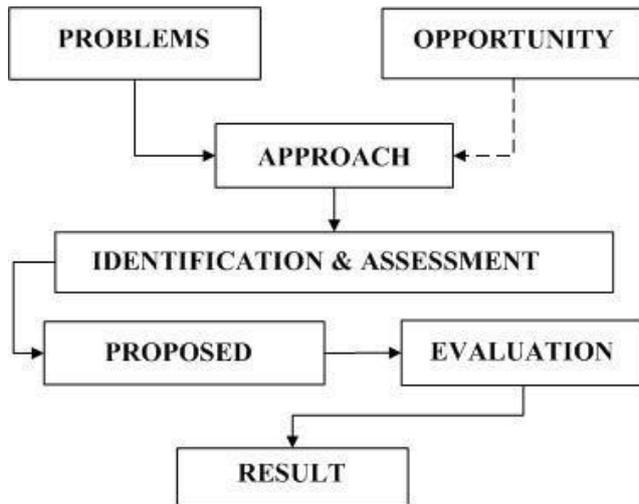
X_{jk} = Koordinat dari obyek j pada dimensi k

G. WEKA

Weka adalah aplikasi *data mining open source* berbasis Java. Aplikasi ini dikembangkan pertama kali oleh Universitas Waikato di Selandia Baru sebelum menjadi bagian dari Pentaho. Weka terdiri dari koleksi algoritma machine learning yang dapat digunakan untuk melakukan generalisasi / formulasi dari sekumpulan data sampling. Walaupun kekuatan Weka terletak pada algoritma yang makin lengkap dan canggih, kesuksesan data mining tetap terletak pada faktor pengetahuan manusia implementornya. Tugas pengumpulan data yang berkualitas tinggi dan pengetahuan pemodelan dan penggunaan algoritma yang tepat diperlukan untuk menjamin keakuratan formulasi yang diharapkan.

3. Kerangka Pemikiran

Kerangka pemikiran merupakan alur utama dari penelitian dengan urutan problem (masalah), opportunity (peluang), approach (pendekatan), identification and assessment (identifikasi dan pemetaan), proposed (pemodelan), evaluasi dan result (hasil). Pada dasarnya penelitian adalah kerangka hubungan antara konsep – konsep yang ingin diamati atau diukur melalui penelitian yang akan dilakukan. Penelitian didasarkan pada kualitas batubara yang diterima yang tidak konsisten yang disebabkan oleh standar kualitas batubara berada dalam kisaran yang cukup jauh. Maka berdasarkan kerangka penelitian yang digunakan secara keseluruhan dapat dilihat pada Gambar 1:

