

Aplikasi Analisa Multivariate dan Analisa Regresi Linier Berganda pada Proses Pengolahan Air Minum

Soehardjoepri

Jurusan Matematika FMIPA ITS, Surabaya

1. Pendahuluan

Air merupakan salah satu unsur penting yang sangat diperlukan bagi kelangsungan hidup manusia, baik digunakan untuk kebutuhan sehari-hari maupun untuk kegiatan lainnya. Agar kebutuhan tersebut dapat terpenuhi sebagaimana mestinya maka perlu diperhatikan masalah kualitas maupun kuantitasnya. Umumnya pengolahan air direncanakan agar dapat dipakai dalam jangka waktu yang cukup lama sehingga kualitas air yang cukup stabil sangat diperlukan. Adanya standard kualitas akan sangat membantu tercapainya kualitas tersebut. Persyaratan kualitas air yang layak dijadikan bahan baku untuk pengolahan air minum haruslah sesuai dengan Peraturan Menteri Kesehatan Republik Indonesia, dalam hal ini diwenangkan ke pihak Perusahaan air minum (PDAM). PDAM yang diambil sebagai tempat riset adalah di PDAM Surabaya yang saat ini mempunyai instalasi khusus yang bertugas untuk mengolah air baku menjadi air minum yang siap disalurkan ke masyarakat luas melalui suatu proses pengolahan. Pada proses pengolahan

ini terdapat variabel-variabel yang mempengaruhinya. Oleh karena karakteristik kualitas pada dasarnya kompleks, maka dilakukan analisis multivariate untuk memeriksa hubungan antar karakteristik maupun untuk pengendalian kualitasnya. Salah satu syarat utama yang sering digunakan sebagai ukuran kualitas air adalah kekeruhan dimana semakin tinggi kekeruhan berarti semakin buruk kualitas air yang dihasilkan dan dapat menyebabkan biaya pengolahan air minum menjadi mahal. Oleh karena itulah maka diperlukan model kekeruhan air dari variabel-variabel pembentuknya.

Dalam penelitian ini digunakan beberapa metode analisis yaitu:

- a. Analisis Multivariate
- b. Analisis Peta Kendali
- c. Analisis Kemampuan Proses
- d. Pembuatan Model Kekeruhan Air

2. Metode Analisis

2.1. Analisis Multivariate

Pada analisis komponen utama dan analisis faktor yang diamati pada PDAM Surabaya adalah X_1 (Prosentase Lumpur sungai), X_2 (kekeruhan intake), X_3 (dosis aluminium sulfat), X_4 (dosis chloor), X_5 (dosis kaporit), X_6 (dosis cupri sulfat), X_7 (dosis kaolin), X_8 (dosis kalium permanganat), X_9 (banyaknya pencucian), X_{10} (prosentase Lumpur accelerator) dan X_{11} (kekeruhan air kota). Untuk menetapkan jumlah komponen utama yang diambil digunakan prosentase kumulatif dari keragaman total yang diterangkan

Hasil perhitungan yang diperoleh dari bantuan SPSS untuk kesebelas variabel diperoleh 3 buah komponen utama. Komponen tersebut adalah kombinasi linier dari variabel asal yang saling berkorelasi, serta memaksimumkan sisa keragaman total.

Dari kesebelas komponen utama yang diamati dengan melihat nilai eigen dan prosentase keragaman maka diperoleh 3 komponen utama yang memegang peranan penting. Ketiga komponen utama ini secara bersama-sama dapat menerangkan keragaman data sebesar 73% dengan rincian sebagai berikut:

- Komponen utama pertama menerangkan keragaman total sebesar 46,5% dengan nilai eigen sebesar $l_1 = 5,17$
- Komponen utama kedua menerangkan keragaman total sebesar 15,7% dengan nilai eigen sebesar $l_2 = 1,73$
- Komponen utama ketiga menerangkan keragaman total sebesar 10,8% dengan nilai eigen sebesar $l_3 = 1,18$

Ketiga komponen utama ini membentuk matriks loading faktor 1, yang tidak berkorelasi satu sama lain dan nilai-nilainya merupakan koefisien korelasi antara variabel dengan faktor-faktor tersebut. Setiap faktor dapat diinterpretasikan sebagai berikut:

| faktor | variabel |
|--------|-------------------------------------|
| 1 | X_3, X_4, X_5, X_6, X_7 dan X_8 |
| 2 | X_1, X_2 dan X_{11} |
| 3 | X_9 dan X_{10} |

Dimana faktor 1 merupakan faktor kimiawi, faktor 2 merupakan faktor fisik dan faktor 3 merupakan faktor filtrasi.

