

# ATURAN ASOSIASI DENGAN STANDAR STORET PADA MODEL PREDIKSI PARAMETER PENDUKUNG UJI KUALITAS AIR BAKU

Diana Purwitasari<sup>1)</sup>, Oktaviandra Pradita Putri<sup>2)</sup>, Wijayanti Nurul Khotimah<sup>3)</sup>

<sup>1)2)3)</sup> *Teknik Informatika, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember (ITS)  
Jl. Teknik Kimia, Ged. Teknik Informatika, Kampus ITS Sukolilo, Surabaya 60111, Indonesia*

<sup>1)</sup>diana@if.its.ac.id

<sup>3)</sup>wijayanti@if.its.ac.id

**Abstrak**—Uji laboratorium tentang kualitas air baku pada penyediaan dan pengolahan air bersih memperhatikan parameter air terkait faktor fisika, kimia dan biologi. Analisis kualitas air di laboratorium membutuhkan waktu. Usulan sistem akan mempercepat waktu dengan menganalisis catatan dataparameter air yang ada dalam rekam data PDAM. Aturan asosiasi pada sistem digunakan untuk melihat hubungan antara parameter air yang didahului praproses dengan mengubah data numerik ke data kategorikal berdasarkan standar STORage and RETrievalData Warehouse (STORET).Selanjutnya model prediksi parameter air yang dihasilkan dari data belajar akan disederhanakan terlebih dahulu sebelum validasi model dengan data uji. Pengujian model menggunakan data belajar menunjukkan rata-rata akurasi 70% dengan minimal *support-confidence* 30% data. Hasil model hubungan parameter air menggunakan rekam data PDAM dapat menjadi pendukung kebijakan di daerah tersebut dalam penyediaan dan pengolahan air bersih sebelum dilakukan uji kualitas laboratorium. Tanpa ada uji laboratorium beberapa nilai parameter faktor kimia tidak dapat diketahui. Meskipun demikian aturan yang dihasilkan sistem usulan tanpa uji laboratorium dapat memberikan akurasi 80%-95% dengan asumsi *missing values* nilai faktor kimiasetelah dicek manual dari narasumber pemilik data. Data uji coba menggunakan dataset kecil untuk mempermudah cek manual.

**Kata Kunci**— prediksi kualitas air, aturan asosiasi, storet

**Abstrak**—Raw Water (*Air Baku*) laboratory analysis is testing physical, chemical and bacteriological characteristics of water to ensure that water supply is clean, safe and ready for drinking water quality. Analyzing raw water quality in laboratorium needs more time. The proposed system could shorten the laboratory processing time by analyzing daily water production log. Association rule in the proposed system was used to generate relation model of water characteristics from the data log provided by local government owned water utilities (*PDAM, Perusahaan Daerah Air Minum*). The data was transformed first from numerical data into categorical data using STORage and RETrieval Data Warehouse (STORET) standard. Generated model needs to be simplified because some prediction rules could have the same interpretation. The generated parameter prediction model was sufficient to be used as the supporting data for any local policy made related to water supply and sanitation without additional costs from standard lab testing of water quality. Some water quality values of chemical characteristics need lab testing. Given the missing values of several chemical characteristics, the generated parameter prediction model still could give better accuracy of 80%-95%. Since *PDAM* staff manually validated the generated model, the experiments used small data set.

**Keywords**— water quality prediction, association rule, storet

## I. PENDAHULUAN

Tingkat pencemaran air Kali Surabaya semakin meningkat menyebabkan kualitas air menurun karena keterbatasan ketersediaan air bersih untuk masyarakat (Kaunang, Hidayat, & Guntarto, 2011). Salah satu contoh kasus pencemaran yang dilaporkan oleh Dinas Komunikasi dan Informatika Jawa Timur terjadi di Kali Surabaya. Pencemaran di duga akibat turbulensi debit air hujan sehingga menyebabkan menurunnya kandungan oksigen

(<http://kominformo.jatimprov.go.id/watch/37468> , Nov 2013). Hal tersebut menyebabkan kematian banyak ikan di sungai. Air Baku adalah air yang memenuhi baku mutu tertentu berasal dari sumber air permukaan, atau air tanah dan air hujan (SNI

6774:2008 Tata Cara Perencanaan Unit Paket Instalasi Pengolahan Air, Badan Standardisasi Nasional). Air baku dipompa menuju ke penampungan untuk pengendapan yang akan dialirkan untuk penyaringan, kemudian dilakukan pembunuhan bakteri sebelum terakhir didistribusikan ke masyarakat.

