# REGRESI LINIER BERGANDA TERMODIFIKASI UNTUK DATA SPEKTRUM PADA LARUTAN KONSENTRASI GLUKOSA, SUKROSA, DAN FRUKTOSA

Yusuf Kurniawan, Leopoldus Ricky Sasongko, Hanna Arini Parhusip
Program Studi Matematika, Fakultas Sains dan Matematika,
Universitas Kristen Satya Wacana
Jl. Diponegoro 52-60, Salatiga 50711, Jawa Tengah
e-mail: 662015001@student.uksw.edu, leopoldus.sasongko@staff.uksw.edu

***Abstrak***

*Penelitian ini bertujuan untuk memperoleh model pendekatan data spektrum suatu larutan yang berisi konsentrasi glukosa, sukrosa, dan fruktosa dengan menggunakan regresi linier berganda. Data spektrum yang didapatkan dari Pusat Studi (Pusdi) Aplikasi Near Infrared (NIR) FSM UKSW merupakan hasil penembakan dua hingga tiga gram dari larutan itu oleh sinar infrared pada alat spektrometer dengan panjang-panjang gelombang tertentu yang selanjutnya data spektrum tiap gelombang yang diperoleh dianalisis melalui analisis regresi linier berganda berdasarkan informasi parameter-parameter regresi, koefisien determinasi, dan signifikansi tiap parameter pada model. Fungsi parameter regresi yang bergantung pada panjang gelombang diperoleh dengan menginterpolasi parameter seluruh model yang telah didapat di setiap panjang gelombang melalui interpolasi polinomial Newton yang selanjutnya diperoleh model pendekatan data spektrum sebagai tujuan penelitian ini. Uji Kolmogorov-Smirnov satu sampel dilakukan untuk menguji distribusi Normal baku pada distribusi error, selisih hasil yang diperoleh model pendekatan terhadap data spektrum yang dimiliki pusdi aplikasi NIR FSM UKSW. Dengan distribusi error yang sebagian besar model mengikuti distribusi Normal baku, model yang selanjutnya disebut model regresi linier berganda termodifikasi dapat dikatakan cukup baik untuk memodelkan data spektrum. Dengan didapatkannya model regresi linier ini dapat diestimasi nilai – nilai data spektrum pada konsentrasi dan panjang gelombang tertentu.*

***Kata Kunci:*** *Data Spektrum; Interpolasi Polinomial; Regresi Linier Berganda.*

## 1. PENDAHULUAN

Pusat Studi (pusdi) Aplikasi *Near Infrared* (NIR) FSM UKSW merupakan pusdi yang didirikan dengan tujuan mengembangkan sistem pengukuran terkait kontrol kualitas suatu produk dengan berbasis spektroskopi NIR (Prasetya 2016) .Dalam pengamatan penulis selama Praktek Kerja Lapangan (PKL) pada bulan Mei sampai dengan Juli tahun 2018 di pusdi NIR, pusdi ini melakukan penelitian mengenai perolehan data spektrum pada larutan yang mengandung konsentrasi glukosa, sukrosa, dan fruktosa. Mula-mula tiga larutan itu, dengan masing-masing larutan memiliki kadar konsentrasi tertentu, dicampurkan lalu diambil dua hingga tiga gram dari campuran tiga larutan itu untuk selanjutnya diunggah pada alat spe-ktrometer guna ditembak oleh sinar *infrared* dengan panjang gelombang tertentu. Hasil pengukuran yang ditampilkan spektrometer disebut data spektrum.

Kendala tenaga dan dana menyebabkan pusdi NIR FSM UKSW hanya melakukan pengukuran data spektrum terhadap campuran tiga larutan dengan konsentrasinya adalah triplet, yang mana $x_i = 0, 0.1, 0.2, 0.4, 0.5, 0.6, 0.8,$ $\forall i = 1, 2, 3,$ dan panjang gelombang *infrared* $w = 4000, 4004,$

$4008, \ldots, 9996$.Masalah selanjutnya dalam penelitian yang dilakukan pusdi NIR adalah bagai-mana menentukan data spektrum, untuk larutan dengan konsentrasi $(x_1, x_2, x_3)$ dan panjang gelombang $w$ tertentu, yang belum diujikan berdasarkan data hasil pengukuran yang dimilikinya.

