Konsep Dasar *Predictive* Data Mining

1

Tujuan utama dari penambangan prediktif adalah untuk memprediksi hasil di masa depan, bukan perilaku data saat ini. Ini melibatkan fungsi pembelajaran terawasi (*supervised learning*) yang digunakan untuk prediksi nilai target. Teknik yang termasuk dalam kategori penambangan ini adalah teknik estimasi atau prediksi dan teknik klasifikasi. Pemodelan data memerlukan analisis prediktif yang bekerja dengan memanfaatkan beberapa variabel untuk mengantisipasi nilai data masa depan yang tidak diketahui untuk variabel lain.

Analisis prediktif dijadikan sebagai alat yang menggunakan teknik statistik, pembelajaran mesin, dan penambangan data untuk menemukan fakta untuk dapat melakukan prediksi tentang peristiwa masa depan yang tidak diketahui. Model prediksi membuat perkiraan tentang nilai data yang tidak dikenal dengan menggunakan nilai yang diidentifikasi. Setiap kasus prediktif harus dianalisis kebutuhannya untuk mengetahui teknik yang paling cocok dalam mengekstraksi pengetahuan yang diinginkan. Analisis prediktif digunakan untuk memberikan informasi tentang "*apa yang mungkin terjadi*?" dan "*mengapa itu bisa terjadi*?". Banyak aplikasi data mining yang diperuntukkan untuk meramalkan keadaan data yang akan datang.

A. TEKNIK ESTIMASI ATAU PREDIKSI

Teknik estimasi merupakan salah satu teknik dalam memperoleh nilai estimasi dalam sebuah data. Proses ini didasarkan pada pola-pola di dalam sekumpulan data. Teknik estimasi menggunakan beberapa variabel untuk memprediksi nilai-nilai variabel masa mendatang yang diperlukan, yang belum diketahui saat ini. Teknik estimasi memiliki dua jenis variabel, yaitu variabel prediktor dan variabel target. Jenis variabel target adalah berupa bilangan numerik (kontinu) dan bukan kategorikal (nominal atau diskrit). Estimasi nilai dari variabel target ditentukan berdasarkan nilai dari variabel *predictor* (atribut).

Teknik prediksi/*forecasting* sama dengan teknik estimasi dimana label/target/*class* bertipe numerik, bedanya adalah data yang digunakan pada teknik prediksi merupakan data runtun waktu (data *time-series*). Istilah prediksi kadang digunakan juga untuk klasifikasi, tidak hanya untuk prediksi *time-series*, karena sifatnya yang bisa menghasilkan *class* berdasarkan berbagai atribut yang disediakan. Semua algoritma estimasi dapat digunakan untuk prediksi/*forecasting*.

REGRESI LINEAR

Apa itu Regresi Linier ?

- Regresi merupakan alat ukur yg digunakan untuk mengetahui ada tidaknya korelasi antarvariabel.
- Analisis regresi lebih akurat dlm analisis korelasi karena tingkat perubahan suatu variabel terhdp variabel lainnya dpt ditentukan). Jadi pada regresi, peramalan atau perkiraan nilai variabel terikat pada nilai variabel bebas lebih akurat pula.
- Regresi linier adalah regresi yang variabel bebasnya (variabel X) berpangkat paling tinggi satu. Utk regresi sederhana, yaitu regresi linier yg hanya melibatkan dua variabel (variabel X dan Y).

Persamaan Regresi Linear dari Y terhadap X

$$Y = a + bX$$

Keterangan :

Y = variabel terikat

X = variabel bebas

a = intersep / konstanta

b = koefisien regresi / slop

Persamaan regresi linear di atas dpt pula dituliskan dlm bentuk

$$Y = \left(\frac{\Sigma xy}{\Sigma x^2} \right) x$$

Mencari nilai a dan b

- Rumus 1

$$a = \frac{(\Sigma Y)(\Sigma X^2) - (\Sigma X)(\Sigma XY)}{(n)(\Sigma X^2) - (\Sigma X)^2}$$

$$b = \frac{(n)(\Sigma XY) - (\Sigma X)(\Sigma Y)}{(n)(\Sigma X^2) - (\Sigma X)^2}$$

