

PENERAPAN ATURAN ASOSIASI DENGAN ALGORITMA APRIORI UNTUK ANALISIS POLUTAN UDARA DI SURABAYA

Ricky Eka Putra¹⁾, Tutuk Indriyani²⁾

¹⁾Jurusan Teknik Informatika, Fakultas Teknik, Universitas Negeri Surabaya

Kampus Ketintang, Surabaya, 60231

Telp : (031) 8299563, Fax : (031) 8299553

²⁾Jurusan Teknik Informatika, Fakultas Teknologi Informasi, ITATS

Jalan Arif Rahman Hakim Nomor 100, Surabaya, 60117

Telp : (031) 5945043, Fax : (031) 5997244

E-mail : rickyeka25@yahoo.com¹⁾, tutuk223@gmail.com²⁾

Abstrak

Analisis asosiasi adalah teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item, atau aturan yang menyatakan asosiasi antara beberapa atribut sering juga disebut market basket analysis yang dapat diketahui melalui dua parameter yakni confidence dan support. Algoritma Apriori termasuk jenis aturan asosiasi pada data mining. Polusi udara memberikan dampak yang negatif terhadap kesehatan, dalam beberapa dekade terakhir polusi udara menjadi salah satu penyebab pemanasan global. Polusi udara di Surabaya disebabkan oleh bermacam-macam polutan baik alami maupun buatan manusia termasuk faktor meteorologis lainnya. Analisis dilakukan menggunakan aturan asosiasi dalam algoritma Apriori dengan data selama periode dua tahun, mulai Januari 2013 sampai dengan Desember 2014 yang diperoleh dari salah satu stasiun pemantau kualitas udara di Badan Lingkungan Hidup Surabaya. Setelah melalui tahapan proses dalam data mining dengan menggunakan algoritma Apriori maka akan dihasilkan beberapa aturan asosiasi dalam berbagai kondisi yang dapat dijadikan bahan kajian untuk beberapa periode selanjutnya.

Kata kunci: polusi udara, polutan udara, aturan asosiasi, algoritma Apriori

1. PENDAHULUAN

Pencemaran udara adalah kehadiran satu atau lebih substansi fisik, kimia atau biologi di atmosfer dalam jumlah yang dapat membahayakan kesehatan manusia, hewan dan tumbuhan, mengganggu estetika dan kenyamanan, atau merusak properti. Pencemaran udara merupakan perubahan keadaan udara karena pengaruh zat kimia atau biologi dalam jumlah yang banyak di atmosfer [1]. Pencemaran udara terjadi dari sumber alami ataupun disebabkan oleh manusia yang tidak memperhatikan keseimbangan ekosistem di sekitarnya. Pencemaran udara meliputi beberapa gangguan fisik, seperti polusi, panas, radiasi, polusi cahaya, dan polusi suara. Sifat alami yang dimiliki udara mengakibatkan dampak pencemaran udara dapat bersifat lokal, regional ataupun global. Pencemaran udara dewasa ini semakin menampakkan kondisi yang sangat memprihatinkan. Dampak dari pencemaran udara tersebut adalah menyebabkan menurunnya kualitas udara, yang berdampak negatif terhadap kesehatan manusia. Menurut perkiraan Depkes, pencemaran udara dan kebisingan akibat kegiatan industri dan kendaraan bermotor akan meningkat 2 kali pada tahun 2000 dari tahun 1990 dan 10 kali pada tahun 2020. Dalam tingkat regional, kualitas udara diukur dan dikontrol oleh pemerintah dalam Peraturan Pemerintah. Indeks standar polutan udara menjadi parameter terhadap kualitas kebersihan kota tersebut. Polutan udara dibagi menjadi 2 primer dan sekunder [2,3]. Penelitian ini akan membahas beberapa polutan primer yang mempengaruhi kualitas udara kota metropolitan Surabaya di antaranya adalah PM₁₀ (partikel debu), CO (karbon monoksida), NO₂ (nitrogen dioksida), SO₂ (sulfur dioksida) dan O₃ (ozon).