Adanya kemungkinan pencemaran air dan baku mutu yang harus dipenuhi, maka uji kualitas air dilakukan pada penyediaan dan pengolahan air bersih untuk mengetahui parameter air yang digunakan dalam tahap penetralan air. Hasil uji kualitas dapat disimpan dan digunakan sebagai informasi untuk pendukung keputusan kebijakan pihak terkait yang menangani penyediaan dan pengolahan air bersih. Berbagai pendekatan telah

dilakukan untuk membuat model analisis dari hasil uji kualitas air seperti pendekatan statistik (Omo-Irabor, Olobaniyi, Oduyemi, & Akunna, 2008), algoritma Genetika (Ayundyahrini, Abdul Kadir, & Gamayanti, 2013), algoritma Fuzzy (Angulo, et al., 2012) serta sistem pendukung keputusan (Pairunan, 2012).

Kualitas air diukur/ diuji berdasarkan parameter tertentu sesuai peraturan perundang-undangan yang berlaku Keputusan Menteri Negara Lingkungan Hidup No. 115/2003 tentang Pedoman Penentuan Status Mutu Air. Penjernihan air berdasarkan beberapa parameter saja belum cukup (Ayundyahrini, Abdul Kadir, & Gamayanti, 2013). Pendekatan statistik dengan analisa keterhubungan banyak parameter telah dilakukan (Omo-Irabor, Olobaniyi, Oduyemi, & Akunna, 2008) namun pendukung keputusan membutuhkan pernyataan berbentuk aturan agar memudahkan pembuat kebijakan (Pairunan, 2012) (Chang, Pongsanonea, & Ernest, 2012). Model analisis kualitas banyak parameter dengan pembentukan fuzzifikasi membutuhkan validasi manual para ahli (Angulo, et al., 2012) sedangkan ketersediaan ahli tersebut belum tentu dapat dipenuhi di setiap daerah.

Makalah ini menjelaskan penggunaan aturan asosiasi sebagai teknik analisa hubungan antar parameter air berdasarkan data-data yang ada. Penggunaan aturan asosiasi pada sistem yang diusulkan dalam makalah ini bertujuan untuk mempercepat waktu analisa catatan data-data parameter air yang ada dalam rekam data PDAM. Hasil model dari usulan sistem dapat dimanfaatkan sebagai pendukung kebijakan di daerah tersebut dalam penyediaan dan pengolahan air bersih sebelum dilakukan uji kualitas laboratorium sehingga mempersingkat waktu dan biaya.

Hasil analisa berbentuk model dapat diimplementasikan dengan mudah dalam pendukung keputusan berbentuk sistem atau kebijakan tertulis. Bahasan berikut menguraikan lebih detil mengenai penelitian terkait dengan pembentukan model prediksi parameter air serta permasalahan yang muncul. Kemudian dilanjutkan dengan bahasan tahapan pembentukan yang memberikan alternatif solusi masalah. Hasil uji coba model disampaikan pada bagian hasil dan pembahasan.

## II. TINJAUAN PUSTAKA

Prediksi parameter pendukung uji kualitas air baku dalam suatu sistem pendukung keputusan dapat menggunakan berbagai macam algoritma komputasi cerdas. Algoritma Genetika digunakan untuk estimasi dosis tawas berdasarkan parameter kekeruhan dan pH data air pada penjernihan air di Kota Surabaya (Ayundyahrini, Abdul Kadir, & Gamayanti, 2013). Uji coba dilakukan pada data-data kualitas air saat musim penghujan dan kemarau menunjukkan bahwa penggunaan tawas

memiliki karakteristik berbeda. Terdapat faktor lain yang mempengaruhi dosis tawas seperti lumpur atau limbah kimia.

Prediksi data numerik parameter kekeruhan dan pH air yang dilakukan belum menunjukkan pola hubungan antar parameter. Terdapat parameter air selain kekeruhan dan pH seperti *suspended solid* (SS), TDS, *chemical oxygen demand* (COD), *biochemical oxygen demand* (BOD), dan konsentrasi unsur kimia. Hubungan yang akan membentuk suatu model membutuhkan data berbentuk kategorikal untuk mempermudah pengenalan relasi antar parameter. Perubahan data numerik menjadi data kategorikal dilakukan dengan standar data STORET untuk menentukan status mutu air (Pairunan, 2012). Model yang digunakan adalah metode *time series moving average* dengan perubahan data menjadi skor STORET agar variasi nilai lebih terjaga pada parameter Sungai Saddang Sulawesi Selatan. Standar STORET (*STOrage and RETrieval*) berdasarkan sistem nilai *United States Environmental Protection Agency* (USEPA) banyak digunakan dalam penentuan klasifikasi mutu air (<http://www.epa.gov/storet/>).

Model serupa namun dengan *multivariate analytical method* karena tidak menggunakan proses transformasi skor STORET seperti *Principal Component Analysis* (PCA), *Factor Analysis* (FA) dan *Hierarchic Cluster Analysis* (HCA) menunjukkan adanya distribusi komponen kimia di daerah sekitar wilayah pengamatan (Omo-Irabor, Olobaniyi, Oduyemi, & Akunna, 2008). Komponen kimia tersebut dimungkinkan ada karena lokasi air berdekatan dengan tempat produksi minyak.

