

Clustering Data Polutan Udara Kota Pekanbaru dengan Menggunakan Metode *K-Means Clustering*

Fathuddin Yazid¹, Muhammad Affandes²

^{1,2}Teknik Informatika UIN Sultan Syarif Kasim Riau

Jl. H.R. Soebrantas no. 155 KM. 18 Simpang Baru, Pekanbaru 28293

fathuddinyazid@gmail.com¹, affandes@uin-suska.ac.id²

Abstrak – Pencemaran udara adalah salah satu permasalahan yang selalu mengganggu kesehatan lingkungan, seperti kabut asap yang belakangan ini terjadi setiap tahunnya di wilayah Kota Pekanbaru yang mengganggu kesehatan manusia bahkan sampai menelan korban jiwa. Untuk itu diterapkan ilmu dari *data mining* dengan metode *K-means clustering* yang dapat mengelompokkan polutan pencemar udara. Pengelompokan *cluster* dilakukan dengan beberapa kali pengujian dengan jumlah *cluster* berbeda dan juga dengan pengolahan data yang berbeda sebagai perbandingan untuk mencari hasil yang optimal. Setelah kelompok polutan dari 5 *cluster* di urutkan berdasarkan kadar polutan yang terkandung dapat disimpulkan bahwa polutan mengalami kenaikan antara bulan Juni dan Juli kemudian turun kembali pada bulan Oktober dan November sehingga diharapkan masyarakat lebih waspada pada rentang bulan-bulan tersebut untuk mencegah efek negatif dari polutan udara seperti ispa dan gangguan pernapasan lainnya bahkan dapat menyebabkan kematian.

Kata kunci – *Clustering, Data mining, K-means, Polutan udara*

PENDAHULUAN

Udara yang bersih merupakan kebutuhan setiap manusia, namun tidak hanya manusia yang membutuhkannya, tumbuhan dan hewan juga membutuhkan udara yang bersih untuk dapat bertahan hidup. Udara yang bersih biasanya terdapat di daerah pedesaan dimana masih banyak dijumpai pepohonan dan selain itu polusi dari asap kendaraan maupun asap dari pabrik masih sangat sedikit, namun berbeda halnya dengan daerah perkotaan dimana seperti yang kita ketahui bersama bahwa udara di daerah perkotaan cenderung lebih kotor dibandingkan dengan udara di wilayah pedesaan, banyaknya jumlah kendaraan di daerah perkotaan serta aktifitas pabrik yang berkontribusi pada pencemaran udara oleh polusi. Namun pencemaran udara tidak selalu berasal dari polusi yang

disebabkan oleh mesin kendaraan maupun aktifitas pabrik, pencemaran udara juga dapat disebabkan oleh bencana kemarau panjang dan akan mengakibatkan kebakaran hutan yang asap nya akan membuat udara tercemar seperti yang belakangan ini terjadi tiap tahunnya di wilayah Sumatera khususnya Kota Pekanbaru. Menurut salah seorang ilmuwan Columbia University yang bekerja untuk NASA mengenai bencana di wilayah Sumatera dan sekitar tersebut, kondisi di Singapura dan tetangga Sumatera serupa dengan 1997, Jika perkiraan cuaca yang memprediksi kemarau panjang bertahan, ini akan membuat kabut asap 2015 termasuk yang paling parah dalam sejarah [1].