Penelitian ini bertujuan untuk memperoleh model pendekatan data spektrum yaitu dinyatakan sebagai suatu fungsi multivariat$f\big(w, (x_1, x_2, x_3)\big)$, fungsi terha-dap panjang gelombang $w$ dan triplet besar konsentrasi glukosa, sukrosa, serta fruktosa $(x_1, x_2, x_3)$. Model regresi linear berganda diusulkan dalam penelitian ini untuk model pendekatan $f\big(w, (x_1, x_2, x_3)\big)$yang mana dimodifikasi sedemikian rupa sehingga model pendekatan data spektrum dinyatakan oleh

$$f\big(w, (x_1, x_2, x_3)\big) \cong b_0(w) + b_1(w)x_1 + b_2(w)x_2 + b_3(w)x_3$$

## 2. METODE PENELITIAN

Data yang digunakan merupakan data simulasi (tiruan) yang dibangkitkan oleh pusdi NIR FSM UKSW berdasarkan data glukosa, sukrosa, dan fruktosa yang telah diteliti.

Data digunakan untuk mengestimasi model regresi linier berganda tiap $w$ tertentu agar diperoleh 1500 model akibat nilai $w$ yang dimiliki adalah sebanyak 1500 nilai yaitu $w = 4000, 4004, 4008, \ldots, 9996$. Model regresi adalah model matematis yang dapat digunakan untuk menaksir nilai-nilai suatu peubah yang tergantung pada nilai-nilai dari satu atau lebih peubah bebas (Disa, 2011). Regresi linier adalah metode yang digunakan untuk menentukan fungsi linier yang paling sesuai dengan kum-pulan titik data bivariat yang diketahui yaitu $(x_1, y_1)$, $(x_2, y_2)$, $(x_3, y_3)$, ..., $(x_n, y_n)$. Pernyataan matematis untuk regresi linier (sederhana) di-nyatakan oleh

$$y = b_0 + b_1 x + \varepsilon$$

dengan $\varepsilon = y - (b_0 + b_1 x)$ adalah galat atau *error* yang mengikuti distribusi pa-rametrik $N(0, \sigma_\varepsilon^2)$, Normal dengan parameter *mean* $\mu = 0$ dan variansi galat $\sigma_\varepsilon^2$. Bentuk umum model regresi linear berganda dengan $k$ peubah bebas adalah

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$$

dengan $Y$ adalah peubah tak bebas yang nilainya bergantung terhadap nilai-nilai dari peubah-peubah bebas $X_1, X_2, X_3, \ldots, X_k$ dan parameter $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ disebut sebagai parameter/koefisien regresi.

Tiap model tersebut akan diperoleh parameter-parameter regresi, koefisien determinasi, dan signifikansi tiap parameter per model. Estimasi parameter regresi dapat dilakukan dengan meminimumkan *residual sum of squares* (RSS) yang di-definisikan oleh

$$\mathrm{RSS}(\boldsymbol{\beta}) = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

Berdasarkan metode *Ordinary Least Square* (OLS) seperti yang dijelaskan (Pratama, 2012; Amrin, 2016; Kutner, 2004), parameter regresi yang me-minimumkan (2.5) adalah

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{Y})$$

Dan koefisien determinasi $R^2$ dihitung oleh

$$R^2 = 1 - \frac{\mathrm{RSS}(\widehat{\boldsymbol{\beta}})}{(\mathbf{Y}-\overline{\mathbf{Y}})^T(\mathbf{Y}-\overline{\mathbf{Y}})}$$

Tanpa memperhatikan signifikansi semua peubah bebas yang dimiliki, penyesuai-an nilai $R^2$, atau disebut $R^2$ *adjusted* (Setiawan, 2017), dihitung oleh

$$R^2\ adjusted = 1 - \frac{n-1}{n-(k+1)}(1-R^2)$$

Sedangkan Untuk $100(1-\alpha)\%$ tingkat kepercayaan interval suatu parameter regresi $\beta_i$, untuk $i = 0, 1, 2, \dots, k$, dapat diperoleh dari

$$\hat{\beta}_i \pm \mathrm{SE}(\hat{\beta}_i)t_{\alpha/2,n-2}$$

Signifikansi $\widehat{\beta}_i$, untuk $i = 0, 1, 2, \dots, k$, dapat ditentukan berdasarkan nilai $p_{value}$ yang bermakna peluang $\beta_i$ ada di $\widehat{\beta}_i \pm \mathrm{SE}(\widehat{\beta}_i)t_{\alpha/2,n-2}$.