2.2. Analisis Peta Kendali

Pengendalian kualitas dapat dilakukan dengan bantuan peta kendali. Dari peta kendali ini dapat dibaca karakteristik proses produksi berdasarkan hasil pengukuran. Sebelum evaluasi produk menggunakan peta kendali multivariate, terlebih dahulu dilakukan pengelompokan data. Dengan analisis komponen utama dan analisis faktor diperoleh 3 faktor data.

Peta kendali multivariate hanya bias dibuat jika data memenuhi asumsi berdistribusi normal multivariate.

Ho : Data berdistribusi normal multivariate

H1 : data tidak berdistribusi normal multivariate

Indikasi distribusi data mengikuti distribusi multivariate normal untuk p variabel adalah:

- Sekurang-kurangnya 50% data dari d_j^2 kurang dari sama dengan $\chi_{(p,50\%)}^2$.
- Plot antara kuadrat jarak d_j^2 dengan $\chi_{(p,50\%)}^2$; $j = 1, 2, \dots, n$ dan p adalah banyaknya variabel.

Berdasarkan data yang ada maka hasil yang didapatkan dengan pengujian asumsi distribusi normal multivariate sebagai berikut:

| $\chi_{(p,50\%)}^2$ | $d_j \leq \chi_{(p,50\%)}^2$ | Prosentase |
|---------------------|------------------------------|------------|
| 10,3410 | 77 | 62,6% |

Jadi data berdistribusi normal multivariate karena $d_j \leq \chi_{(p,50\%)}^2$ lebih dari 50%.

Tahap selanjutnya adalah pembentukan peta kendali, sebagai berikut:

(a) **Peta Kendali Faktor 1**

Pada faktor 1 terdapat tujuh titik berada di atas batas kendali $T^2 = 18,92$

yaitu pada obeservasi ke-6, ke-25, ke-48, ke-51, ke-80, ke-115 dan ke-119. Hal ini menunjukkan adanya indikasi out of control. Jika kondisi ini terjadi maka perlu menyelidiki adanya variasi yang terjadi melalui selisih terbesar antara data yang keluar dengan nilai rata-rata dari masing-masing variabel. Adapun hasil perhitungan nilai selisih pengamatan faktor 1 adalah sebagai berikut:

| No. | X_3 | X_4 | X_5 | X_6 | X_7 | X_8 |
|-----|-------|--------|--------|---------|--------|---------|
| 6 | 12,1 | 0,1344 | 0,1634 | 0,02608 | 2,0711 | 0,03488 |
| 25 | 3,3 | 0,0056 | 0,0066 | 0,00392 | 0,0889 | 0,00512 |
| 48 | 0,1 | 0,0744 | 0,0834 | 0,0608 | 1,0889 | 0,01488 |
| 51 | 2,7 | 0,0244 | 0,0966 | 0,00392 | 0,0111 | 0,00488 |
| 80 | 8,4 | 0,0956 | 0,2356 | 0,2392 | 0,3889 | 0,03512 |
| 115 | 13,5 | 0,0844 | 0,1034 | 0,02606 | 0,8111 | 0,03488 |
| 119 | 5,1 | 0,0444 | 0,0534 | 0,01608 | 0,3889 | 0,02488 |

Dari tabel di atas menunjukkan bahwa X_3 yang menyebabkan adanya kondisi out of control. Adapun sumber penyebabnya adalah karena karakteristik air baku yang berubah-ubah sehingga mengganggu proses koagulan bersama-sama dengan aluminium sulfat. Oleh karena penyebab sumber variasi berasal dari assignable causes maka data tersebut dapat dikeluarkan dan dibuat peta kendali multivariate penyesuaiannya.

(b) **Peta Kendali Faktor 2**

Pada faktor 2 terdapat dua titik yang berada di luar batas kendali $T^2 = 13,52$ yaitu pada observasi ke-5 dan ke-79. Hal ini menunjukkan adanya kondisi out of control maka perlu dicari sumber variasinya, seperti pada faktor 1 sehingga didapatkan hasil perhitungan nilai selisih pengamatan faktor 2 sebagai berikut:

| No. | X_1 | X_2 | X_{11} |
|-----|--------|--------|----------|
| 5 | 0,0752 | 14,826 | 0,0946 |
| 79 | 0,4248 | 7,084 | 0,0346 |

Dari tabel di atas menunjukkan bahwa penyebab proses tidak terkendali adalah variabel X_2 . Adapun sumber penyebabnya adalah karakteristik air baku yang berubah-ubah. Setelah diketahui sumber variasinya maka dibuat peta kendali multivariate penyesuaiannya.