Hubungan antara parameter air seperti pH, jumlah mineral larut dalam air, kekeruhan, amonia dan *dissolved oxygen* (DO) dengan pendekatan fuzzy memberikan informasi yang diterjemahkan ahli menjadi aturan bagi penyedia air bersih (Angulo, et al., 2012). Sebelum *fuzzy inference system* tersebut diimplementasikan, validasi purwa-rupa aturan dilakukan manual oleh ahli. Aturan fuzzy yang dihasilkan akan mendeteksi adanya pembuangan limbah termasuk eutrofikasi akibat limbah fosfat berlebih yang dapat ditandai dengan banyaknya tanaman eceng gondok serta limbah yang menyebabkan peristiwa kematian ikan-ikan.

Oleh karena itu diperlukan suatu model yang dapat menunjukkan relasi antar banyak parameter kualitas air. Aturan asosiasi digunakan dalam makalah ini untuk melihat keberadaan relasi tersebut. Implementasi aturan asosiasi sering menggunakan Algoritma Apriori (Han, Kamber, & Pei, 2011). Namun demi menghasilkan model yang mudah dibaca maka data-data parameter air akan diubah dari numerikal menjadi kategorikal berdasarkan aturan STORET. Oleh karena itu

sebelum tahap pencarian hubungan antar parameter maka tahap pertama adalah pengolahan data dari numerikal menjadi kategorikal.

### III. METODE PENELITIAN

Dua tahapan utama dalam pembentukan model adalah transformasi data kategorikal dan pencarian hubungan antara parameter air.

#### A. Pengolahan Data Air dengan Standar STORET

Pengelompokkan data dilihat dari parameter air terkait faktor *fisika* (penglihatan, sentuhan, rasa, bau), *kimia* (bahan-bahan yang larut dalam air), dan *biologi* (keberadaan mikroorganisme dalam air). Data yang digunakan dalam makalah ini diambil dari data air sungai di Surabaya sebelum diolah selama 3 tahun (2010-2013) hasil pencatatan harian Pusat Penelitian dan Pengembangan (puslitbang) PDAM. Sebelum pengolahan air baku menjadi air bersih, PDAM melakukan uji kualitas sampel air di laboratorium untuk mengetahui jumlah kandungan setiap parameter air. Pada sumber data tercatat 20 parameter-parameter air yang meliputi:

1) 4 parameter terkait **faktor fisika** meliputi suhu, kekeruhan, warna dan suspended solid (SS). Nilai SS adalah padatan tersuspensi dalam air berupa bahan-bahan organik/ inorganik yang mempengaruhi kekeruhan air.

2) 16 parameter terkait **faktor kimia** meliputi alkalinitas, CO<sub>2</sub> bebas, dissolved oxygen (DO) atau kelarutan oksigen, nitrit, amonia, tembaga, fosfat, sulfida, besi, krom hexavalen, mangan, seng, timbal, chemical oxygen demand (COD), dan detergen. Nilai DO adalah jumlah oksigen terlarut dalam air. Air dengan oksigen kurang membuat hidup mikroorganisme tidak bertahan lama dan akan menimbulkan bau.

Asumsi bahwa kebutuhan air kota Surabaya tidak melebihi kebutuhan kota Jakarta, maka data acuan yang digunakan sebagai baku mutu air diambil dari Keputusan Gubernur Provinsi DKI Jakarta No.582/1995 tentang Penetapan Peruntukan dan Baku Mutu Air Sungai / Badan Air serta Baku Mutu Limbah Cair di Wilayah DKI Jakarta dan PP No.82/2001 tentang Pengelolaan Kualitas Air dan Pengendalian Pencemaran Air. Golongan air sesuai ketentuan yang berlaku diambil Golongan B yaitu air yang digunakan sebagai air baku air minum.

Tahap pengolahan data air dengan Standar STORET-USEPA untuk persiapan data numerik menjadi kategorikal adalah sebagai berikut:

1) Tentukan nilai batas minimal dan maksimal setiap parameter air berdasarkan baku mutu air Golongan B dalam Kep.Gub. DKI Jakarta No.582/1995.

2) Hitung nilai minimal, maksimal dan rata-rata per 10 hari data puslitbang PDAM untuk setiap parameter air: Terdapat total 720 data

(hari) dengan 500 data sebagai data belajar dan 220 sebagai data uji. Untuk data belajar akan dihitung 72 nilai minimal, maksimal dan rata-rata dari 20 parameter air.