Kesehatan lingkungan adalah suatu keseimbangan ekologi yang harus ada antara manusia dan lingkungan agar dapat menjamin keadaan sehat dari manusia. Apabila kesehatan lingkungan terganggu maka juga akan berpengaruh kepada kesehatan manusia. Seperti yang terjadi di Kota Pekanbaru dimana terjadi kabut asap yang tidak hanya mengganggu kesehatan lingkungan tetapi juga mengganggu kesehatan manusianya, bahkan sampai menelan korban jiwa. Dikutip dari Liputan6.com, Kabut asap di Riau akibat kebakaran hutan dan lahan kembali menelan korban jiwa, Ramadhani Lutfi Aerli (9) menghembuskan napas terakhir pada Rabu (21/10/2015) dini hari akibat gangguan pernapasan. Sebelumnya murid kelas 3 SD di Jalan Sumatera, Pekanbaru, itu sempat dirawat intensif di Rumah Sakit Santa Maria. Sejumlah perawatan medis tak mampu menyelamatkan nyawanya karena paru-parunya sudah dipenuhi asap. Mereka menambahkan meninggalnya Ramadhani kian memperpanjang daftar korban tewas akibat kabut asap. Sebelumnya ada bocah 12 tahun, Muhanum Angriawati dan seorang PNS meninggal karena kabut asap [2]. Selain mengganggu kesehatan, kabut asap di Kota Pekanbaru juga mengganggu proses belajar mengajar, dari mulai tingkat TK sampai tingkat Perguruan Tinggi. Dikutip dari Majalah Lingkungan Hidup Lestari, Menteri Pendidikan, seperti yang dilansir beberapa media massa telah memberikan edaran untuk mengatasi dampak kabut asap bagi dunia pendidikan. Kesehatan serta keselamatan peserta didik dan pendidik adalah prioritas utama. Dan akan diberlakukan perlakuan khusus bagi daerah yang

terkena dampak kabut asap. Surat edaran Menteri Pendidikan dan Kebudayaan tersebut bertanggal 23 Oktober 2015 dan berisi sembilan poin yang memuat penanganan pada proses belajar mengajar selama dan sesudah kabut asap, salah satunya kebijaksanaan meliburkan siswa dari kegiatan belajar mengajar.

Penelitian terdahulu yang membahas tentang polutan udara yaitu dengan judul *Clustering* Data Pencemaran Udara Sektor Industri Di Jawa Tengah Dengan Kohonen Neural Network. Hasil *clustering* pada data beban pencemaran udara sektor industri di Jawa Tengah tahun 2006 yang termasuk *Cluster I* yaitu industri kecil, yang termasuk *Cluster II* yaitu industri makanan, industri minuman, industri kimia dasar, industri non logam, industri semen, industri kapur dan gips, industri logam dasar, industri hasil-hasil olahan logam, industri rumah sakit, dan industri perhotelan, yang termasuk *Cluster III* yaitu industri kayu, industri olahan kayu, dan industri kertas [3].

Penelitian berjudul *A Cluster Analysis Of Constant Ambient Air Monitoring Data from the Kanto Region of Japan* meng*cluster* stasiun monitoring data wilayah Kanto Jepang berdasarkan 4 polutan yaitu NOx, Ox, NMHC, dan SPM dengan tujuan agar dapat mengurangi jumlah stasiun monitoring yang berada di sekitar wilayah Kanto Jepang. Kesimpulan dari penelitian tersebut yaitu stasiun monitoring di *cluster* kedalam 8 *cluster*, dan dapat dikonfirmasi bahwa stasiun monitoring dapat di *cluster* secara topologi untuk NOx dan Ox menggunakan analisa *cluster*. Jika stasiun

monitoring udara ambien memungkinkan untuk di kelompokkan maka begitu juga dengan mengurangi jumlah stasiun monitoring dapat dimungkinkan dengan melihat beberapa kriteria [4].

Dengan adanya data-data tersebut masyarakat sepatutnya mewaspadai akibat dari masalah polusi udara ini dan tidak meremehkannya, masih banyak di antara masyarakat yang tetap memaksakan diri untuk keluar rumah disaat bencana kabut asap seperti yang telah terjadi, padahal mereka mengetahui bahwa indikator polusi udara yang terpampang di jalan dan juga disebar melalui surat kabar cetak maupun elektronik menyatakan bahwa udara sangatlah tidak sehat dan berbahaya bagi kesehatan. Untuk mewaspadai akibat dari masalah polusi, terlebih dahulu kita harus mengetahui kapan saja kondisi polusi udara berbahaya bagi kesehatan manusia, salah satu caranya yaitu dengan memprediksi dengan melihat pola dari polusi udara yang telah terjadi beberapa tahun terakhir sehingga masyarakat dapat lebih waspada dan dapat mencegah akibat negatif dari polusi udara.