- Pendekatan Matriks

$$\begin{pmatrix} n & \Sigma X \\ \Sigma X & \Sigma X^2 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \Sigma Y \\ \Sigma XY \end{pmatrix}$$

$$a = \frac{\det A_1}{\det A} \qquad b = \frac{\det A_2}{\det A}$$

$$A = \begin{pmatrix} n & \Sigma X \\ \Sigma X & \Sigma X^2 \end{pmatrix} \; A_1 = \begin{pmatrix} \Sigma Y & \Sigma X \\ \Sigma XY & \Sigma X^2 \end{pmatrix} \; A_2 = \begin{pmatrix} \Sigma n & \Sigma Y \\ \Sigma X & \Sigma XY \end{pmatrix}$$

$$\det A = (n)(\Sigma X^2) - (\Sigma X)(\Sigma X)$$

$$\det A_1 = (\Sigma Y)(\Sigma X^2) - (\Sigma X)(\Sigma XY)$$

$$\det A_2 = (n)(\Sigma XY) - (\Sigma Y)(\Sigma X)$$

- Rumus II

$$b = \frac{(n)(\Sigma XY) - (\Sigma X)(\Sigma Y)}{(n)(\Sigma X^2) - (\Sigma X)^2}$$

$$a = \overline{Y} - b.\overline{X}$$

Contoh Soal

- Berikut ini data mengenai pengalaman kerja dan penjualan
- X=pengalaman kerja (tahun)
- Y=omzet penjualan (ribuan)

X	2	3	2	5	6	1	4	1
Y	5	8	8	7	11	3	10	4

- Tentukan nilai a dan b (gunakan ketiga cara)!
- Buatkan persamaan regresinya!
- Berapa omzet pengjualan dari seorang karyawan yg pengalaman kerjanya 3,5 tahun

Penyelesaian :

X	Y	X2	Y2	XY
2	5	4	25	10
3	8	9	64	24
2	8	4	64	16
5	7	25	49	35
6	11	36	121	66
1	3	1	9	3
4	10	16	100	40
1	4	1	16	4
24	56	96	448	198

$$\overline{X} = \frac{24}{8} = 3 \qquad \overline{Y} = \frac{56}{8} = 7$$

Cara 1.

$$a = \frac{(56)(96) - (24)(198)}{(8)(96) - (24)^2}$$

$$a = \frac{5.376 - 4.752}{768 - 576} = 3,25$$

$$b = \frac{(8)(198) - (24)(56)}{(8)(96) - (24)^2}$$

$$b = \frac{1.584 - 1.344}{768 - 576} = 1,25$$

Cara 2.

$$\begin{pmatrix} 8 & 24 \\ 24 & 96 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 56 \\ 198 \end{pmatrix}$$

$$A = \begin{pmatrix} 8 & 24 \\ 24 & 96 \end{pmatrix} \quad A_1 = \begin{pmatrix} 56 & 24 \\ 198 & 96 \end{pmatrix} \quad A_2 = \begin{pmatrix} 8 & 56 \\ 24 & 198 \end{pmatrix}$$

$$\det A = (8)(96) - (24 - 24) = 192$$

$$\det A_1 = (56)(96) - (24)(198) = 624$$

$$\det A_2 = (8)(198) - (56)(24) = 240$$

$$a = \frac{624}{192} = 3,25 \qquad b = \frac{240}{192} = 1,25$$

Cara 3

$$b = \frac{(8)(198) - (24)(56)}{(8)(96) - (24)^2}$$

$$b = \frac{1.548 - 1.344}{768 - 576} = 1,25$$

$$a = 7 - 1,25(3)$$

$$a = 3,25$$

a. Dari ketiga cara pengerjaan tersebut diperoleh nilai a = 3,25 dan nilai b = 1,25
b. Persamaan regresi linearnya adalah Y=3,25+1,25X
c. Nilai duga Y, jika X=3,5 adalah Y=3,25+1,25X

   Y=3,25+1,25(3,5)

   =7,625

Koefisien Determinasi (R^2)

$$R^2 = \frac{((n)(\Sigma XY) - (\Sigma X)(\Sigma Y))^2}{(n(\Sigma X^2) - (\Sigma X)^2\ (n(\Sigma Y^2) - (\Sigma Y)^2)}$$

$$R^2 = \frac{((8)(198) - (24)(56))^2}{(8(96) - (24)^2\ \ (8(448) - (56)^2)}$$

$$R^2 = \frac{(1.584 - 1.344)^2}{(768 - 576)(3.584 - 3.136)}$$

$$R^2 = \frac{(240)^2}{(192)(448)} \quad = \frac{57.600}{86.016} = 0,6696$$

Nilai determinasi (R2) sebesar 0,6696, artinya sumbangan atau pengaruh pegalaman Kerja terhadap naik turunnya omzet penjualan adalah sebesar 66,96%. Sisanya 33,04% Disebabkan oleh faktor lain yang tidak dimasukkan dalam model.