3) Hitung skor setiap nilai minimal, maksimal dan rata-rata parameter air tersebut dengan standar STORET-USEPA sebagai acuan : Untuk semua nilai minimal, maksimal dan rata-rata per 10 data parameter air (Langkah 2) yang masih ada dalam batas nilai minimal-maksimal acuan (Langkah 1) akan diberikan skor 0. Apabila nilai diluar batas akan diberikan skor sebagai berikut:

Skor (Kategorikal)	Parameter Air Faktor		
	Fisika	Kimia	Biologi
Minimal/ Maks	-1	-2	-3
Rata-rata	-3	-6	-9

Sebagai contoh, nilai minimal-maksimal dari 10 data (hari) parameter kekeruhan=12,80-233 dan rata-rata 45,19. Nilai data acuan parameter kekeruhan (Langkah 1) adalah 0-100. Nilai minimal, maksimal dan rata-rata kekeruhan air dalam rentang waktu observasi 10 hari masih dalam batas data acuan. Maka total skor akhir dihitung dari min+maks+rata parameter suhu yang termasuk faktor fisika = 0+(-1)+0 = -1.

3) Klasifikasikan skor akhir menurut kelas mutu air STORET- USEPA (A, B, C, D) :

Kelas	Total Skor	Mutu Air
A	0	Baik Sekali
B	-1 s/d -10	Baik
C	-11 s/d -30	Sedang
D	>-30	Buruk

Jadi pada contoh 10 hari tersebut (Langkah 3) dilihat dari parameter kekeruhan air memiliki kelas mutu B atau tergolong baik. Langkah pengolahan data air dengan standar STORET dilakukan untuk menghasilkan model prediksi parameter yang mudah dibaca sehingga perlu perubahan data-data parameter air dari numerikal menjadi kategorikal.

#### B. Penentuan Hubungan antar Parameter Air dengan Aturan Asosiasi

Pada data-data catatan air akan dicari/ digali (*data mining*) pola yang sering muncul (*frequent pattern*) dengan algoritma apriori untuk mencari hubungan antar parameter atau disebut dengan aturan asosiasi (Han, Kamber, & Pei, 2011). Inisialisasi tahap penentuan hubungan antar parameter air tersebut adalah mengubah nilai numerik menjadi nilai kategorikal.

Pada contoh nilai parameter kekeruhan yang telah dibahas sebelumnya, variasi enam kategori yang mungkin adalah sebagai berikut:

Skor	min+maks+rata	Kategori
0	0+0+0	kekeruhan-1, C <sub>1</sub>
-1	(-1)+0+0 atau 0+(-1)+0	kekeruhan-2, C <sub>2</sub>
-2	(-1)+(-1)+0	kekeruhan-3, C <sub>3</sub>
-3	0+0+(-3)	kekeruhan-4, C <sub>4</sub>
-4	(-1)+0+(-3) atau 0+(-1)+(-3)	kekeruhan-5, C <sub>5</sub>
-5	(-1)+(-1)+(-3)	kekeruhan-6, C <sub>6</sub>

Sehingga contoh nilai numerik suatu data yang telah dirubah menjadi nilai kateogikal adalah data  $d1=\{\text{suhu-1, kekeruhan-2, warna-1, ss-6, ph-1, alkalinitas-1, co2-6...}\}$ .

Sebagai catatan, semua 20 parameter air memiliki enam jenis kategori berdasarkan skor STORET dengan variasi nilai minimal, maksimal dan rata-rata berada didalam atau diluar batas normal sesuai data acuan (Tabel 1).

TABEL 1. KATEGORI PARAMETER AIR

Kategori	Keterangan
$c_1$	Semua nilai <i>dalam</i> batas normal Contoh: alkalinitas-1
$c_2$	Nilai minimal <i>atau</i> maksimal <i>diluar</i> batas normal Contoh: do-2, kekeruhan-2
$c_3$	Nilai minimal <i>dan</i> maksimal <i>diluar</i> batas normal
$c_4$	Nilai rata-rata <i>diluar</i> batas normal
$c_5$	Nilai minimal <i>dan</i> rata-rata <i>atau</i> Nilai maksimal <i>dan</i> rata-rata <i>diluar</i> batas Contoh: kekeruhan-5
$c_6$	Semua nilai <i>diluar</i> batas normal

Kemudian akan dihasilkan kombinasi sejumlah  $n$  parameter air ( $n$ -frequent-itemset) yang masih memenuhi nilai ambang  $min$ -support dan  $min$ -confidence. Tahap pembentukan model parameter:

1) Tentukan kandidat  $n$ -itemset dari nilai kategori setiap parameter dengan nilai ambang support : Untuk setiap parameter air ( $p_{x \in 1..20}$ ) yang memiliki nilai kategori  $c_i$  masing-masing, misal parameter kekeruhan  $p_{kekeruhan} = \{c_1 \dots c_6\}$ , akan dihitung nilai supports ( $p_x \geq min_{support}$ ) dengan Eq. 1. Kombinasi parameter yang ditunjukkan pada Eq. 1 bisa ditambahkan sampai maksimal 20 parameter selama masih memenuhi syarat. Namun pada uji coba hanya 10-frequent-itemset yang memenuhi nilai support  $\geq 70\%$ .

$$support(p_x = c_i) = \frac{\# data_{p_x=c_i}}{\# data}$$

$$support(p_x = c_i \cap p_y = c_j) = \frac{\# data_{p_x=c_i \cap p_y=c_j}}{\# data} \tag{1}$$

Apabila pada parameter kekeruhan dengan kategori kekeruhan-2 dan kekeruhan-3 memenuhi  $min$ -support. Maka hanya dua kategori tersebut yang menjadi kandidat  $1$ -itemset parameter kekeruhan.