Setelah didapatkan 1500 model regresi linier berganda, dilakukan analisis terhadap parameter-parameter regresi dengan memperhatikan signifikansi para-meter, apabila terdapat parameter yang tidak signifikan, langkah selanjutnya ada-lah memperhatikan nilai koefisien determinasi. Jika nilai koefisien determinasinya tin-ggi, maka paramater itu tetap digunakan pada model. Jika nilai koefisien de-terminasi tergolong rendah, maka parameter tidak digunakan pada model, atau dihapus.

Interpolasi fungsi parameter yang bergantung pada panjang gelombang dilakukan dengan mengambil beberapa titik data (tidak sebanyak 1500 titik). Interpolasi dilakukan menggunakan metode interpolasi Newton. Interpolasi poli-nomial adalah sebuah metode untuk menaksir (mengestimasi) nilai di antara titik-titik data yang tepat (Nugroho,2009). Bentuk umum fungsi polinomial derajat (order) *n* adalah:

$$P_n(z) = a_0 + a_1 z + a_2 z^2 + \dots + a_n z^n$$

Secara umum bentuk interpolasi polinomial, untuk $n + 1$ titik data, ($z_0$, $f_0$), ($z_1$, $f_1$), ... , ($z_n$, $f_n$) dapat dicocokkan dengan suatu polinomial berderajat *n* yang mempunyai bentuk.

$$y = a_0 + a_1(z-z_0) + a_2(z-z_0)(z-z_1) + \cdots + a_n\prod_{i=0}^{n}(z-z_i)$$

Validasi di tahap ini dilakukan guna mengukur seberapa baik model data spektrum terhadap data yang dimiliki. Model data spektrum yang diperoleh dapat dikatakan baik jika *error* yang diperoleh dari model data spektrum dan data yang dimiliki pada triplet $(x_1, x_2, x_3)$ dan $w$ tertentu berdistribusi Normal dengan $\mu = 0$ dan suatu variansi $\sigma^2$. Uji yang dilakukan untuk *error* tersebut dilakukan melalui uji Kolmogorov-Smirnov satu sampel. Uji *Kolmogorov Smirnov* satu sampel mer-upakan *goodnest of fit test* yang menguji hipotesis, dengan hipotesis $\mathbf{H_0}$: data me-ngikuti distribusi parametrik $\widehat{G}(t; \Omega)$, sedangkan $\mathbf{H_1}$: data tidak mengikuti dis-tribusi parametrik $\widehat{G}(t; \Omega)$ dimana $\Omega$ telah terlebih dahulu telah diestimasi (Greselda,2015). Statistik uji *Kolomogorov-Smirnov* dinotasikan $D_n$ yang me-nyatakan perbedaan terbesar antara fungsi distribusi empirik dan fungsi distribusi paramerik dugaan, yaitu

$$D_n = max\{D_n^-, D_n^+\}$$

dengan

$$D_n^- = \max_{i=1,2,\ldots,n}\left[\frac{i}{n} - \widehat{G}(y_i; \Omega)\right] \text{ dan } D_n^+ = \max_{i=1,2,\ldots,n}\left[\widehat{G}(y_i; \Omega) - \frac{i-1}{n}\right]$$

dan $y_i, i = 1,2,\ldots n$ adalah data yang diurutkan kecil ke besar. Pengambilan kesimpulan uji ini adalah $\mathbf{H_0}$ ditolak jika $D_n$ melebihi batas kuantil Kolmogorov-Smirnov, $D = d_\alpha\left(n^{-1/2} + 0.11n^{-1/2} + 0.12\right)^{-1}$, yang mana nilai $d_\alpha = 1.224, 1.358;$ atau $1.628$ untuk $\alpha$ berurutan $\alpha = 0.10, 0.05;$ atau $0.01$. Atau, $\mathbf{H_0}$ ditolak jika $p_{value} = \Pr[D_n \leq D] < \alpha$.

Diharapkan model data spektrum yang berupa model regresi linier bergan-da termodifikasi dapat diperoleh sehingga dapat mengestimasi data spektrum untuk triplet konsentrasi glukosa, sukrosa, dan fruktosa $(x_1, x_2, x_3)$ dan panja-ng gelombang $w$ tertentu.