Berbagai penelitian mengenai estimasi atau prediksi pencemaran udara telah banyak dilakukan. Beberapa penelitian di antaranya adalah penggunaan jaringan syaraf tiruan [4,5,6]. Namun, beberapa penelitian tersebut masih berfokus pada 1-2 variabel polutan. Hal ini berbeda dengan penelitian ini yang menggunakan beberapa variabel polutan primer sebagai acuannya.

Dalam penerapan data mining, terdapat beberapa metode untuk menganalisis, memantau atau memperkirakan data hasil ukur selanjutnya ataupun dampaknya. Dalam penelitian ini, aturan asosiasi akan digunakan untuk menemukan aturan pada berbagai macam kondisi lingkungan terhadap udara dengan beberapa polutan. Dalam beberapa kasus, metode ini banyak digunakan untuk analisis pasar [7], namun hal tersebut hanya sebagian kecil contoh dari segudang tumpukan data yang ada. Oleh karena itu, penerapan aturan asosiasi dengan algoritma

Apriori ini akan digunakan untuk analisis lingkungan yakni polusi udara dan hubungannya dengan variabel meteorologi [8]. Penelitian ini menggunakan variabel-variabel polutan primer dengan variabel meteorologi di dalam-nya. Penelitian ini diharapkan mampu menemukan keterkaitan antara variabel pencemar tersebut dan mengevaluasi sumber polutan apa yang sering muncul serta bagaimana mengatasi dampak polusi tersebut dalam jangka waktu yang lama.

2. TINJAUAN PUSTAKA

Subbab ini akan membahas dua teori yang akan digunakan dalam penelitian ini. Dua teori tersebut antara lain algoritma Apriori serta variabel-variabel polutan udara yang digunakan dalam penelitian ini.

2.1 Algoritma Apriori

Data mining adalah suatu proses untuk menemukan informasi yang bermanfaat dari sekumpulan database besar yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik, matematika, kecerdasan buatan, dan *machine learning* [9]. Salah satu teknik *data mining* yang populer adalah *association rule*.

Association rule adalah salah satu teknik utama atau prosedur dalam Market Basket Analysis untuk mencari hubungan antar item dalam suatu data set dan menampilkan dalam bentuk *association rule* [10]. *Association rule* (Aturan Asosiatif) akan menemukan pola tertentu yang mengasosiasikan data yang satu dengan data yang lain. Untuk mencari *association rule* dari suatu kumpulan data, langkah pertama yang harus dilakukan adalah mencari *frequent itemset* terlebih dahulu. *Frequent itemset* adalah sekumpulan item yang sering muncul secara bersamaan. Setelah semua pola *frequent itemset* ditemukan, barulah mencari aturan asosiatif atau aturan keterkaitan yang memenuhi syarat yang telah ditentukan.

Association rule memerlukan suatu variabel ukuran yang ditentukan sendiri oleh pengguna untuk menentukan batasan sejauh mana atau sebanyak apa output yang diinginkan pengguna. *Support* dan *confidence* adalah sebuah ukuran kepercayaan dan kegunaan suatu pola yang telah ditemukan. Algoritma Apriori menggunakan pengetahuan frekuensi atribut yang telah diketahui sebelumnya untuk memproses informasi selanjutnya. Pada algoritma Apriori menentukan kandidat yang mungkin muncul dengan cara memperhatikan *minimum support* dan *minimum confidence* [11]. *Support* adalah nilai pengunjung atau prosentase kombinasi sebuah *item* dalam *database*. *Confidence* adalah nilai kepastian yaitu kuatnya hubungan antar *item* dalam sebuah Apriori. *Confidence* dapat dicari setelah pola frekuensi munculnya sebuah item ditemukan. Persamaan *support* dan *confidence* terlihat pada persamaan (1) dan (2).

$$\text{Support (A)} = \frac{\text{Jumlah transaksi mengandung A}}{\text{Total Transaksi}} \times 100\% \quad (1)$$

$$\text{Confidence } P(B|A) = \frac{\text{Total transaksi mengandung A dan B}}{\text{Transaksi mengandung A}} \times 100\% \quad (2)$$

Algoritma Apriori untuk mendapat *frequent itemset* [9] yaitu :

1. *Join* (penggabungan)
Proses ini dilakukan dengan cara pengkombinasian *item* dengan *item* yang lainnya hingga tidak dapat terbentuk kombinasi lagi.
2. *Prune* (pemangkasan)
Proses pemangkasan yaitu hasil dari *item* yang telah dikombinasikan kemudian dipangkas dengan menggunakan *minimum support* yang telah ditentukan oleh *user*.