LANDASAN TEORI

Gas polutan adalah gas-gas yang menyebabkan terjadinya polusi atau pencemaran udara. Gas-gas polutan ini dapat menimbulkan berbagai bibit penyakit bagi semua makhluk hidup baik manusia, hewan dan juga tumbuhan.

| Kategori | Rentang | Carbon Monoksida (CO) | Nitrogen (NO ₂) | Ozon (O ₃) | Sulfur Dioksida (SO ₂) | Partikulat |
|--------------------|-------------|--|--|---|---|--|
| Baik | 0 - 50 | Tidak ada efek | Sedikit berbau | Luka pada beberapa spesies tumbuhan akibat kombinasi dengan SO ₂ (selama 4 jam) | Luka pada beberapa spesies tumbuhan akibat kombinasi dengan O ₃ (selama 4 jam) | Tidak ada efek |
| Sedang | 51 - 100 | Perubahan kimia darah tapi tidak terdeteksi | Berbau | Luka pada beberapa spesies tumbuhan | Luka pada beberapa spesies tumbuhan | Terjadi penurunan pada jarak pandang |
| Tidak Sehat | 101 - 199 | Peningkatan pada kardiovaskular pada perokok yang sakit jantung | Bau dan kehilangan warna. Peningkatan reaktivitas pembuluh tenggorokan pada penderita asma | Penurunan kemampuan pada atlet yang berlatih keras | Bau, meningkatnya kerusakan tanaman | Jarak pandang turun dan terjadi pengotoran debu dimana-mana |
| Sangat Tidak Sehat | 200 - 299 | Meningkatnya kardiovaskular pada orang bukan perokok yang berpenyakit jantung, dan akan tampak beberapa kelemahan yang terlihat secara nyata | Meningkatnya sensitivitas pasien yang berpenyakit asma dan bronhitis | Olah raga ringan mengakibatkan pengaruh pemafasan pada pasien yang berpenyakit paru-paru kronis | Meningkatnya sensitivitas pada pasien berpenyakit asma dan bronhitis | Meningkatnya sensitivitas pada pasien berpenyakit asma dan bronhitis |
| Berbahaya | 300 - lebih | Tingkat yang berbahaya bagi semua populasi yang terpapar | | | | |

Gambar 1. Pengaruh untuk setiap parameter pencemar

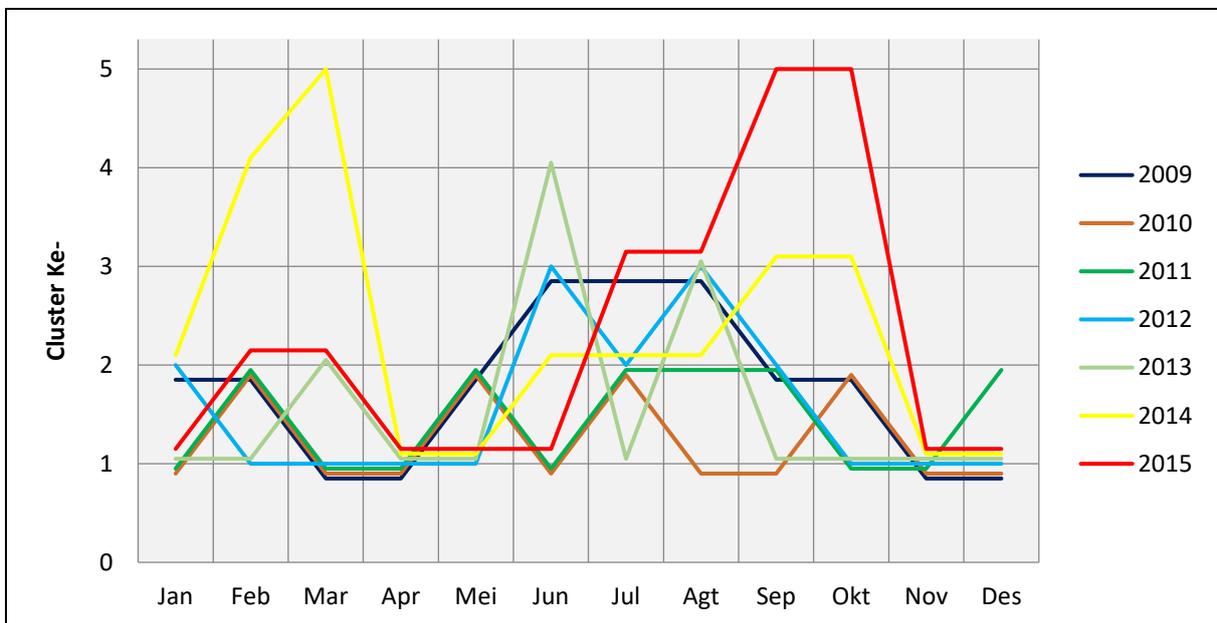
Berdasarkan peraturan pemerintah RI No.41 Tahun 1999 mengenai Pengendalian Pencemaran Udara, yang dimaksud dengan pencemaran udara adalah masuknya atau dimasukkannya zat, energi dan/atau komponen lain ke dalam udara ambient oleh kegiatan manusia sehingga mutu udara ambient turun sampai ke tingkat tertentu yang menyebabkan udara ambient tidak memenuhi fungsinya.