## 3. HASIL PENELITIAN DAN PEMBAHASAN

Dalam penentuan model regresi linier berganda pada data spektrum ini me-mperhatikan perolehan parameter-parameter dan signifikansinya, serta koefisien korelasi. Misalkan untuk panjang gelombang $w = 4000$ dan permutasi triplet $(x_1, x_2, x_3)$, dapat diperoleh model regresi linier bergandanya sebagai berikut:

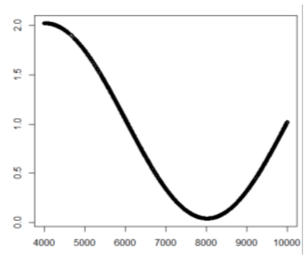
```
Call:
lm(formula = gel ~ x1 + x2 + x3)
Residuals:
Min        1Q    Median        3Q       Max
-0.029747 -0.014070  0.001695  0.014093  0.029779

Coefficients:
Estimate Std. Error  t value Pr(>|t|)
(Intercept)   2.022627   0.002186  925.307   <2e-16 ***
x1            0.149719   0.003145   47.607   <2e-16 ***
x2           -0.350219   0.003145 -111.361   <2e-16 ***
x3           -0.001552   0.003153   -0.492    0.623
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.01548 on 339 degrees of freedom
Multiple R-squared:  0.9774,    Adjusted R-squared:  0.9772
F-statistic:  4889 on 3 and 339 DF,  p-value: < 2.2e-16
```

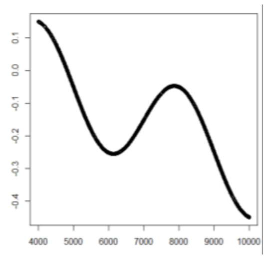
Pada `Call:` di atas diperoleh parameter model regresi linier berganda saat panjang gelombang $w = 4000$ adalah $b_0 = 2.02263$, $b_1 = 0.14972$, $b_2 = -0.3502$, dan $b_3 = -0.0016$. Parameter $b_0$, $b_1$, dan $b_2$ signifikan di tingkat $\alpha = 0.001$ yang di-tandai tiga bintang " *** " yang berarti interval kepercayaan $100(1-\alpha)\%$ para-meter-parameter tersebut tidak memuat 0 (nol), namun $b_3$ tidak signifikan dika-renakan interval kepercayaannya memuat nilai 0 (nol). Koefisien determinasi ya-ng diperoleh dari `Multiple R-squared:0.9774` dan `Adjusted R-squared:0.9772` yang tergolong tinggi (dekat ke nilai 1) yang berarti model regresi linier berganda baik, terlepas peubah bebas $x_3$ memiliki parameter $b_3$ yang tidak signifikan. Dengan begitu, parameter $b_3$ tetap digunakan dalam model data spektrum saat panjang gelombang $w = 4000$.

Langkah yang dijelaskan di atas kemudian dilakukan untuk data spektrum saat panjang gelombang yang lain hasilnya ditampilkan pada Tabel 3.1. Ditun-jukkan pada Tabel 3.1 bahwa terdapat 25 model yang parameternya tidak signifikan namun memiliki kasus yang sama dengan $w = 4000$ maka tetap digunakan.

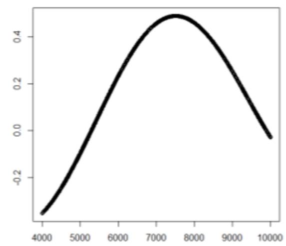
Selanjutnya, berdasarkan Tabel 3.1. diperoleh 1500 titik data yang nantinya ditentukan fungsi-fungsi parameter $b_0(w)$, $b_1(w)$, $b_2(w)$, dan $b_3(w)$ yang di-gambarkan pada Gambar 3.1.