2.2 Variabel Polutan Udara

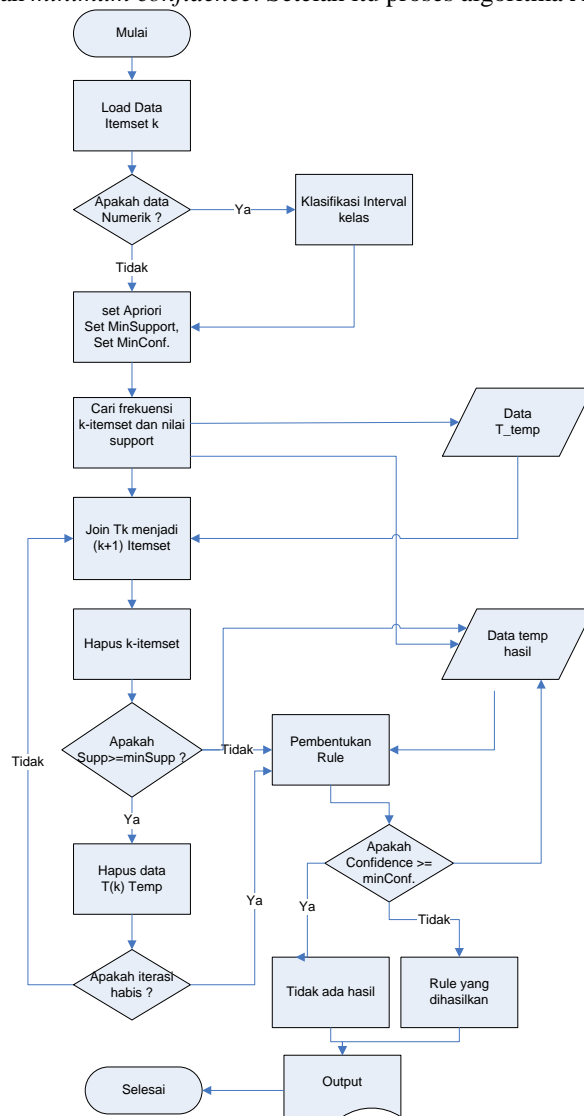
Udara sebagai komponen lingkungan yang penting dalam kehidupan perlu dipelihara dan dijaga kualitasnya. Sehingga dapat memberikan daya dukungan bagi makhluk hidup untuk hidup didalamnya. Pencemaran udara saat ini semakin menampakkan kondisi yang mengkhawatirkan. Sumber pencemaran udara dapat berasal dari berbagai kegiatan industri antara lain, industri transportasi, perkantoran dan perumahan. Secara global dampak polusi udara dalam skala yang lebih luas dapat menimbulkan pemanasan global sebagai dampak dari efek rumah kaca. Efek rumah kaca tersebut dapat menimbulkan anomali iklim atau perubahan iklim seperti saat ini yang banyak menyebabkan terjadinya bencana dan kegagalan produksi pertanian. Untuk itu perlu dilakukan pengaturan dan pengawasan terhadap polusi udara.

Perubahan iklim selain berdampak buruk pada lingkungan, juga berdampak buruk bagi kesehatan. Diantarnya buruk untuk jantung, lebih mudah terkena alergi, peristiwa alam ekstrim, kekeringan, pertumbuhan bakteri dan penyebaran penyakit. Maka dari itu untuk mengendalikan dampak pencemaran udara ditetapkan peraturan pemerintah [2,3]. Dari peraturan pemerintah tersebut, beberapa polutan dasar ditetapkan sebagai alat ukur

kualitas udara di daerah perkotaan diantaranya partikulat/debu (PM₁₀), Karbon Monoksida (CO), Sulfur dioksida (SO₂), Nitrogen Dioksida (NO₂) dan Ozon (O₃). Selain itu, penelitian ini juga menggunakan variabel meteorologi untuk melengkapi variabel-variabel polutan primer tersebut.