Data mining maksudnya adalah mengekstrak atau menambang pengetahuan dari sekumpulan data yang sangat banyak. Istilah yang lebih tepat untuk data mining seharusnya *knowledge mining from data* atau menambang pengetahuan dari data. Banyak istilah lain untuk data mining yang memiliki makna yang sama atau sedikit berbeda seperti *knowledge mining from data*, *knowledge extraction*, *data/pattern analysis*, *data archaeology*, and *data dredging* [5].

Algoritma *K-means* merupakan algoritma klusterisasi yang mengelompokkan data berdasarkan titik pusat kluster (*centroid*) terdekat dengan data. Tujuan dari *K-means* adalah pengelompokkan data dengan memaksimalkan kemiripan data dalam satu kluster dan meminimalkan kemiripan data antar kluster. Ukuran kemiripan yang digunakan dalam kluster adalah fungsi jarak. Sehingga pemaksimalan kemiripan data didapatkan berdasarkan jarak terpendek antara data terhadap titik *centroid* [6].

Tahap-tahap dari algoritma *K-means clustering* adalah sebagai berikut [5] :

1. Tentukan jumlah *cluster* (*k*) yang diinginkan sebagai *input*.
2. Tentukan titik pusat *cluster/centroid* secara acak sebanyak jumlah *cluster* yang di inginkan dari data yang tersedia.



Gambar 2. Kelompok dengan 5 cluster

3. Hitung jarak antara data dengan *centroid*. Pada penelitian ini menggunakan *Euclidean Distance* yaitu metode paling populer untuk mencari jarak terpendek antara data dengan *centroid* dengan rumus yaitu :

$$D(X_i, Y_j) = \sqrt{(P1_i - Q1_j)^2}$$

Dimana :

- $D(X_i, Y_j)$ = Jarak data *i* ke *centroid* *j*
- $P1_i$ = Variabel ke 1 pada data ke *i*
- $Q1_j$ = Variabel ke 1 pada *centroid* ke *j*

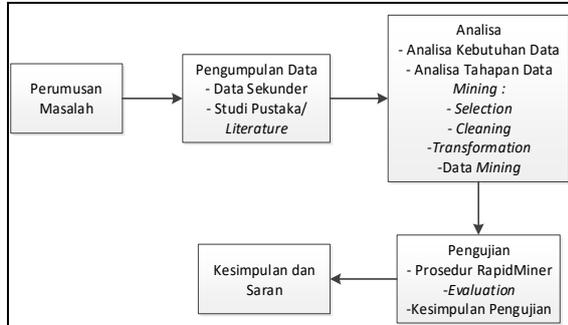
4. Kelompokkan data berdasarkan jarak terpendeknya antara data dengan *centroid* menjadi sebuah kelompok *cluster*.
5. Hitung rata-rata tiap kelompok *cluster* yang terbentuk untuk dijadikan sebagai *centroid*

yang baru dan ulangi perhitungan mencari jarak terpendek antara data dan *centroid* apabila *centroid* berubah dan perhitungan akan berhenti apabila *centroid* tidak mengalami perubahan.