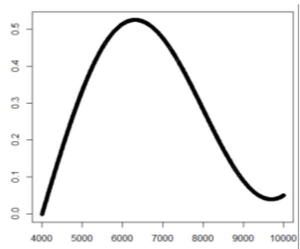
**Gambar 3.1..a.** Parameter $b_0(w)$ untuk setiap panjang gelombang *w*



**Gambar 3.1..b.** Parameter $b_1(w)$ untuk setiap panjang gelombang *w*



**Gambar 3.1..c.** Parameter $b_2(w)$ untuk setiap panjang gelombang *w*



**Gambar 3.1..d.** Parameter $b_3(w)$ untuk setiap panjang gelombang *w*

**Tabel 3.1. Ringkasan Model Linier Berganda Data Spektrun**

| $w$ | $R^2$ | $R^2\ adj.$ | $b_0$ | Sig. $b_0$ | $b_1$ | Sig. $b_1$ | $b_2$ | Sig. $b_2$ | $b_3$ | Sig. $b_3$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 4000 | 0,9774 | 0,9772 | 2,0226 | Y | 0,1497 | Y | -0,3502 | Y | -0,0016 | T |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 4016 | 0,9771 | 0,9769 | 2,0228 | Y | 0,1488 | Y | -0,3475 | Y | 0,0043 | T |
| 4020 | 0,9770 | 0,9768 | 2,0228 | Y | 0,1486 | Y | -0,3468 | Y | 0,0058 | Y |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 4792 | 0,9605 | 0,9601 | 1,8469 | Y | 0,0062 | Y | -0,1592 | Y | 0,2733 | Y |
| 4796 | 0,9605 | 0,9601 | 1,8451 | Y | 0,0051 | T | -0,1580 | Y | 0,2745 | Y |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 4832 | 0,9609 | 0,9606 | 1,8287 | Y | -0,0049 | T | -0,1470 | Y | 0,2856 | Y |
| 4836 | 0,9610 | 0,9606 | 1,8268 | Y | -0,0060 | Y | -0,1458 | Y | 0,2868 | Y |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 9868 | 0,9840 | 0,9839 | 0,9222 | Y | -0,4405 | Y | 0,0057 | Y | 0,0428 | Y |
| 9872 | 0,9840 | 0,9839 | 0,9253 | Y | -0,4409 | Y | 0,0046 | T | 0,0430 | Y |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 9908 | 0,9842 | 0,9841 | 0,9535 | Y | -0,4441 | Y | -0,0050 | T | 0,0446 | Y |
| 9912 | 0,9843 | 0,9841 | 0,9566 | Y | -0,4444 | Y | -0,0060 | Y | 0,0448 | Y |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 9992 | 0,9847 | 0,9845 | 1,0193 | Y | -0,4498 | Y | -0,0268 | Y | 0,0497 | Y |
| 9996 | 0,9847 | 0,9845 | 1,0225 | Y | -0,4500 | Y | -0,0278 | Y | 0,0499 | Y |

Keterangan:
Y : Nilai parameter signifikan.
T : Nilai Parameter tidak signifikan

Berdasarkan Tabel 3.1. dan Gambar 3.1. Estimasi fungsi parameter yang bergantung pada panjang gelombang dilakukan dengan mengambil beberapa titik data (tidak sebanyak 1500 titik). Titik yang diambil pada penelitian ini hanya 15 titik karena dengan lebih dari 15 titik, penghitungan pada program R tidak dapat membantu akibat parameter pada fungsi parameter akan semakin menuju 0 (nol) saat derajat polino-mial semakin tinggi. Dengan membagi 1500 titik menjadi 15 titik, maka $w = 4000, 4400, 4800, \ldots, 9600$). Berikut keluaran program R, dituliskan ulang dalam bentuk fungsi, adalah

$$\begin{aligned} b_0(z) = 2.022627 &+ 0.001577551z - 0.003091673z^2 - 1.150406\left(10^{-6}\right)z^3 + 1.590542\left(10^{-6}\right)z^4 \\ &+ 3.356677\left(10^{-9}\right)z^5 - 5.815351\left(10^{-10}\right)z^6 + 1.212603\left(10^{-11}\right)z^7 - 4.224554\left(10^{-13}\right)z^8 \\ &+ 1.319841\left(10^{-14}\right)z^9 - 2.702218\left(10^{-16}\right)z^{10} + 3.675723\left(10^{-18}\right)z^{11} \\ &- 3.26534\left(10^{-20}\right)z^{12} + 1.727665\left(10^{-22}\right)z^{13} - 4.146674\left(10^{-25}\right)z^{14} \end{aligned}$$