3. DESAIN DAN IMPLEMENTASI SISTEM

Diagram alir dan alur sistem penelitian ini terlihat pada gambar 1 dan gambar 2. Pencarian asosiasi pada polutan dimulai dari mengambil data ukur polutan dari *database* dengan memilih waktu pengukuran dan variabel yang akan diasosiasikan. Kemudian menentukan pengaturan algoritma Apriori dengan menentukan batasan *minimum support*, *maximum support* dan *minimum confidence*. Setelah itu proses algoritma Apriori dapat dilakukan.



Gambar 1. Diagram Alir Sistem Apriori Pada Polutan

Praproses data merupakan tahapan yang paling penting dalam *data mining*, data nyata yang didapat cenderung kotor (*noisy*) dan tidak konsisten. Praproses data antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan cetak (tipografi). Dalam data ini, data yang dianggap tidak normal jauh melebihi ambang batas akan disesuaikan sehingga diperlukan kalibrasi pada waktu tertentu sehingga ada kemungkinan data tersebut *misssing values*. Salah satu solusi jika terjadi *noisy* (data tidak valid) atau *missing values* yaitu dengan memberikan nilai rata-rata yang sama seperti pada periode sebelumnya atau setelahnya. Penelitian ini menggunakan solusi tersebut dalam mengantisipasi data yang kurang sesuai.

Transformasi data merupakan proses pengubahan atau penggabungan data kedalam format yang sesuai untuk diproses dalam *data mining*. Sering kali data yang akan digunakan dalam proses data mining belum dapat digunakan secara langsung. Oleh karena itu data tersebut harus dirubah formatnya atau digabungkan dengan data yang lain. Penentuan data polutan yang sesuai dengan peraturan daerah, nasional ataupun internasional dilakukan pada tahap ini. Ada enam variabel yang digunakan dalam penelitian ini. Enam variabel tersebut antara lain :

1. CO (Carbon Monoksida)

Data hasil pengukuran konsentrasi gas CO per 1 jam yang diukur dalam satuan $\mu\text{g}/\text{m}^3$ (mikrogram per meter kubik) atau biasanya mg/m^3 (miligram per meter kubik). Nilai pada polutan ini merupakan tipe data numerik sehingga harus di rubah dulu dalam bentuk nominal sesuai dengan standar interval data pada batas ambang polutan CO. Batas ambang polutan CO terlihat pada tabel 1.

2. PM₁₀ (Partikulat Matter)

Data hasil pengukuran konsentrasi PM10 atau debu berukuran kurang lebih 10 mikron per 1 jam yang di ukur dalam satuan $\mu\text{g}/\text{m}^3$ (mikrogram per meter kubik). Nilai pada polutan ini merupakan tipe data numerik sehingga harus di rubah dulu dalam bentuk nominal sesuai dengan standar interval data pada batas ambang polutan PM10. Batas ambang polutan PM10 terlihat pada tabel 2.

Tabel 1. Batas Ambang Polutan CO

Nomor	Nilai CO ($\mu\text{g}/\text{m}^3$)	Kategori ISPU
1	0 – 5	50
2	6 – 10	100
3	11 – 17	200
4	18 – 35	300
5	> 35	400

Tabel 2. Batas Ambang Polutan PM₁₀

Nomor	Nilai PM ₁₀ ($\mu\text{g}/\text{m}^3$)	Kategori ISPU
1	0 – 50	50
2	51 – 150	100
3	151 – 350	200
4	351 – 420	300
5	> 420	400

3. SO₂ (Sulfur Dioksida)

Data hasil pengukuran konsentrasi gas SO₂ per 1 jam yang diukur dalam satuan $\mu\text{g}/\text{m}^3$ (mikrogram per meter kubik). Nilai pada polutan ini merupakan data dengan tipe numerik sehingga perlu dirubah dahulu ke tipe nominal sesuai dengan standar interval data pada batas ambang polutan SO₂. Batas ambang polutan SO₂ terlihat pada tabel 3.