(c) **Peta kendali faktor 3**

Pada faktor 3 terdapat satu titik di luar kendali $T^2 = 11,36$ yaitu titik pada observasi ke-76. Hal ini menunjukkan kondisi out of control sehingga dicari sumber variasinya, seperti halnya pada faktor 1 dan faktor 2 yaitu didapatkan hasil sebagai berikut:

| No. | X_9 | X_{10} |
|-----|-------|----------|
| 76 | 0,691 | 2,4317 |

Dari tabel tersebut menunjukkan bahwa X_{10} yang menyebabkan kondisi out of control. Kondisi ini disebabkan oleh karakteristik air baku yang berubah-ubah sehingga menyebabkan prosentase Lumpur kurang atau lebih dari batas ketentuan. Oleh karena penyebab kondisi tersebut berasal dari assignable causes maka dapat dibuat peta kendali multivariate penyesuaiannya.

2.3. Analisis Kemampuan Proses

Adanya situasi dimana perusahaan memiliki batas atas dan bawah untuk hasil pengukuran. Analisa yang digunakan guna mengetahui kemampuan perusahaan dalam rangka mengevaluasi produk apakah sesuai dengan standar perusahaan adalah dengan menggunakan indeks kemampuan proses (Cp) dan indeks performa proses (Cpk). Adapun pembahasan hasil analisis tersebut adalah sebagai berikut:

- *Kemampuan proses faktor 1.*

Nilai Cp dan Cpk faktor 1 yang terdiri dari variabel X_3 , X_4 , X_5 , X_6 , X_7 dan X_8 adalah cukup besar (> 1), hal ini menunjukkan bahwa proses berjalan baik (kapabel), akan tetapi masih perlu perhatian khusus pada variabel X_3 yaitu aluminium sulfat karena proses dalam kondisi hati-hati. Proses ini juga didukung dengan adanya peta kendali multivariate yang berada dalam keadaan terkendali setelah dilakukan penyesuaian, sehingga secara umum proses dapat dikatakan kapabel.

- *Kemampuan proses faktor 2.*

Hasil perhitungan menunjukkan bahwa proses kapabel karena nilai Cp untuk variabel X_1 , X_2 dan X_{11} cukup bagus (> 1) dan peta kendali multivariate setelah penyesuaian satu kali dalam keadaan terkendali.

- *Kemampuan proses faktor 3.*

Pada faktor 3 yang terdiri dari dua variabel yaitu X_9 dan X_{10} terlihat bahwa nilai Cp untuk X_9 cukup besar (> 1), sedangkan untuk X_{10} dengan Cp yang < 1 menunjukkan bahwa proses tidak kapabel. Dalam hal ini berarti bahwa proses filtrasi prosentase Lumpur melebihi batas ketentuan akibat karakteristik air baku yang berubah-ubah.

2.4. Pemodelan Kekeruhan Air Kota

Pada analisis regresi ini pemilihan model terbaik didasarkan pada prosedur eliminasi langkah mundur (Backward elimination) dengan memasukkan seluruh variabel yang terkait dan mengeluarkan variabel-variabel yang tidak signifikan dalam persamaan agar efisien.

Model regresi dikatakan terbaik jika variable-variabel dalam regresi tersebut berpengaruh nyata terhadap variable respon. Metode Backward elimination ini digunakan khususnya untuk melihat semua variable dalam persamaan tanpa kehilangan sesuatu.

Tahap I

Pada tahap awal dalam metode ini adalah dengan meregresikan semua variabel terhadap variabel respon dan diperoleh persamaan:

$$Y = 1,13 - 0,185X_1 + 0,000979X_2 - 0,0237X_3 + 0,323X_4 - 0,00135X_5$$

dengan

- Y : Kekeruhan air kota
- X_1 : pemakaian aluminium sulfat
- X_2 : prosentase Lumpur sungai
- X_3 : kekeruhan intake
- X_4 : prosesntase Lumpur predicanteur
- X_5 : banyaknya pencucian

Prosedur ini pada hakikatnya mencoba membuang semua variable X_i yang tidak signifikan, untuk itu maka perlu dihitung sumbangan masing-masing X_i kepada jumlah kuadrat regresi seolah-olah masing-masing variable tersebut dimasukkan terakhir kali dalam persamaan regresi.