$$\begin{aligned} b_1(z) = 0.1497189 &- 0.005312329z - 0.001850678z^2 + 1.225606\left(10^{-6}\right)z^3 + 3.708668\left(10^{-6}\right)z^4 \\ &+ 1.148875\left(10^{-8}\right)z^5 - 4.198582\left(10^{-9}\right)z^6 + 6.592321\left(10^{-11}\right)z^7 - 1.487258\left(10^{-12}\right)z^8 \\ &+ 8.940348\left(10^{-14}\right)z^9 - 2.331202\left(10^{-15}\right)z^{10} + 2.796992\left(10^{-17}\right)z^{11} \\ &- 1.46334\left(10^{-19}\right)z^{12} + 9.539565\left(10^{-23}\right)z^{13} + 1.303204\left(10^{-24}\right)z^{14} \end{aligned}$$

$$\begin{aligned} b_2(z) = -0.350219 &+ 0.01663493z + 0.001093313z^2 - 1.475338\left(10^{-5}\right)z^3 - 5.254096\left(10^{-7}\right)z^4 \\ &+ 7.070654\left(10^{-10}\right)z^5 + 4.143365\left(10^{-10}\right)z^6 - 1.772412\left(10^{-11}\right)z^7 \\ &+ 6.777027\left(10^{-13}\right)z^8 - 1.994545\left(10^{-14}\right)z^9 + 4.153612\left(10^{-16}\right)z^{10} \\ &- 6.003462\left(10^{-18}\right)z^{11} + 5.750507\left(10^{-20}\right)z^{12} - 3.278508\left(10^{-22}\right)z^{13} \\ &+ 8.408282\left(10^{-25}\right)z^{14} \end{aligned}$$

$$\begin{aligned} b_3(z) = -0.001552071 &+ 0.03657963z - 1.678984\left(10^{-6}\right)z^2 - 3.036475\left(10^{-5}\right)z^3 + 1.434888\left(10^{-8}\right)z^4 \\ &+ 7.705993\left(10^{-9}\right)z^5 + 1.266465\left(10^{-10}\right)z^6 - 8.251182\left(10^{-12}\right)z^7 + 2.729091\left(10^{-13}\right)z^8 \\ &- 7.79575\left(10^{-15}\right)z^9 + 1.650688\left(10^{-16}\right)z^{10} - 2.414774\left(10^{-18}\right)z^{11} \\ &+ 2.331532\left(10^{-20}\right)z^{12} - 1.342437\left(10^{-22}\right)z^{13} + 3.490347\left(10^{-25}\right)z^{14} \end{aligned}$$

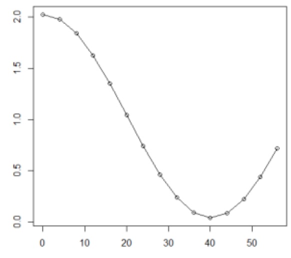
dengan

$$z = \frac{w - 4000}{100}$$

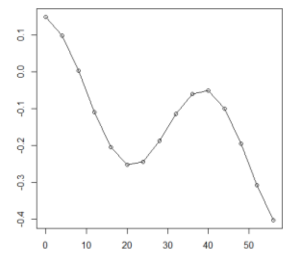
Untuk maka fungsi-fungsi di atas diberikan nomor persamaan (3.1) yaitu

$$b_0(z), b_1(z), b_2(z), b_3(z), \text{untuk } z = \frac{w - 4000}{100} \tag{3.1}$$

Gambar IV.2 adalah grafik fungsi-fungsi polinomial pada (3.1) yang dicocokkan dengan 15 nilai $w$ dan telah ditransformasi ke peubah $z$.

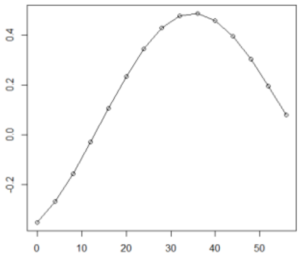



**Gambar IV.2.a.** fungsi polinomial $b_0(z)$ pada (4.1) yang dicocokkan

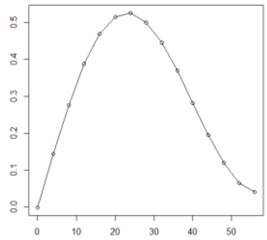



**Gambar IV.2.b.** fungsi polinomial $b_1(z)$ pada (4.1) yang dicocokkan dengan 15 nilai $z$

dengan 15 nilai $z$



**Gambar IV.2.c.** fungsi polinomial $b_2(z)$ pada (4.1) yang dicocokkan dengan 15 nilai $z$



**Gambar IV.2.d.** fungsi polinomial $b_3(z)$ pada (4.1) yang dicocokkan dengan 15 nilai $z$

Setelah fungsi-fungsi parameter model regresi linier berganda untuk setiap nilai $w$ ditentukan, maka dengan menggunakan persamaan (3.1) diperoleh model data spektrum melalui model regresi linier berganda termodifikasi adalah