4. NO₂ (Nitrogen Dioksida)

Data hasil pengukuran konsentrasi gas NO₂ per 1 jam yang diukur dalam satuan $\mu\text{g}/\text{m}^3$ (mikrogram per meter kubik). Nilai pada polutan ini merupakan data dengan tipe numerik sehingga perlu dirubah dahulu ke tipe nominal sesuai dengan standar interval data pada batas ambang polutan NO₂. Batas ambang polutan NO₂ terlihat pada tabel 4.

Tabel 3. Batas Ambang Polutan SO₂

Nomor	Nilai SO ₂ ($\mu\text{g}/\text{m}^3$)	Kategori ISPU
1	0 – 19	50
2	20 – 49	100
3	50 – 79	200
4	80 – 129	300
5	> 129	400

Tabel 4. Batas Ambang Polutan NO₂

Nomor	Nilai NO ₂ ($\mu\text{g}/\text{m}^3$)	Kategori ISPU
1	0 – 200	50
2	201 – 400	100
3	401 – 800	200
4	801 – 2260	300
5	> 2260	400

5. O₃ (Ozon)

Data hasil pengukuran konsentrasi Ozon per 1 jam yang diukur dalam satuan $\mu\text{g}/\text{m}^3$ (mikrogram per meter kubik). Nilai pada polutan ini merupakan data dengan tipe numerik sehingga perlu dirubah dahulu ke tipe nominal sesuai dengan standar interval data pada batas ambang polutan O₃. Batas ambang polutan O₃ terlihat pada tabel 5.

6. Variabel meteorologi (kecepatan angin, arah angin, dan lain-lain)

Data hasil pengukuran variabel lain seperti kecepatan angin dan arah angin disekitar area tersebut yang diukur dalam ukuran meter per detik dengan range interval 0-5, 6-10, 11-15 m/s dan 16-20 serta arah angin yang diukur dalam satuan derajat 0-360 derajat. Data pada variabel ini tidak mempunyai ketetapan sehingga data yang diambil adalah nilai minimal, nilai tengah dan nilai maksimal pada suatu periode kelas intervalnya (misalnya satu tahun). Contoh kecepatan angin pada tabel 6 dilakukan dengan mengambil nilai rata-rata pada periode pengukuran yang terjadi selama setahun.

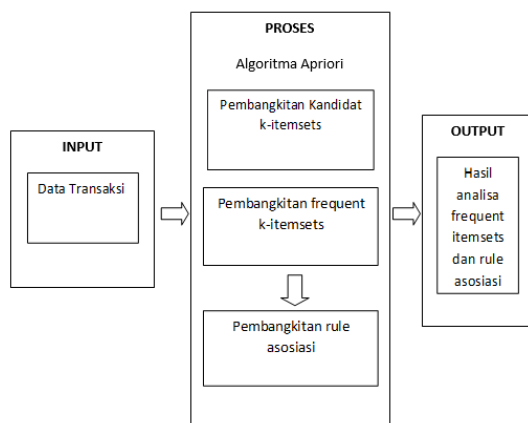
Praproses data yang telah dijelaskan sebelumnya dilakukan terlebih dahulu terhadap data. Tampilan praproses pada data terlihat pada gambar 3. Transformasi data dilakukan untuk mengubah data menjadi bentuk yang sesuai (seperti pada tabel masing-masing variabel data). Proses transformasi data terlihat pada gambar 4. Proses akhir dari penelitian ini yakni penerapan algoritma Apriori untuk membangun sebuah aturan berdasarkan data yang telah dirubah dapat dilihat pada gambar 5. Sistem tersebut akan dilakukan beberapa percobaan untuk mendapatkan hasil yang maksimal. Hasil uji coba dan evaluasi pada sistem dapat dilihat pada subbab berikutnya.