Melalui uji t dengan hipotesa: $H_0 : \beta_i = 0$ dan $H_1 : \beta_i \neq 0$; didapatkan nilai t_{hitung} sebagai berikut:

| Variabel | t_{hitung} | $t_{tabel} = t_{121, 0,025}$ | Kesimpulan |
|----------|--------------|------------------------------|--------------|
| X_1 | 7,68 | 1,979 | Tolak H_0 |
| X_2 | 1,12 | 1,979 | Terima H_0 |
| X_3 | 0,50 | 1,979 | Terima H_0 |
| X_4 | 10,10 | 1,979 | Tolak H_0 |
| X_5 | 0,40 | 1,979 | Terima H_0 |

Dari uji t terlihat bahwa tidak semua variable X_i signifikan atau memberikan kontribusi yang berarti terhadap model.

Tahap II

Setelah dilakukan uji t pada tahap awal, selanjutnya dicari variabel yang signifikan untuk mendapatkan model yang terbaik. Model analisis regresi linier berganda didapatkan sebagai berikut:

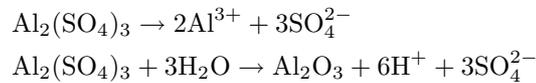
| No. | Model |
|-----|---|
| 1 | $Y = 1,13 - 0,185X_1 + 0,000979X_2 - 0,0237X_3 + 0,323X_4 - 0,00135X_5$ |
| 2 | $Y = 1,10 - 0,183X_1 + 0,00105X_2 - 0,0247X_3 + 0,324X_4$ |
| 3 | $Y = 1,07 - 0,183X_1 + 0,000751X_2 + 0,323X_4$ |
| 4 | $Y = 1,11 - 0,129X_1 + 0,324X_4$ |

Berdasarkan uji t maka variabel yang signifikan adalah X_1 dan X_4 dan ditinjau dengan tabel di atas maka model sementara yang digunakan adalah:

$$Y = 1,11 - 0,129X_1 + 0,324X_4$$

Model yang didapatkan tersebut belum tentu mewakili keadaan yang sebenarnya. Untuk itu maka dilakukan pengujian model dengan menggunakan serentak dan uji individu. Dari kedua uji tersebut ternyata model yang terpilih tersebut merupakan model terbaik dengan variasi keragaman data yang mendukung model yaitu sebesar 76%. Kejernihan air didapatkan pada saat X_i terletak antara 0,573 – 5,78125 dimana 0,573 didapatkan pada saat $y = 0$ dan 5,78125 didapatkan pada saat $y = 1$. Sedangkan untuk X_4 semakin mendekati 0 maka kekeruhan air kota yang dihasilkan semakin baik. Pada saat $y = 1$ (kondisi air buruk) maka batasan untuk X_4 yang dapat ditoleransi adalah 0,0005 – 3,0864.

Model ini dapat diinterpretasikan bahwa semakin tinggi prosentase Lumpur predicator maka semakin tinggi pula kekeruhan air kota yang dihasilkan dan semakin banyak aluminium sulfat yang digunakan maka semakin rendah kekeruhan air kota yang dihasilkan karena aluminium sulfat adalah zat koagulan yang dapat mengikat partikel koloid sehingga dapat menurunkan kekeruhan air. Adapun reaksi ionnya adalah sebagai berikut:



3. Kesimpulan

Dari pembahasan penelitian ini dapatlah diperoleh kesimpulan bahwa:

- Berdasarkan hasil analisis factor dengan analisis komponen utama dari 11 variabel dapat direduksi menjadi tiga factor. Faktor-faktor yang berpengaruh yaitu factor kimiawi, faktor fisik dan factor filtrasi.
- Pemodelan kekeruhan air kota dari variable-variabel yang mempengaruhinya diterangkan oleh model sebagai berikut:

$$Y = 1,11 - 0,129X_1 + 0,324X_4$$

Dengan model ini dapatlah ditentukan ukuran kualitas air.

Pustaka

- [1] Drapper, Norman (1992), *Analisis Regresi Terapan*, Edisi Kedua, PT. Gramedia, Pustaka Utama, Jakarta
- [2] Johnson, Wichren., (1992), *Applied Multivariate Statistical Analysis*, Third Edition, Prentice Hall, New Jersey
- [3] Montgomery, Douglass C., (1998), *Pengantar Pengendalian Kualitas Statistik*, Edisi Kelima, Gadjah Mada University Press, Yogyakarta
- [4] Morrison, Donald F., (1990), *Multivariate Statistical Methods*, Third Edition, Mc Graw-Hill, New York