$$\hat{f}(w,(x_1,x_2,x_3)) = b_0(z) + b_1(z)x_1 + b_2(z)x_2 + b_3(z)x_3 \qquad (3.2)$$

untuk $z = \dfrac{w-4000}{100}$.

Setelah didapatkan model data spektrum seperti pada (3.2), mengukur seberapa baik model tersebut untuk mewakili data yang dimiliki. Didasarkan pada galat atau *error* antara data spektrum yang dimiliki terhadap nilai yang dihasilkan model data spektrum pada (3.2) untuk triplet $(x_1,x_2,x_3)$ dan untuk suatu panjang gelombang $w$ yang bersesuaian adalah berdistribusi Normal dengan mean 0 (nol) dan variansi *error* tertentu. *Error* saat triplet $(x_1,x_2,x_3)$ dan $w$ dapat dihitung oleh

$$\varepsilon_{x_1,x_2,x_3,w} = y_{x_1,x_2,x_3,w} - \hat{f}(w,(x_1,x_2,x_3)) \qquad (3.3)$$

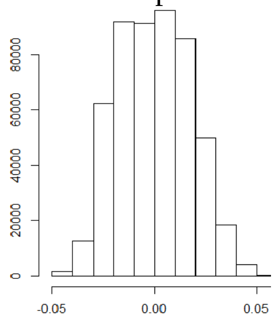
dengan $y_{x_1,x_2,x_3,w}$ adalah nilai pada data spektrum dan $\hat{f}(w,(x_1,x_2,x_3))$ adalah model data spektrum seperti pada (3.2) saat triplet $(x_1,x_2,x_3)$ dan $w$ yang berse-suaian. Gambar 3.3. merupakan histogram *error* model. Rata-rata *error* diperoleh sebesar $-1.5234(10^{-5})$ yang dekat dengan nilai 0 (nol). Dilakukan uji Ko-lmogorov-Smirnov pada *error* model berdistribusi Normal di panjang gelombang $w = 5972$. Hasil *running* program di atas adalah sebagai berikut

```
> v_e5972
[1] 0.0004491117
> ks.test(eij[,494],"pnorm",0,sd(eij[,494]))

        One-sample Kolmogorov-Smirnov test
data:  eij[, 494]
D = 0.030013, p-value = 0.9168
```

```
alternative hypothesis: two-sided
```

Tertampil $p_{value}$ dari hasil uji Kolmogorov-Smirnov adalah $0.9168 \geq 0.05$ yang berarti hipotesis $\mathbf{H_0}$: *error* mengikuti sebaran distribusi parametrik Normal de-ngan mean 0 (nol) dan variansinya $4.491117(10^{-4})$, diterima. Langkah ini dila-kukan untuk 1500 model. Dari hasil uji Kolmogorov-Smirnov untuk 1500 model diperoleh 990 model memiliki $p_{value} > 0.1$, 1078 model memiliki $p_{value} > 0.05$, 1235 model memiliki $p_{value} > 0.02$, 1458 model memiliki $p_{value} > 0.01$, 1500 model memiliki $p_{value} > 0.001$ , sehingga dapat dikatakan model ini cukup baik.




**Gambar 3.3** *Erorr* dari model data spektrum pada persamaan (3.2)

Setelah diperoleh model data spektrum, dapat diestimasi nilai data spektrum dari model di triplet konsentrasi $(x_1, x_2, x_3)$ seperti di nilai-nilai $x_i = 0.25, 0.3, 0.7, 0.9, 1$, dan lainnya, serta panjang gelombang tertentu seperti $w = 4001, 4002, 4003$, dan lainnya, yang mana tidak dimiliki data di nilai-nilai tersebut.