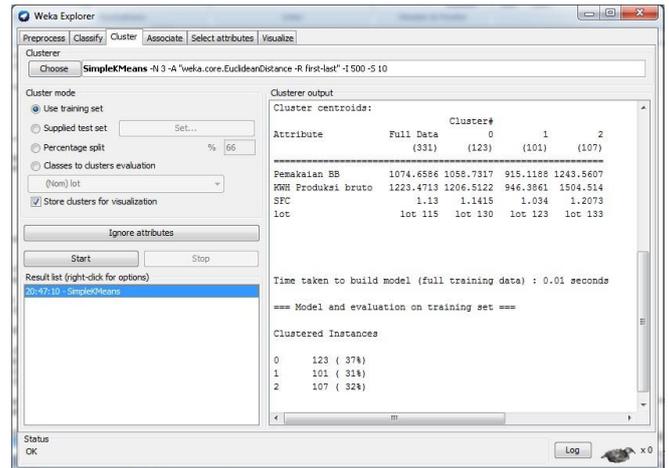


Gambar 1. Kerangka pemikiran

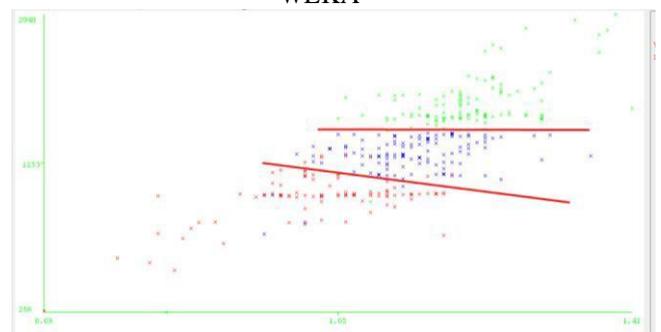
Di PT PLN (Persero) Sektor Tarahan belum evaluasi pengaruh kualitas batubara terhadap kinerja PLTU. Selain itu belum adanya standar baku untuk menentukan kualitas batubara yang diperlukan oleh PLTU agar dapat kinerja yang maksimal. Standar batubara yang diterima masih dalam jarak / kisaran yang cukup besar sehingga pemasok sering mengirimkan batubara dengan perbedaan kualitas yang cukup jauh. Hal ini dapat menyebabkan produksi listrik menjadi tidak stabil dan juga kinerja PLTU menjadi tidak maksimal karena tidak diketahui kualitas batubara yang paling ideal untuk PLTU Sebalang. Penanganan terhadap masalah tersebut dapat dilakukan dengan penerapan cluster menggunakan algoritma K- Means. Pendekatan ini dilakukan dengan menentukan variabel – variabel yang paling mempengaruhi kinerja PLTU. Tahapan validasi dilakukan sebagai pengujian terhadap pendekatan yang dilakukan yaitu meliputi pengujian manual dan pengujian data olahan. Setelah divalidasi maka akan menghasilkan sebuah cluster terhadap kualitas batubara pada PLTU Sebalang.

4. Penerapan Algoritma

Penerapan k-means dengan menggunakan weka digunakan untuk membuat cluster secara keseluruhan dan menampilkan visualisasi cluster yang terbentuk dalam mengklaster data pembangkit. Dimana data yang digunakan sebanyak 331 record. Berikut gambar hasil implementasi kedua menggunakan weka:



Gambar 2. Implementasi metode k-means dengan WEKA



Gambar 3. Visualisasi hasil cluster

A. Evaluasi

Pengujian metode algoritma k-means dilakukan untuk mengetahui internal homogeneity dan eksternal heterogeneity menggunakan data hasil visualisasi implementasi clustering:

Tabel 1. Data pengujian berdasarkan data terluar dari masing masing cluster

Centroid		Data cluster		C
Produksi KWH bruto	Pemakaian BB	Produksi KWH bruto	Pemakaian BB	
946.3861	915.1188	1043	942	C1
1206.5122	1058.7317	1091	997	C0
1206.5122	1058.7317	1321	1198	C0
1504.514	1243.5607	1346	1215	C2

Jarak data 1 dengan centroid 1-3 adalah:

$$d(X_1, C_1) = \sqrt{(a_j - c_j)^2 + (b_j - c_j)^2}$$

$$d(X_1, C_1) = \sqrt{(1043 - 946.3861)^2 + (942 - 915.1188)^2}$$

$$= 100.28382$$

$$d(X_1, C_0) = \sqrt{(1043 - 1206.5122)^2 + (942 - 1058.7317)^2}$$

$$= 200.90427$$

$$d(X_1, C_2) = \sqrt{(1043 - 1504.514)^2 + (942 - 1243.5607)^2}$$

$$= 551.30212$$

Jarak data 2 dengan centroid 1-3 adalah:

$$d(X_i, C_j) = \sqrt{(a_j - c_j)^2 + (b_j - c_j)^2}$$

$$d(X_2, C_1) = \sqrt{(1091 - 946.3861)^2 + (997 - 915.1188)^2} = 285.84149$$

$$d(X_2, C_0) = \sqrt{(1091 - 1206.5122)^2 + (997 - 1058.7317)^2} = 130.972787$$

$$d(X_2, C_2) = \sqrt{(1091 - 1504.514)^2 + (997 - 1243.5607)^2} = 481.441592$$

Jarak data 3 dengan centroid 1-3 adalah:

$$d(X_i, C_j) = \sqrt{(a_j - c_j)^2 + (b_j - c_j)^2}$$

$$d(X_3, C_1) = \sqrt{(1321 - 946.3861)^2 + (1198 - 915.1188)^2} = 469.422355$$

$$d(X_3, C_0) = \sqrt{(1321 - 1206.5122)^2 + (1198 - 1058.7317)^2} = 180.286205$$

$$d(X_3, C_2) = \sqrt{(1321 - 1504.514)^2 + (1198 - 1243.5607)^2} = 189.085074$$

Jarak data 4 dengan centroid 1-3 adalah:

$$d(X_i, C_j) = \sqrt{(a_j - c_j)^2 + (b_j - c_j)^2}$$

$$d(X_4, C_1) = \sqrt{(1346 - 946.3861)^2 + (1215 - 915.1188)^2} = 499.619858$$

$$d(X_4, C_0) = \sqrt{(1346 - 1206.5122)^2 + (1215 - 1058.7317)^2} = 209.46748$$

$$d(X_4, C_2) = \sqrt{(1346 - 1504.514)^2 + (1215 - 1243.5607)^2} = 161.066451$$

Jarak eksternal heterogeneity antara cluster 0, cluster 1, dan cluster 2 sebagai berikut:

$$d(C_0, C_1) = \sqrt{(1091 - 1043)^2 + (997 - 942)^2} = 73$$

$$d(C_1, C_2) = \sqrt{(1043 - 1346)^2 + (942 - 1215)^2} = 407.8455$$

$$d(C_0, C_2) = \sqrt{(1321 - 1346)^2 + (1198 - 1215)^2} = 30,2324$$

B. Analisis dan Penerapan WEKA

Pada penerapan k-means menggunakan weka, didapatkan 3 centroid dari 3 sluster yang dibentuk yaitu:

1. Cluster 0
 - Pemakaian batu bara : 1058.7317
 - Kwh produksi bruto : 1206.5122
 - SFC : 1.1415
 - Lot : lot 130
2. Cluster 1
 - Pemakaian batu bara : 915.1188
 - Kwh produksi bruto : 946.3861
 - SFC : 1.034
 - Lot : lot 123
3. Cluster 2
 - Pemakaian batu bara : 1243.5607
 - Kwh produksi bruto : 1504.514
 - SFC : 1.2073
 - Lot : lot 133

Pada perhitungan evaluasi implementasi clustering, posisi data yang terletak paling luar tidak mengalami perubahan cluster yang artinya proses cluster di weka berjalan dengan baik. Cluster 1 memiliki internal homogeneity yang baik karena jarak data terluar ke centroidnya lebih pendek dibandingkan cluster 0 dan cluster 2.

Pada perhitungan jarak eksternal heterogeneity, Cluster 1 dengan cluster 2 memiliki eksternal heterogeneity yang baik karena jarak data terluarnya paling jauh

dibandingkan jarak antara cluster 0 dengan cluster 1, dan jarak antara cluster 0 dengan cluster 2.

5. Kesimpulan

Dari penelitian ini dapat disimpulkan beberapa hal berikut :

1. Penerapan data mining dengan metode k-means clustering telah dilakukan dan ditentukan 3 cluster sebagai komoditas kualitas batubara.
2. Variable cluster yang digunakan hanya 4 variabel, dimana 13 variabel lainnya sudah diwakilkan oleh variable lot batu bara. Variabel yang digunakan hanya variabel yang paling berpengaruh terhadap hasil cluster.
3. Dari pengujian yang telah dilakukan, maka dapat disimpulkan bahwa cluster 1 memiliki internal homogeneity yang paling baik karena jarak antara data terluar dengan centroid cluster paling dekat dan eksternal heterogeneity antara cluster 1 dengan 2 paling baik karena jarak antar data terluar cluster paling jauh.

Daftar Pustaka

- [1] Saleh, A. Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. 2, 207–217 (2015).
- [2] Han, J. & Kamber, M. Data Mining: Concepts and Techniques. Ann. Phys. (N. Y). 54, 770 (2006).
- [3] Cole, R. Running Head : PROACTIVE RESEARCH APPROACHES Being Proactive : Where Action Research meets Design Research. 1–21 (2005).