SELISIH TAKSIR STANDAR (STANDAR DEVIASI)

- Angka indeks yg digunakan utk mengukur ketepatan suatu penduga atau mengukur jumlah variasi titik-titik observasi di sekitar garis regresi.
- Jika semua titik observasi berada tepat pada garis regresi, selisih taksir standar sama dengan nol. Menunjukkan pencaran data.
- Selisih taksir standar berguna mengetahui batasan seberapa jauh melesetnya perkiraan dalam meramal data.

Rumus

$$S_{y/x} = S_{x.y} = Se = \sqrt{\frac{\Sigma(Y - Y')^2}{n-2}}$$

$$atau$$

$$S_{x/y} = S_{y.x} = Se = \sqrt{\frac{\Sigma(X - Y')^2}{n-2}}$$

Keterangan :

Sy/x = Sx/y = Selisih taksir standar

Y = X = nilai variabel sebenarnya

Y' = X' = nilai variabel yang diperkirakan

n = jumlah frekuensi

Contoh :

- Hubungan antara variabel X dan variabel Y

X	1	2	3	4	5	6
Y	6	4	3	5	4	2

a. Buatkan persamaan regresinya
b. Tentukan nilai duga Y, jika X = 8
c. Tentukan selisih taksir standarnya

Penyelesaian

X	Y	X^2	Y^2	XY
1	6	1	36	6
2	4	4	16	8
3	3	9	9	9
4	5	16	25	20
5	4	25	16	20
6	2	36	4	12
21	24	91	106	75

$$b = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)^2}$$

$$b = \frac{6(75) - (21)(24)}{6(91) - (21)^2}$$

$$b = \frac{-54}{105} = -0{,}5$$

$$a = \overline{Y} - b.\overline{X}$$

$$a = \frac{24}{6} - (0{,}5)\left(\frac{21}{6}\right)$$

a. Persamaan garis regresinya:

Y’ = 5,75 – 0,5 X

b. Nilai duga Y’, jika X=8

Y’ = 5,75 – 0,5 (8)

Y’ = 1,75

c. Selisih taksir standar

X	Y	Y'	Y-Y'	$(Y-Y')^2$
1	6	5.25	0.75	0.5625
2	4	4.75	-0.8	0.5625
3	3	4.25	-1.3	1.5625
4	5	3.75	1.25	1.5625
5	4	3.25	0.75	0.5625
6	2	2.75	-0.8	0.5625
				5.375

$$S_{y/x} = \sqrt{\frac{\sum (Y - Y')^2}{n-2}}$$

$$S_{y/x} = \sqrt{\frac{5,375}{6-2}} = 1,2$$

Tabel 4.1
Cacat Produk Makanan Akibat Suhu

No	Rata-rata suhu	Jumlah cacat
1	27	13
2	28	16
3	25	12
4	26	14
5	24	12
6	27	16
7	23	9
8	24	13
9	23	11
10	22	7
11	21	5
12	26	12
13	25	11
14	26	13
15	27	14
16	24	11
17	22	5
18	21	6
19	20	3
20	22	6
21	19	4
22	20	5
23	23	9
24	24	10
25	25	13
26	21	7
27	20	4
28	20	6
29	19	3
30	25	12

DATA MINING
REGRESI LINIER BERGANDA

DENTIK KARYANINGSIH. M.KOM

Regresi Linier Berganda

- Regresi linier berganda yaitu suatu model persamaan yang mengkaji antara sebuah variabel terikat/response dengan dua atau lebih variabel bebas/predictor.
- Menurut (sugiyono,2014) “Analisis linier berganda bermaksud meramalkan bagaimana keadaan (naik turunnya)variabel dependen (kriterium), bila dua atau lebih variabel independent sebagai factor prediator dimanipulasi (dinaik turunkan nilainya). Jadi analisis regresi berganda akan dilakukan bila jumlah variable independennya minimal 2.
- Regresi linier memiliki syarat atau asumsi klasik yang harus terpenuhi. Syarat tersebut ditujukan agar model prediksi yang dihasilkan bersifat BLUE (Best Linier Unbiased Estimation).