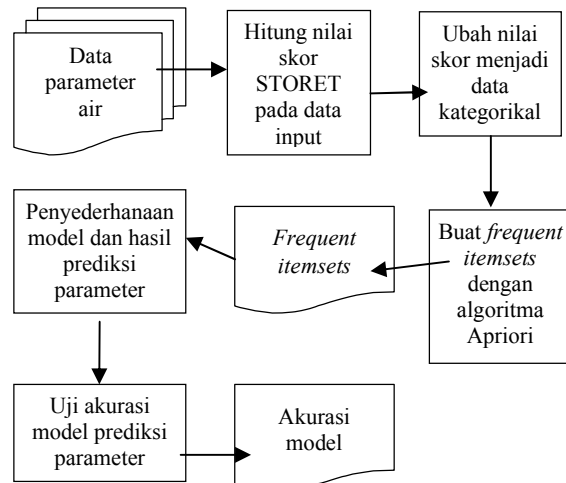
Sebagai contoh pada 3 parameter air berikut hanya ada 4 nilai kategori yang menjadi kandidat  $1$ -itemset sehingga untuk  $2$ -itemset terdapat variasi parameter menjadi 5 kandidat.

- |                  |                              |
|------------------|------------------------------|
| <b>1-itemset</b> | <b>2-itemset</b>             |
| • alkalinitas-1  | - alkalinitas-1, do-2        |
|                  | - alkalinitas-1, kekeruhan-2 |
|                  | - alkalinitas-1, kekeruhan-5 |
| • do-2           | - do-2, kekeruhan-2          |
|                  | - do-2, kekeruhan-5          |
| • kekeruhan-2    |                              |
| • kekeruhan-5    |                              |

2) Tentukan kandidat itemset berikut dengan  $n > 1$  dengan nilai ambang confidence (Eq. 2) : Pada Eq. 2 ditunjukkan hitung confidence untuk  $2$ -itemset dan  $3$ -itemset. Cara yang sama dapat dilakukan sampai  $n$ -itemset.

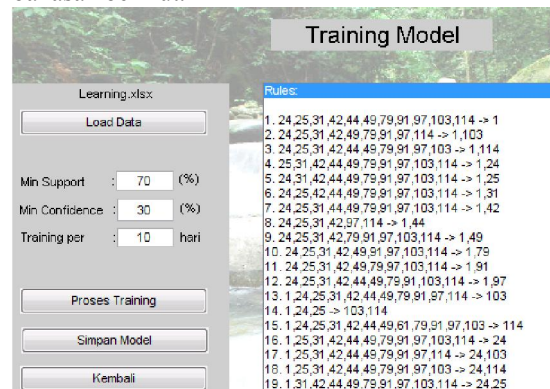
$$confidence(p_x = c_i, p_y = c_j) = \frac{\# data_{p_x=c_i \cap p_y=c_j}}{\# data_{p_x=c_i}}$$

$$confidence(p_x = c_i \cap p_y = c_j, p_z = c_k) = \frac{\# data_{p_x=c_i \cap p_y=c_j \cap p_z=c_k}}{\# data_{p_x=c_i \cap p_y=c_j}} \tag{2}$$

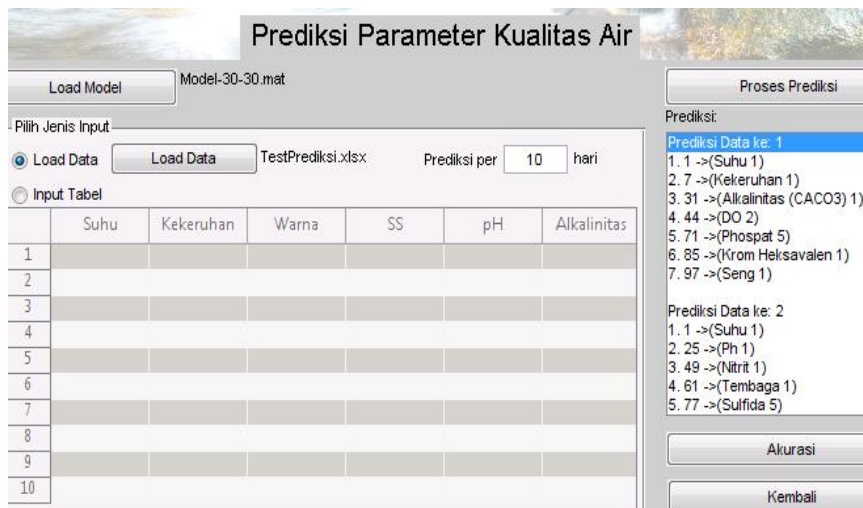


Gambar 1. Diagram alir tahapan pembentukan model prediksi parameter pendukung uji kualitas air baku