Rapidminer merupakan perangkat lunak yang bersifat *opensource*. Rapidminer sebelumnya bernama YALE (*Yet Another Learning Environment*). Rapidminer dikembangkan pada tahun 2001 oleh Ralf Klein Kenberg, Ingo Mierswa dan Simon Fischer dari University of Dortmund ditulis dalam bahasa java.

METODOLOGI PENELITIAN

Metodologi penelitian adalah sistematis keseluruhan tahapan yang akan di laksanakan selama penelitian. Langkah-langkah dari metodologi penelitian yaitu sebagai berikut :

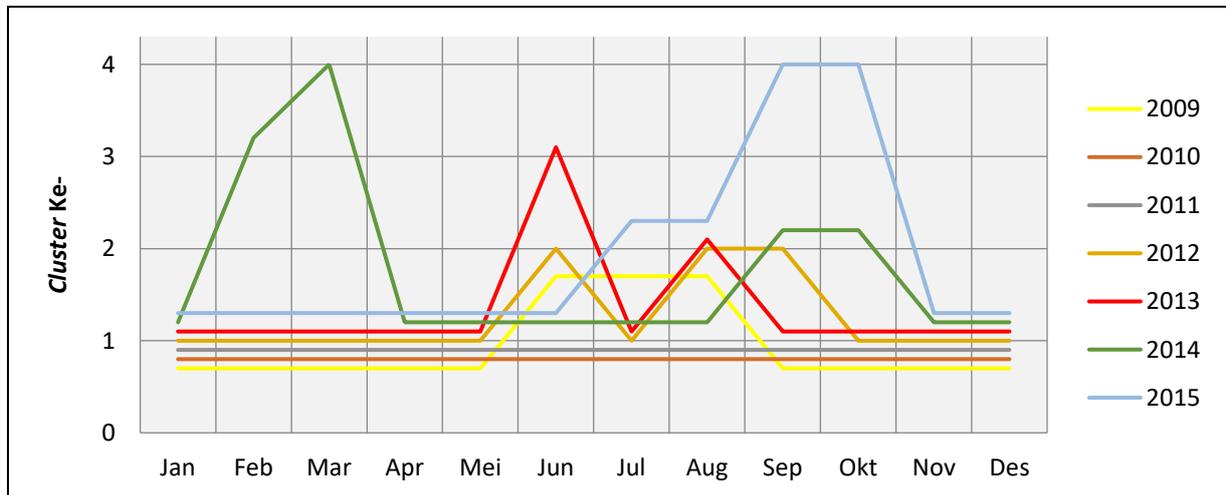


Gambar 3. Tahapan metodologi penelitian

ANALISA

Analisa merupakan suatu tahapan dimana dilakukan pembahasan untuk memahami permasalahan apa saja yang dihadapi dalam penelitian yang dilakukan. Dengan adanya tahapan analisa ini maka inti permasalahan yang ada dapat dijelaskan secara detail dan dapat dipelajari secara mendalam sehingga akan dapat dipahami dengan baik.

Data yang didapat dari Badan Lingkungan Hidup Kota Pekanbaru yaitu data dari tahun 2009 hingga 2015 dimana terdapat total berjumlah 16 variabel dan tiap variabel mempunyai *record* data setiap 30 menit.



Gambar 4. Kelompok dengan 4 cluster

Dari jumlah total 16 variabel data yang tersedia, yang digunakan untuk penelitian yaitu data dengan variabel PM_{10} dari tahun 2009 hingga tahun 2015 dimana tiap variabelnya mempunyai *record* data setiap 30 menit. Sehingga total jumlah data hari yaitu sebanyak 2.556 hari dan jika data polutan di *record* setiap 30 menit setiap harinya maka jumlah keseluruhan data polutan udara yaitu 122.688 *record* data.