A. REGRESI LINIER BERGANDA

Tujuan dari regresi linier berganda adalah untuk mengetahui nilai variabel terikat/*response* (Y) jika variabel bebas/*predictor* ($X_1, X_2, ..., X_n$) diketahui. Regresi linier berganda juga dilakukan untuk mengetahui masing-masing variabel apakah memiliki hubungan positif atau negatif. Untuk itu maka berikut diberikan persamaan matematis untuk regresi linier berganda:

$$\boldsymbol{Y = a + b_1X_1 + b_2X_2 + \cdots + b_nX_n}$$

Keterangan:

Y	= Variabel terikat (nilai yang diprediksikan)
X_1, X_2, X_n	= Variabel bebas
a	= Konstanta (nilai Y' apabila semua X bernilai 0)
b	= koefisien regresi (nilai peningkatan ataupun penurunan)

Apabila terdapat dua variabel bebas/*predictor* yaitu X_1 dan X_2, mka bentuk persamaan regresinya adalah

$$Y = a + b_1X_1 + b_2X_2$$

Berikut adalah kondisi koefisien regresi yang mungkin terjadi, b_1 dan b_2 memiliki suatu nilai:

1. Nilai 0, hal tersebut terjadi apabila Y tidak dipengaruhi variabel X_1 dan X_2.
2. Bernilai negatif (-), hal tersebut disebabkan hubungan terbalik antara variabel Y dengan variabel X_1 dan X_2.
3. Bernilai positif (+), hal tersebut disebabkan hubungan searah antara variabel Y dengan variabel X_1 dan X_2

Konstanta dan koefisien-koefisien regresi yaitu a, b_1 dan b_2 dihitung dengan persamaan berikut:

$$b_1 = \frac{[(\sum x_2^2 \times \sum x_1 y) - (\sum x_2 y \times \sum x_1 x_2)]}{[(\sum x_1^2 \times \sum x_2^2) - (\sum x_1 x_2)^2]}$$

$$b_2 = \frac{[(\sum x_1^2 \times \sum x_2 y) - (\sum x_1 y \times \sum x_1 x_2)]}{[(\sum x_1^2 \times \sum x_2^2) - (\sum x_1 x_2)^2]}$$

$$a = \frac{(\sum Y) - (b_1 \times \sum X_1) - (b_2 \times \sum X_2)}{n}$$

Nilai masing-masing dari x_1, x_2, dan y dapat dilihat berdasarkan persamaan matematika berikut:

$$\sum x_1^2 = \sum X_1^2 - \frac{(\sum X_1)^2}{n}$$

$$\sum x_2^2 = \sum X_2^2 - \frac{(\sum X_2)^2}{n}$$

$$\sum y^2 = \sum y^2 - \frac{(\sum Y)^2}{n}$$

$$\sum x_1 y = \sum X_1 Y - \frac{\sum X_1 \sum Y}{n}$$

$$\sum x_2 y = \sum X_2 Y - \frac{\sum X_2 \sum Y}{n}$$

$$\sum x_1 x_2 = \sum X_1 X_2 - \frac{\sum X_1 \sum X_2}{n}$$

Selain cara tersebut terdapat cara lain yang digunakan untuk menentukan nilai a, b_1 dan b_2. Adapun caranya adalah dengan menggunakan perhitungan matriks (metode kuadrat terkecil).

$$an + b_1\sum X_1 + b_2\sum X_2 = \sum Y$$

$$a\sum X_1 + b_1\sum X_1^2 + b_2\sum X_1 X_2 = \sum X_1 Y$$

$$a\sum X_2 + b_1\sum X_1 X_2 + b_2\sum X_2^2 = \sum X_2 Y$$

Berdasarkan tiga persamaan dengan tiga buah variabel maka akan dibentuk matriks. Matriks tersebut kemudian dicari nilai determinannya sehingga dapat ditentukan nilai a, b_1 dan b_2. Untuk pembentukan matriks dari tiga persamaan tersebut dapat dilihat sebagai berikut:

$$m_{11}a + m_{12}b_1 + m_{13}b_2 = h_1$$

$$m_{21}a + m_{22}b_1 + m_{23}b_2 = h_2$$

$$m_{31}a + m_{32}b_1 + m_{33}b_2 = h_3$$

$$\begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{bmatrix} a \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix}$$

$$M_1 = \begin{bmatrix} h_1 & m_{12} & m_{13} \\ h_2 & m_{22} & m_{23} \\ h_3 & m_{32} & m_{33} \end{bmatrix} \qquad a = \frac{\det M_1}{\det M}$$

$$M_2 = \begin{bmatrix} m_{11} & h_1 & m_{13} \\ m_{21} & h_2 & m_{23} \\ m_{31} & h_3 & m_{33} \end{bmatrix} \qquad b_1 = \frac{\det M_2}{\det M}$$

$$M_3 = \begin{bmatrix} m_{11} & m_{12} & h_1 \\ m_{21} & m_{22} & h_2 \\ m_{31} & m_{32} & h_3 \end{bmatrix} \qquad b_2 = \frac{\det M_3}{\det M}$$

Contoh:

$$\begin{cases} 2a + b_1 + 4b_2 = 8 \\ 3a + 2b_1 + b_2 = 5 \\ a + 3b_1 + 3b_2 = 8 \end{cases}$$

$$M = \begin{bmatrix} 2 & 1 & 4 \\ 3 & 2 & 1 \\ 1 & 3 & 3 \end{bmatrix}$$

$$M_1 = \begin{bmatrix} 8 & 1 & 4 \\ 5 & 2 & 1 \\ 8 & 3 & 3 \end{bmatrix}$$

$$M_2 = \begin{bmatrix} 2 & 8 & 4 \\ 3 & 5 & 1 \\ 1 & 8 & 3 \end{bmatrix}$$

$$M_3 = \begin{bmatrix} 2 & 1 & 8 \\ 3 & 2 & 5 \\ 1 & 3 & 8 \end{bmatrix}$$

$$a = \frac{\det M_1}{\det M} = \frac{13}{26} = 0{,}5$$

$$b_1 = \frac{\det M_2}{\det M} = \frac{26}{26} = 1 b_2 = \frac{\det M_3}{\det M} = \frac{39}{26} = 1{,}5$$

B. KOEFISIEN DETERMINISI (r^2)

Koefisien determinisi adalah koefisien yang digunakan untuk mengetahui tingkat pengaruh variabel-variabel X_1 dan X_2 terhadap variabel Y. Untuk itu koefisien determinisi dapat dihitung dengan persamaan sebagai berikut:

$$r^2 = \frac{(b_1 \sum x_1 y) + (b_2 \sum x_2 y)}{\sum y^2}$$

Beberapa ketentuan untuk hasil koefisien determinasi dapat digunakan untuk menjelaskan keterhubungan antara variabel bebas/*predictor* Y dengan variabel terikat/*response* X_1 dan X_2. Apabila koefisien determinasi (r^2) bernilai 0 , maka dalam persamaan regresi menunjukkan bahwa variabel Y tidak dapat dijelaskan oleh variabel X_1 dan X_2. Hal tersebut juga berlaku sebaliknya, apabila koefisien determinasi (r^2) bernilai 1, maka variabel Y secara penuh dapat dijelaskan oleh variabel X_1 dan X_2.

C. KOEFISIEN KORELASI GANDA (r)

Koefisien korelasi ganda merupakan koefisien yang dapat digunakan untuk menjelaskan korelasi yang terjadi antara variabel X_1, X_2,..., X_n dengan variabel Y. Koefisien korelasi ganda merupakan akar dari koefisien determinasi. Sehingga untuk membuat persamaan koefisien korelasi tersebut maka persamaan pada koefisien determinasi tinggal diakarkan saja. Berikut persamaan dari koefisien korelasi ganda:

$$r = \sqrt{r^2} = \sqrt{\frac{(b_1 \sum x_1 y) + (b_2 \sum x_2 y)}{\sum y^2}}$$

Terdapat aturan untuk menjelaskan nilai r tersebut. Apabila nilai r mendekati nilai +1 atau – 1, maka menunjukkan semakin kuatnya korelasi yang antara variabel bebas dengan variabel terikat. Sebaliknya, apabila nilai r mendekati 0, maka semakin lemahnya korelasi yang terjadi antara variabel bebas dengan variabel terikat.