Tahapan metodologi penelitian ditunjukkan pada Gambar 1. Proses dimulai dari unggah data ke sistem prediksi yang akan diubah dari numerik ke kategorikal. Proses penggalian data yang dilakukan berikutnya dengan aturan asosiasi menggunakan Algoritma Apriori akan membentuk  $n$ -frequent-itemset. Tahap berikutnya adalah uji coba akurasi model yang akan diuraikan detail pada bahasan berikut.



Gambar 2. Contoh antar muka unggah data untuk identifikasi fitur parameter air



Gambar 3. Contoh antar muka hasil model prediksi dengan aturan asosiasi

#### IV. HASIL DAN PEMBAHASAN

Untuk keperluan uji coba digunakan antar muka pada Gambar 2 saat unggah data dan Gambar 3 saat pembentukan model prediksi. Tahap uji coba menggunakan dua jenis data yaitu data belajar (500 data) untuk pembentukan model prediksi dan data uji (220 data) untuk validasi model prediksi. Data-data dituliskan dalam file xls sesuai yang akan diunggah dengan antar muka Gambar 2. Data akan ditransformasi ke skor STORET per 10 data sehingga terdapat 50 data belajar dan 22 data uji. Pada eksperimen ini menggunakan dataset kecil yaitu kurang dari 30 data namun masih memenuhi minimum kebutuhan data standar distribusi normal (Walpole, Myers, Myers, & Ye, 2011). Dataset kecil juga akan mempermudah validasi model secara manual.

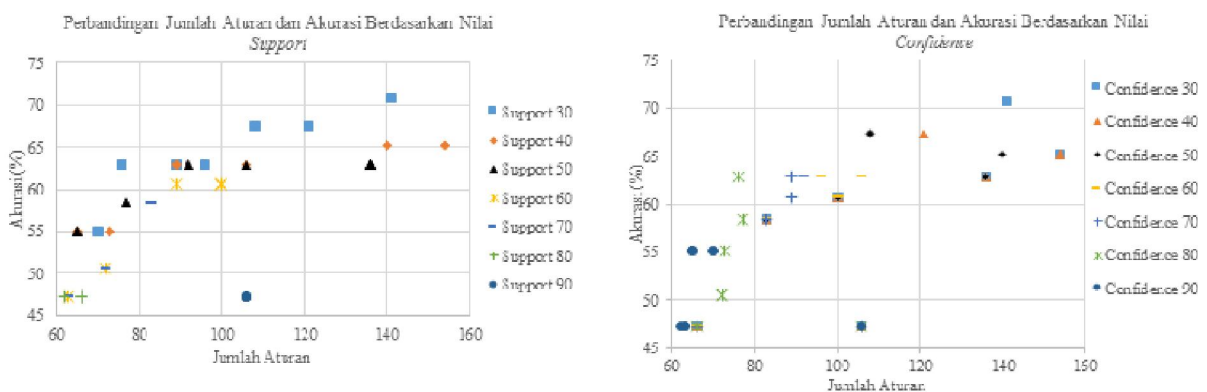
##### A. Penyederhanaan Model Prediksi

Skenario uji coba pembentukan aturan asosiasi menggunakan variasi nilai ambang *support* dan *confidence* dari penggunaan 30%...90% data belajar. Aturan yang dihasilkan bahkan dengan nilai

ambang *support*=70% dan *confidence*=70% sejumlah >1000. Untuk penyederhanaan aturan maka setiap kategori parameter air akan dikodefikasikan menjadi angka seperti yang ditunjukkan pada antar muka Gambar 2.

Sebagai contoh terdapat tiga aturan berbeda yang dihasilkan dengan nilai ambang bervariasi untuk *support* dan *confidence*. Untuk keperluan penyederhanaan maka akan diambil bentuk umum aturannya seperti yang ditunjukkan berikut. Ketiga aturan tersebut memiliki kodifikasi sama di bagian kiri yaitu 1,24,25 sehingga dapat dianggap sebagai aturan sama. Kemudian hasil di sebelah kanan terdapat kodifikasi yang sama 31 sehingga hasil dapat disederhanakan menjadi 1, 24, 25 → 31. Implementasi modul penyederhanaan merupakan bagian dari sistem. Validasi manual dilakukan sebelum implementasi modul penyederhanaan.

Catatan: 1 adalah kodifikasi dari suhu-1 yang berarti suhu pada 10 hari data masih dalam batas normal baik dilihat dari nilai minimal, maksimal dan rata-rata suhu air (Tabel 1).