Tabel 5. Batas Ambang Polutan O₃

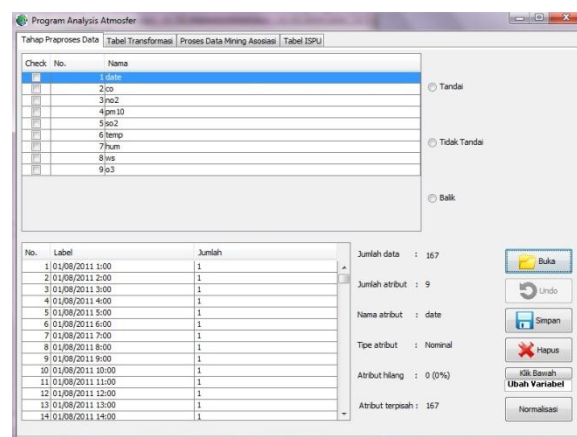
Nomor	Nilai O ₃ (µg/m ³)	Kategori ISPU
1	0 – 19	50
2	20 – 49	100
3	50 – 79	200
4	80 – 129	300
5	> 129	400

Tabel 6. Batas Ambang Polutan Variabel Meteorologi

Nomor	Nilai Ws (m/s)	Kategori
1	0 – 2	Rendah
2	3 – 4	Sedang
3	5 – 6	Tinggi



Gambar 2. Alur Sistem



Gambar 3. Praproses Data

Gambar 4. Transformasi Data

Gambar 5. Hasil Pembuatan Aturan Berdasarkan Algoritma Apriori

4. UJI COBA DAN EVALUASI

Analisis akhir yang dihasilkan antara bulan Januari 2013 dan Januari 2014 menunjukkan beberapa hal yakni :

1. Jumlah aturan yang dihasilkan berbeda, pada bulan Januari 2013 sebanyak 12 aturan dan pada bulan Januari 2014 sebanyak 50 aturan.
2. Level konsentrasi dari tiap polutan menunjukkan bahwa rata-rata kondisi udara menunjukkan pada level baik karena tidak terdapat konsentrasi polutan yang melebihi dari nilai paling rendah.
3. Contoh analisis/ pembacaan aturan ((IF)co='CO(0-5)' 744 (THEN) ==> so2='SO2(0-20)' 744 nilai **support 0.9** nilai **confidence:(1)** artinya Jika hasil ukur Konsentrasi CO perjam pada interval 0-20 maka juga akan diikuti SO2 pada konsentrasi (0-20).
4. Data hasil informasi yang didapat diatas dapat disimpulkan bahwa pada tahun 2014 terdapat penambahan PM10 dengan frekuensi pada level indeks tertentu yang tinggi.

Tabel 7. Large Itemset Akhir Bulan Januari 2013 dan Januari 2014

Nomor	Waktu	Large itemset akhir
1	Januari 2013	o3='O3(0-120)' so2='SO2(0-80)' co='CO(0-5)' frekuensi=744
2	Januari 2014	o3='O3(0-120)' pm10='PM10(0-50)' so2='SO2(0-80)' co='CO(0-5)' frekuensi=690

Hal ini berarti ada peningkatan atau penambahan parameter lain yang mempengaruhi tingkat kualitas udara di Surabaya. *Large itemset* akhir yang diperoleh dari penelitian ini dapat dilihat pada tabel 7.

5. SIMPULAN DAN SARAN

Dari hasil penelitian ini, kami dapat menarik beberapa simpulan dan saran mengenai perbaikan-perbaikan yang perlu dilakukan dalam penelitian selanjutnya. Simpulan dan saran ini diharapkan dapat berguna bagi peneliti lain dalam melakukan penelitian yang sejenis dan memberikan kontribusi lebih terhadap kelemahan-kelemahan di penelitian ini.