The Integrated Air Quality Management for Metropolitan Areas
 Laporan harian kualitas udara di stasiun pemantau
 Daily Air Quality Report for Monitoring Stations

Tanggal / Date: 01.01.2009
 Kota / City: Pekanbaru
 Stasiun / Station: PEF2 (Sukajadi)

Laporan disiapkan di / Report prepared at: Regional Center Pekanbaru
 Laporan berdasarkan data setiap jam rata-rata / Report based on half-hour mean values

| Waktu / Time GMT+7 (WB) | PM10 ug/m3 | SO2 ug/m3 | CO mg/m3 | O3 ug/m3 | NO2 ug/m3 | NO ug/m3 |
|----------------------------|---------------|--------------|-------------|-------------|--------------|-------------|
| 15:30 | --- | 0,00 | 0,84 | 17,40 | 23,95 | 4,38 |
| 16:00 | 133,29 | 0,00 | 0,51 | 18,41 | 51,23 | 18,50 |
| 16:30 | 40,81 | 0,00 | 0,08 | 14,00 | 24,46 | 2,52 |
| 17:00 | 16,55 | 0,00 | 0,03 | 11,61 | 29,22 | 2,47 |
| 17:30 | 20,86 | 0,00 | 0,18 | 10,70 | 34,75 | 1,50 |
| 18:00 | 27,61 | 0,00 | 0,46 | 2,29 | 46,92 | 3,55 |
| 18:30 | 29,88 | 0,00 | 0,68 | 0,00 | 60,32 | 7,08 |
| 19:00 | 44,46 | 0,00 | 0,16 | 0,00 | 53,25 | 6,05 |
| 19:30 | 39,01 | 0,00 | 0,00 | 0,00 | 43,58 | 8,07 |
| 20:00 | 41,29 | 0,00 | --- | 0,91 | 37,22 | 8,13 |
| 20:30 | 27,45 | 0,00 | --- | 3,85 | 32,56 | 5,01 |
| 21:00 | --- | --- | --- | --- | --- | --- |
| 21:30 | --- | --- | --- | --- | --- | --- |
| 22:00 | --- | --- | --- | --- | --- | --- |
| 22:30 | --- | --- | --- | --- | --- | --- |

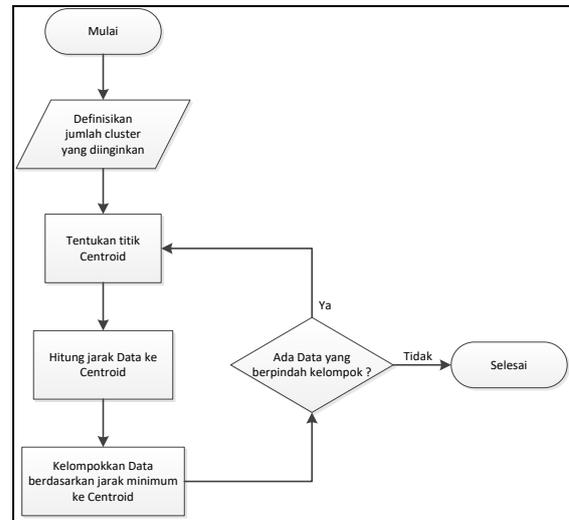
Gambar 5. Data variabel PM_{10} , SO_2 , CO, O_3 , NO_2 , NO

Setelah memilih data variabel PM_{10} dari tahun 2009 - 2015 kemudian di ambil rata-rata data variabel gas polutan per harinya. Data hasil dari tahapan *Selection* kemudian dilakukan pengecekan untuk mencari data yang masih mengandung *missing*