Gambar 4. Grafik perbandingan jumlah aturan dan akurasi pada model prediksi parameter berdasarkan variasi nilai *support* dan *confidence*

$min_{support} = 70\%$	$min_{conf}$	penyederhanaan
1, 24, 25→31	100%	1, 24, 25→31
1, 24, 25→31, 103	90%	suhu & pH normal serta
1, 24, 25→31, 44	70%	SS tidak normal→ alkalinitas normal

Sehingga aturan awal untuk  $support=70\%$  dan  $confidence=70\%$  berkurang sampai sejumlah 10% dari 1000 lebih aturan menjadi kurang dari 100 aturan.

**B. Penyederhanaan Hasil Prediksi**

Saat uji coba prediksi meskipun model aturan sudah disederhanakan, namun aturan berbeda dapat menghasilkan prediksi parameter yang sama dengan kategori berbeda. Oleh karena itu penyederhanaan hasil prediksi juga perlu dilakukan berdasarkan nilai  $support$  yang lebih besar seperti contoh berikut. Pada parameter kekeruhan terdapat kategori berbeda yang dihasilkan yaitu

- semua nilai dalam batas normal (kekeruhan-1)
- nilai minimal dan rata-rata atau nilai maksimal dan rata-rata diluar batas (kekeruhan-5)

Pada kondisi tersebut maka langkah selanjutnya adalah melihat nilai  $support$  parameter dan memilih hasil dengan nilai terbesar.

prediksi	hasilawal	hasilakhir
7	kekeruhan-1	kekeruhan-1
11	kekeruhan-5	(kekeruhannormal)
		$min_{support} (kekeruhan-1)$ $> min_{support} (kekeruhan-5)$

**C. Skenario Uji Coba**

Skenario uji coba yang digunakan memiliki tujuan untuk mengetahui hubungan

- akurasi dan variasi nilai  $support-confidence$
- akurasi dan variasi jumlah aturan model prediksi parameter

Uji coba menggunakan data yang telah dihilangkan beberapa nilai parameter air sebagai *missing values*. Kondisi tersebut mungkin terjadi jika ada catatan parameter tidak lengkap atau tidak jelas sehingga bisa disalahartikan.

**D. Diskusi**

Model prediksi parameter yang dibentuk dari 50 data belajar telah diuji dengan 22 data dengan variasi nilai  $support$  dan  $confidence$  antara 30%-90% data. Setelah disederhanakan model hubungan parameter maka jumlah aturan yang dihasilkan berkisar di angka 100an (Tabel 2). Semakin banyak data menjadi prasyarat pembentukan parameter *n-frequent-itemset* yang ditandai dengan nilai ambang besar (misal 90%) pada  $support$  dan atau  $confidence$  maka jumlah aturan cenderung semakin sedikit dengan rata-rata akurasi menurun. Namun aturan tersebut memiliki banyak parameter yang mencapai *10-frequent-itemset*.

Sebagai contoh adalah 24, 25, 31, 42, 49, 79, 91, 97, 114→1,103 untuk nilai minimal  $support$

*Aturan Asosiasi dengan Standar STORET pada Model Prediksi Parameter Pendukung Uji Kualitas*

dan  $confidence$  menggunakan 90% data atau setidaknya 40 data memiliki nilai parameter air demikian.

Penjelasan aturan tersebut adalah jika kualitas air memiliki alkalinitas, besi, mangan, nitrit, seng, pH normal dan CO2, COD serta SS tidak normal maka parameter suhu dan timbal pada air tersebut akan normal. Namun uji coba validasi aturan tersebut dengan data uji menunjukkan akurasi <50% sehingga terlihat kecenderungan penurunan akurasi.

Semakin besar nilai ambang sebagai prasyarat maka aturan menjadi terlalu detil dengan data belajar (kondisi *overfitting*) sehingga cenderung tidak fleksibel terhadap data baru sebagai uji coba.

TABEL 2. JUMLAH ATURAN DAN AKURASI DENGAN VARIASI NILAI SUPPORT DAN CONFIDENCE PADA MODEL PREDIKSI PARAMETER

$Y_i$	$min_{support} = Y_i$		$min_{confidence} = Y_i$	
	Jumlah Aturan	Akurasi Rata (%)	JumlahAt uran	Akurasi Rata (%)
30	101	64,21	113	58,91
40	112	61,64	110	58,43
50	107	61,16	106	60,30
60	90	57,30	95	57,46
70	79	55,70	88	57,46
80	66	47,19	78	53,13
90	106	47,19	71	50,56

Pada Tabel 2 kecenderungan akurasi yang menurun juga dapat dilihat pada grafik Gambar 4. Tingkat akurasi yang tertinggi atau mencapai lebih dari 65% namun tidak melebihi 75% terjadi jika model dihasilkan dengan syarat nilai ambang  $support$  dan  $confidence$  menggunakan minimal 30% atau 15 data dari 50 data belajar. Oleh karena itu pada Tabel 3 ditunjukkan lebih detil variasi jumlah aturan dan akurasi dihasilkan dengan nilai ambang  $min_{support} = \{30\%, 40\%\}; min_{confidence} = \{30\% - 50\%\}$ .