KOEFISIEN PARSIAL

Korelasi parsial adalah korelasi yang digunakan untuk menjelaskan hubungan salah satu variabel dengan tepat satu variabel lainnya. Untuk variabel lain gap sebagai nilai konstan atau konstanta. Untuk regresi linier dengan dua variabel, at tiga jenis koefisien variabel parsial. Adapun tiga jenis variabel tersebut dapat sebagai berikut:

Korelasi X_1 dan X_2, (Y dianggap sebagai konstan)

$$r_{12.y} = \frac{r_{12} - (r_{y1}r_{y2})}{\sqrt{(1 - r_{y1}^2)(1 - r_{y2}^2)}}$$

Korelasi Y dan X_1, (X_2 dianggap sebagai konstan)

$$r_{y1.2} = \frac{r_{y1} - (r_{y2}r_{12})}{\sqrt{(1 - r_{y2}^2)(1 - r_{12}^2)}}$$

Korelasi Y dan X_1, (X_2 dianggap sebagai konstan)

$$r_{y2.1} = \frac{r_{y2} - (r_{y1}r_{12})}{\sqrt{(1 - r_{y1}^2)(1 - r_{12}^2)}}$$

Berdasarkan tiga korelasi parsial tersebut, perlu diketahui untuk masing-masing r_{12}, r_{y1}, dan r_{y2}. Persamaannya dapat dilihat sebagai berikut:

$$r_{12} = \frac{n \times \sum X_1 X_2 - (\sum X_1 \times \sum X_2)}{\sqrt{[(n \times \sum X_1^2) - (\sum X_1^2)] \times [(n \times \sum X_2^3) - (\sum X_2)^2]}}$$

$$r_{y1} = \frac{n \times \sum X_1 Y - (\sum Y \times \sum X_1)}{\sqrt{[(n \times \sum Y^2) - (\sum Y^2)] \times [(n \times \sum X_1^3) - (\sum X_1)^2]}}$$

$$r_{y2} = \frac{n \times \sum X_2 Y - (\sum Y \times \sum X_2)}{\sqrt{[(n \times \sum Y^2) - (\sum Y^2)] \times [(n \times \sum X_2^3) - (\sum X_2)^2]}}$$

E. KESALAHAN BAKU ESTIMASI (*Standard Error Estimate*)

Kesalahan baku estimasi merupakan standard yang digunakan untuk mengukur tepat atau kurang tepatnya persamaan regresi untuk melakukan prediksi terhadap variabel terikat/ *response* Y. Jika kesalahan bakunya besar maka persamaan regresi yang digunakan untuk melakukan prediksi kurang tepat. Hal tersebut terjadi akibat besarnya selisih antara variabel Y estimasi dengan variabel Y. Persamaan yang digunakan untuk mengukur keslahan baku dapat ditunjukkan dengan:

$$S_e(S_{yx})\sqrt{\frac{\sum Y^2 - (a\sum Y) - \left(b_1 \sum X_1 Y\right) - (b_2 \sum X_2 Y)}{N-3}}$$

F. PENGUJIAN HIPOTESIS

Kegiatan belajar sebelumnya telah memberikan deskripsi mengenai pengujian hipotesis. Sama seperti materi regresi linier sederhana, pada materi regresi linier berganda juga dilakukan pengujian hipotesis. Pengujian ini tetap memperhatikan populasi. Untuk mempermudah dapat dilakukan dengan menggunakan random sampling. pengujian yang dilakukan juga menggunakan koefisien determinasi (r^2). Setelah mendapatkan nilai r^2 maka selanjutnya uji signifikansi dapat dilakukan. pengujian dapat dilakukan dengan menggunakan Uji-*F*.

Uji-*F*

Uji-*F* merupakan salah satu jenis pengujian signifikansi. Pengujian ini bertujuan untuk mengetahui tingkat keberpengaruhan variabel bebas/predictor (X_1 dan X_2) terhadap variabel terikat/response Y. Langkah-langkah yang dilakukan dalam Uji-*F* dapat dilihat sebagai berikut:

a. Penentun hipotesis:

$H_0 : \beta = 0$; variabel bebas (X_1 dan X_2) tidak berpengaruh signifikan atau nyata terhadap variabel terikat (Y)

$H_1 : \beta \neq 0$; variabel bebas (X_1 dan X_2) berpengaruh signifikan atau nyata terhadap variabel terikat (Y)

b. Penentuan signifikansi:

Taraf signifikansi umumnya memiliki nilai 0,05 atau dalam bentuk matematisnya adalah $\alpha = 5\%$ ($\alpha = 0{,}05$).

c. Menghitung nilai F hitung

Rumus yang digunakan untuk menghitung F hitung ditunjukkan pada persamaan sebagai berikut:

$$F_{hit} = \frac{r^2/k}{(1 - r^2)/(n - k - 1)} = \frac{r^2(n - k - 1)}{k(1 - r^2)}$$

d. Menentukan F_{tab} dengan menggunakan tabel uji-t.