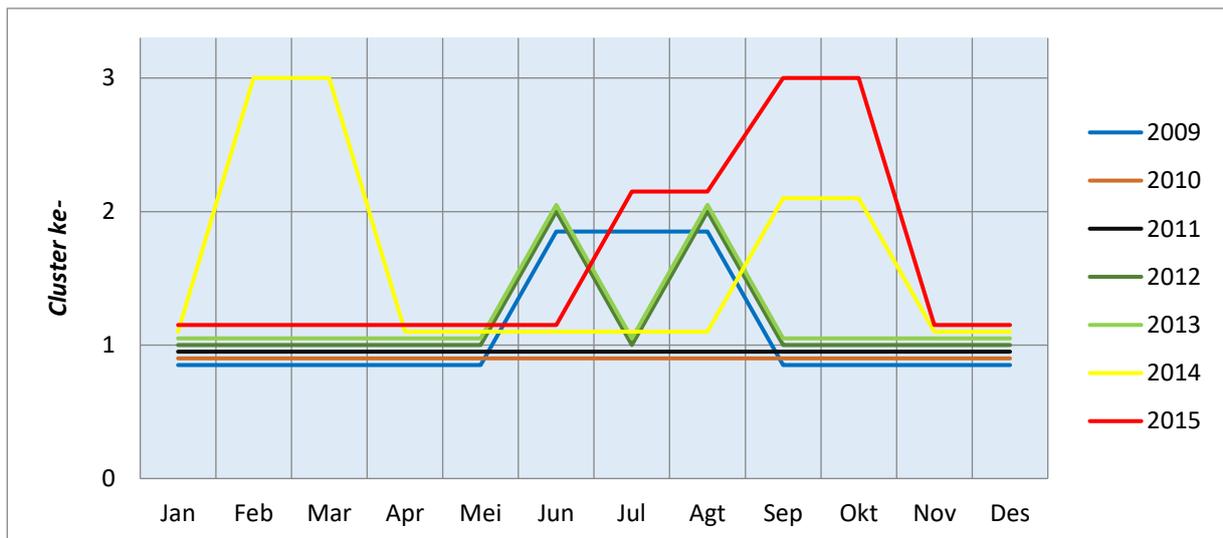
value. Data polutan udara yang memiliki *missing value* dapat disebabkan oleh beberapa hal seperti *maintenance* rutin, kalibrasi alat, gangguan teknis, anomali kerja alat, mati lampu dan sebagainya. Proses *cleaning* dari data yang memiliki *missing value* yaitu dengan menginputkan data rata-rata tiap variabel perbulan kedalam data yang memiliki *missing value*. Jumlah data polutan udara yang memiliki *missing value* dari tahun 2009-2015 berjumlah 397 record dari 122.688 record data. Setelah semua data yang memiliki *missing value* hilang barulah menghitung rata-rata tiap variabel per bulan yang akan di jadikan data masukan untuk menghitung kelompok *cluster* menggunakan metode *K-means clustering* yang akan di cari pada aplikasi Rapidminer 5.3.

Berdasarkan dari data variabel yang telah didapatkan setelah melalui beberapa proses sebelumnya maka selanjutnya implementasi dari data variabel tersebut kedalam metode yang digunakan yaitu *K-means* untuk mendapatkan

kelompok *cluster* yang di inginkan seperti yang telah di rencanakan sebelumnya.



Gambar 6. Diagram alir metode *K-means*



Gambar 7. Kelompok dengan 3 cluster

PENGUJIAN

Pengujian pertama dilakukan dengan mengujikan data polutan PM10 dengan menentukan *cluster* sebanyak 5 *cluster* dimana jumlah sebanyak 5 *cluster* berdasarkan jumlah kategori udara pada perhitungan Indeks Standar Polutan Udara. Hasil pengujian sebanyak 5 *cluster* dapat dilihat pada Gambar 4, dari hasil pengujian 5 *cluster* dapat dilihat bahwa *cluster* 1 merupakan kelompok dengan *volume* polutan antara 0 hingga 28,9. *Cluster* 2 merupakan kelompok dengan *volume* polutan antara 30 hingga 52,4. *Cluster* 3 merupakan kelompok

dengan *volume* polutan antara 58,3 hingga 100,2. *Cluster* 4 merupakan kelompok dengan *volume* polutan antara 130,5 hingga 174,8. *Cluster* 5 merupakan kelompok dengan *volume* polutan antara 247,1 hingga 310,3.