5.1 Simpulan

1. Algoritma Apriori ini mampu menemukan pola keterkaitan antar variabel polutan dengan nilai dukungan dan kepercayaan yang tinggi yang ditentukan pengguna.
2. Dengan inputan nilai kepercayaan (*minimum confidence*) yang rendah maka aturan yang dihasilkan semakin banyak. Sebaliknya semakin tinggi nilai kepercayaan yang diberikan akan sedikit pola asosiasi terbaik.
3. Dari hasil uji coba kombinasi antara *minimum confidence* dan *threshold support* antara 0.1 sampai dengan 1 (10-100%) maka dihasilkan nilai *minimum support* tertinggi adalah 0.95 atau 95 %.
4. Dari data percobaan perbulan Menunjukkan bahwa dengan nilai *support* 90% dan *minimum confidence* 90% maka akan dihasilkan sedikit aturan yang terbaik, dengan pola asosiasi dengan kombinasi *itemset* tiap periode menunjukkan perbedaan dan persamaan. Hasil analisis pertahun menunjukkan tidak ada penambahan variabel polutan yang berbeda.

5.2 Saran

Data polutan jika ditransformasikan sesuai standar polutan nasional kurang menunjukkan tren yang baik, karena hasil standar terlalu tinggi. maka penulis berinisiatif mengkombinasikan dengan standar internasional EPA dan CAI-Asia dengan ketentuan hasil ukur per jam. Penyebab tidak ditemukannya pola asosiasi yang variatif karena kurangnya variabel lain, seperti radiasi matahari, kelembaban udara, temperatur udara, curah hujan, dsb. Jika ada penambahan variabel lain maka kemungkinan aturan yang dihasilkan akan lebih baik. Oleh karena itu, perlu adanya penelitian lain untuk mengantisipasi kekurangan-kekurangan yang ada di dalam penelitian ini. Selain itu, penerapan metode lain dapat dilakukan dalam analisis dan prediksi polutan ini sebagai pembandingan dalam penelitian ini.

6. DAFTAR RUJUKAN

- [1] Redman, P., 2006. *Top Ten Countries Turning The Corner On Toxic Pollution 2014*. Blacksmith Institute.
- [2] Republik Indonesia. 1997. KEPKA BAPPEDAL No. 107 Tahun 1997 tentang Perhitungan dan Pelaporan serta Indeks Standar Pencemar Udara. Sekretariat Kabinet RI. Jakarta.
- [3] Republik Indonesia. 1997. KEPMEN Lingkungan Hidup No. 45 Tahun 1997 tentang Indeks Standar Pencemar Udara. Sekretariat Kabinet RI. Jakarta.
- [4]. Arifien, N.F., Arifin, S., Widjiantoro, B.L. 2012. *Prediksi Kadar Polutan Menggunakan Jaringan Syaraf Tiruan (Jst) Untuk Pemantauan Kualitas Udara Di Kota Surabaya*. Skripsi Sarjana ITS. Surabaya.
- [5]. Ozbay, B., Keskin, G. A., Dogruparmak, S. C., & Ayberk, S. 2011. *Predicting Tropospheric Ozon Concentration in Different Temporal Scales by Using Multilayer Perceptrons Models*. Ecological Informatics, 242- 247.
- [6] Warsito, B., Rusgiyono, A., & Amirillah, M. 2008. Permodelan General Regression Neural Network Untuk Prediksi Tingkat Pencemaran Udara Kota Semarang. Media Statistika, 43-51
- [7] Agung, Okky Priandi. 2005. *Penerapan Algoritma Apriori untuk Mendapatkan Pola Asosiasi Penjualan pada PT. Ladangku Lestari Subur Surabaya*. STIKOM Surabaya.
- [8] Parvinnia, Elham. 2007. *The Application of Association Rule Mining a Case Study: The Effect of Atmospheric Parameters on Air Pollution*. IADIS. ISBN: 978-972-8924-30-0.
- [9] Han, J. Kamber, M. 2001. *Data Mining : Concepts and Techniques*. San Fransisco: Morgan Kaufmann Publisher.
- [10] Kusriani, Taufik Luthfi, Emha. 2009. *Algoritma Data Mining*. Yogyakarta: Penerbit Andi.
- [11] Septyasmoro, Adam. 2007. *Implementasi Multiple Minimum Support dalam Mining Association Rules dengan Algoritma MSApriori*. STIKOM Surabaya.