TABEL 3. VARIASI NILAI SUPPORT DAN CONFIDENCE DENGAN AKURASI >65% PADA MODEL PREDIKSI PARAMETER

$min_{support}$	$min_{confidence}$	JumlahAturan	Akurasi (%)
30	30	141	70,79
	40	121	67,42
	50	108	67,42
40	30	154	65,17
	40	154	65,17
	50	140	65,17

Pengamatan lebih detil nilai minimal  $support=30\%$  menunjukkan akurasi yang lebih baik. Tujuan penelitian ini adalah menghasilkan model hubungan parameter air sehingga mengurangi uji laboratorium. Parameter air yang membutuhkan uji laboratorium adalah parameter terkait faktor kimia. Sehingga uji coba selanjutnya adalah menghilangkan beberapa nilai parameter

terkait faktor kimia yaitu Alkalinitas, CO<sub>2</sub> Bebas, DO, Nitrit, Amonia, Tembaga, Fospat, Sulfida, Besi, Krom Hexavalen, Mangan Seng, Timbal, dan COD. Hasil uji coba ditunjukkan pada Tabel 4 yang memperlihatkan bahwa model masih tetap dapat memberikan rata-rata akurasi >80% bahkan sampai 95% untuk beberapa parameter.

TABEL 4. AKURASI PARAMETER DENGAN MISSING VALUES PADA MODEL PREDIKSI PARAMETER

Rata-rata Akurasi (%)	Parameter Air Faktor Kimia (support=30%, confidence=30%-50%)
95.65	Alkalinitas, CO <sub>2</sub> Bebas, COD, Mangan, Seng, Timbal
86.96	Besi, Nitrit
82.61	Tembaga
52.17	Krom Heksavalen
36.96	Sulfida
34.78	dissolved oxygen (DO)
9.81	Fosfat

#### V. KESIMPULAN DAN SARAN

Makalah ini menguraikan model prediksi parameter air dengan aturan asosiasi berdasarkan data tanpa uji kualitas laboratorium. Prediksi membutuhkan transformasi data menjadi kategori dengan standar STORET. Uji coba menunjukkan nilai *support-confidence* didukung 30% data belajar untuk rata-rata akurasi 70%. Uji coba *missing values* untuk menekankan aspek tanpa uji laboratorium menghasilkan akurasi yang bagus sampai 95%. Penyederhanaan model prediksi parameter dengan teknik pengelompokan akan menjadi penelitian lebih lanjut.

#### DAFTAR PUSTAKA

Angulo, C., Cabestany, J., Rodríguez, P., Batlle, M., González, A., & Campos, S. (2012). Fuzzy expert system for the detection of episodes of poor water quality through continuous

measurement. *Expert Systems with Applications Vol 39*, 1011–1020.

Ayundyahrini, M., Abdul Kadir, R., & Gamayanti, N. (2013). Estimasi Dosis Aluminium Sulfat pada Proses Penjernihan Air menggunakan Metode Genetic Algorithm. *Jurnal Teknik POMITS Vol. 2, No. 2*, F390-F395.

Chang, N., Pongsanonea, N., & Ernest, A. (2012). A rule-based decision support system for sensor deployment in small drinking water networks. *Journal of Cleaner Production, Volumes 29–30*, 28-37.

Han, J., Kamber, M., & Pei, J. (2011). Ch.6 Mining Frequent Patterns, Associations, and Correlations: Basic Concepts and Methods. In J. Han, M. Kamber, & J. Pei, *Data Mining: Concepts and Techniques: Concepts and Techniques* (pp. 243-278). Morgan Kaufmann.

Kaunang, S., Hidayat, A., & Guntarto, B. (2011, Desember). Umbulan Water Supply: Memaksimalkan Sumber Air Jawa Timur. *Sustaining Partnership: Media Informasi Kerjasama Pemerintah dan Swasta, Bappenas*, pp. 18-19.

Omo-Irabor, O., Olobaniyi, S., Oduyemi, K., & Akunna, J. (2008). Surface and groundwater water quality assessment using multivariate analytical methods: A case study of the Western Niger Delta, Nigeria. *Physics and Chemistry of the Earth Vol. 33*, 666–673.

Pairunan, T. (2012). Perangkat Lunak Pendukung Keputusan Analisis Pengelolaan Kualitas dan Pengendalian Pencemaran Air Sungai. *Jurnal Ilmiah Sains Vol. 12 No. 2, Oktober 2012*, 105-111.

Walpole, R., Myers, R., Myers, S., & Ye, K. (2011). Ch6. Some Continuous Probability Distributions. In R. H. Ronald E. Walpole, *Probability and Statistics for Engineers and Scientists 9th*. Pearson.

[This Page Intentionally Left Blank]