Pengujian selanjutnya dilakukan dengan jumlah *cluster* sebanyak 4 *cluster* dimana dengan menggunakan 4 *cluster* kelompok *cluster* yang terbentuk memiliki anggota dengan rentang *volume* polutan seperti pada Gambar 1 Pengaruh ISPU Untuk Setiap Parameter Pencemar. Dari hasil pengujian dengan menggunakan 4 *cluster* dapat dilihat bahwa *cluster* 1 merupakan kelompok dengan

volume polutan antara 0 hingga 50. *Cluster* 2 merupakan kelompok dengan *volume* polutan antara 50 hingga 100. *Cluster* 3 merupakan kelompok dengan *volume* polutan antara 100 hingga 200. *Cluster* 4 merupakan kelompok dengan *volume* polutan antara 200 hingga 300.

Pengujian selanjutnya dilakukan dengan jumlah *cluster* sebanyak 3 *cluster* sebagai perbandingan untuk melihat pola yang dihasilkan. Dari hasil pengujian dengan menggunakan 3 *cluster* dapat dilihat bahwa *cluster* 1 merupakan kelompok dengan *volume* polutan antara 0 hingga 52,4. *Cluster* 2 merupakan kelompok dengan *volume* polutan antara 58,3 hingga 130,5. *Cluster* 3 merupakan kelompok dengan *volume* polutan antara 174,8 hingga 310,3.

KESIMPULAN

Berdasarkan hasil analisa dan juga pengujian yang telah dilakukan maka beberapa hal yang dapat disimpulkan adalah sebagai berikut :

1. Dari hasil penelitian diketahui bahwa terdapat 16 bulan dari 84 bulan dari tahun 2009 – 2015 yang memiliki kadar polutan PM₁₀ yang termasuk ke dalam kategori sedang hingga sangat tidak sehat.
2. Dari hasil penelitian dapat dilihat bahwa rata-rata *volume* polutan PM₁₀ mengalami kenaikan dimulai pada bulan Juni, *volume* polutan rata-rata turun kembali pada bulan Oktober dan November.

Berdasarkan hasil yang didapatkan, saran dari penulis untuk pengembangan dari penelitian ini selanjutnya adalah:

1. Penelitian selanjutnya dapat dikembangkan dengan mencari pengaruh antara polutan udara dengan suatu kondisi seperti pengaruh polutan udara terhadap suatu penyakit tertentu, pengaruh polutan udara terhadap pertumbuhan tanaman tertentu, dan kondisi lainnya.
2. Penelitian selanjutnya diharapkan dapat menggunakan metode yang berbeda dengan penelitian ini untuk mendapatkan hasil yang lebih optimal terutama dalam pengujian menggunakan data polutan udara sesudah normalisasi.

REFERENSI

- [1] R. Field, "Indonesia's Dangerous Haze," 26 October 2015. [Online]. Available: <http://www.policyforum.net/indonesias-dangerous-haze/0>.
- [2] M. Syukur, 21 Oktober 2015. [Online]. Available: <http://news.liputan6.com/read/2345710/paru-paru-penuh-asap-bocah-di-riau-meninggal>. [Diakses 8 Oktober 2016].
- [3] B. Warsito, D. Ispriyanti dan H. Widayanti, "Clustering Data Pencemaran Udara Sektor Industri di Jawa Tengah Dengan Kohonen Neural Network," *Jurnal PRESIPITASI*, 2008.
- [4] A. Iizuka, S. Shirato, A. Mizukoshi, M. Noguchi, A. Yamasaki dan Y. Yanagisawa, "A Cluster Analysis of Constant Ambient Air Monitoring Data from the Kanto Region of Japan," *International Journal of Environmental Research and Public Health*, pp. 6844-6855, 2014.
- [5] J. Han dan M. Kamber, *Data Mining Concepts and Techniques Second Edition*, San Francisco: Morgan Kaufmann, 2006.
- [6] Asroni dan R. Adrian, "Penerapan Metode K-means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," *Jurnal Ilmiah Semesta Teknika*, pp. 76-82, 2015.