**Tabel IV.2** Nilai Model Data Spektrum di $(x_1, x_2, x_3)$ dan $w$ Tertentu

| $x_1$ | $x_2$ | $x_3$ | $w$ | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | 4000 | 4001 | 4002 | 4003 | 4004 | 4005 |
| 0 | 0,3 | 0,25 | 1,9171 | 1,9173 | 1,9175 | 1,9176 | 1,9178 | 1,9179 |
| 0,1 | 0,4 | 0,65 | 1,8965 | 1,8968 | 1,8971 | 1,8974 | 1,8977 | 1,8980 |
| 0,3 | 0,25 | 0,6 | 1,9790 | 1,9793 | 1,9795 | 1,9798 | 1,9800 | 1,9805 |
| 0,4 | 0,35 | 0,5 | 1,9591 | 1,9593 | 1,9596 | 1,9598 | 1,9600 | 1,9603 |
| 0,7 | 0,3 | 0,5 | 2,0215 | 2,0217 | 2,0220 | 2,0222 | 2,0224 | 2,0226 |
| 0,9 | 0,7 | 0,75 | 1,9110 | 1,9114 | 1,9117 | 1,9121 | 1,9124 | 1,9128 |
| 1 | 1 | 0,9 | 1,8207 | 1,8211 | 1,8216 | 1,8221 | 1,8225 | 1,8230 |

Dari hasil Tabel 3.2 menunjukan bahwa model data spektrum dapat digunakan untuk menentukan nilai data spektrum di konsentrasi-konsentrasi dan panjang gelombang tertentu, yang mana tidak dimiliki data di konsentrasi-konsentrasi dan panjang gelombang tersebut.

## 4. SIMPULAN

Pada penelitian ini menjelelaskan tentang penemuan model regresi linier berganda yang termodifikasi guna menemukan sebuah pedekaan pada suatu data spektrum, sehingga dapat ditarik beberapa kesimpulan sebagai berikut:

1. Dari penelitian ini diperoleh model regresi berganda termodifikasi, seperti pada (3.2) dengan memperhatikan parameter dan signifikansinya.
2. Interpolasi polinomial Newton digunakan dalam penentuan fungsi-fungsi polinomial yang dapat mewakili plot-plot parameter.
3. Berdasarkan 1500 model, didapatkan nilai error tiap modelnya berdistribusi Normal dengan mean 0 (nol) dan variansi $4.491117(10^{-4})$. Untuk signifikansi $p_{value} > 0.1$ diperoleh 990 model, $p_{value} > 0.05$ diperoleh 1078 model, $p_{value} > 0.02$ diperoleh 1235 model, 1458 model memiliki $p_{value} > 0.01$, 1500 model memiliki $p_{value} > 0.001$.
4. Dengan telah ditemukannya model regresi linier berganda, Penulis dapat mengestimasi data-data konsentrasi data spektrum $x_i = 0.25, 0.3, 0.7, 0.9, 1$ dan juga pada gelombang $w = 4001, 4002, 4003$ dengan kenaikan gelombang 1.

## 5. DAFTAR PUSTAKA


Prasetya, A. T., & Wibowo, N. A. (2016). Klasifikasi Mutu Teh Hitam (Camellia sinensis) Menggunakan Spektroskopi Difusi-Refleksi Inframerah Dekat dengan Analisis Komponen-komponen Utama. *Prosiding Seminar Nasional Quantum* .

Nugroho,D. B. (2009). *Metode Numerik.* Salatiga.

Pratama, F. W. (2012). Identifikasi Saham Perusahaan Penghitung Indeks LQ45 berdasarkan Koefisien Regresi Linier Berganda yang Signifikan beserta Peringkatnya Menggunakan P/E Rasio. *PROSIDING*.

Amrin. (2016). Data Mining dengan Regresi Linier Berganda untuk Peramalan Tingkat Inflasi. *Jurnal Techno Nusa Mandiri*.

Kutner, M., Nachtsheim., C., & Neter., J. (2004). Applied Linear Regression Models. *4th ed. New York: McGraw-Hill Companies, Inc***.**

Greselda, E. (2015, November 14). Model Biaya Garansi Satu Dimensi Polis FRW (NON-RENEWING FREE REPLACEMENT WARRANTY). *SEMINAR NASIONAL MATEMATIKA DAN PENDIDIKAN MATEMATIKA UNY* .

Setiawan, A. (2017). *Analisis Data Statistik.* Salatiga: Research Gate.