Berikut akan diberikan tabel uji-t dengan nilai signifikan $\alpha = 5\%$. Nilai derajat kebebasan pembilang dihitung dengan persamaan $(numerator, df) = k - 1$; Untuk penyebut $(denominator, df) = n - k$;

Keterangan:

$df = derajat\ kebebasan$

$n = jumlah\ sampel$

$k = jumlah\ variabel\ (variabel\ predictor, variabel\ response)$

e. Pengujian nilai t hitung dan t tabel adalah sebagai berikut:

Bila nilai $F_{hit} < F_{tab}$, maka H_0 diterima, H_1 ditolak

Bila nilai $F_{hit} > F_{tab}$, maka H_0 ditolak, H_1 diterima

f. Penentuan kesimpulan uji signifikansi

Uji Koefisien Parsial (Uji-*t*)

Pengujian koefisien parsial merupakan pengujian yang berfungsi untuk mengetahui variabel bebas/*predictor* secara parsial berpengaruh terhadap variabel terikat/terikat. Pengujian koefisien secara parsial ini sama dengan Uji-*t*. Maka pengujian ini memiliki tahapan yang sama dengan Uji-*t* yang sudah dijabarkan pada kegiatan belajar sebelumnya yaitu pada materi regresi linier sederhana.

Contoh 1:

Sebuah penelitian dilakukan untuk mengetahui apakah pemberian mpasi (X_1) dan tinggi balita (X_2) mempengaruhi berat badan balita (Y) di Bali. Data pengamatan pada sepuluh balita diberikan sebagai berikut:

Tabel 4.11
Data Berat Badan Balita

Balita ke-	X_1	X_2	Y
1	8	97	15
2	10	111	18
3	7	95	14
4	12	109	20
5	9	105	16
6	10	108	19
7	7	98	12
8	8	100	16
9	11	109	19
10	9	105	17

1. Untuk menentukan persamaan regresi linier maka harus dicari nilai konstanta a, nilai b_1 dan b_2.

Tabel 4.12
Perhitungan X_1^2, X_2^2, X_1X_2, X_1Y, X_2Y

Balita ke-	X_1	X_2	Y	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y
1	8	97	15	64	9409	776	120	1455
2	10	111	18	100	12321	1110	180	1998
3	7	95	14	49	9025	665	98	1330
4	12	109	20	144	11881	1308	240	2180
5	9	105	16	81	11025	945	144	1680
6	10	108	19	100	11664	1080	190	2052
7	7	98	12	49	9604	686	84	1176
8	8	100	16	64	10000	800	128	1600
9	11	109	19	121	11881	1199	209	2071
10	9	105	17	81	11025	945	153	1785
jumlah	91	1037	166	853	107835	9514	1546	17327

Sebelum menentukan nilai a, b_1, dan b_2 terlebih dahulu dicari niai $\sum x_1^2$, $\sum x_2^2$, $\sum y^2$, $\sum x_1 y$, $\sum x_2 y$, dan $\sum x_1 x_2$ sebagai berikut:

a. Menentukan nilai $\sum \boldsymbol{x_1^2}$

$$\sum x_1^2 = \sum X_1^2 - \frac{(\sum X_1)^2}{n}$$

$$\sum x_1^2 = 853 - \frac{8281}{10}$$

$$\sum x_1^2 = 24{,}9$$

b. Menentukan nilai $\sum \boldsymbol{x_2^2}$

$$\sum x_2^2 = \sum X_2^2 - \frac{(\sum X_2)^2}{n}$$

$$\sum x_2^2 = 107835 - \frac{1075369}{10}$$

$$\sum x_2^2 = 298.1$$

c. Menentukan nilai $\sum \boldsymbol{y^2}$

$$\sum y^2 = \sum Y^2 - \frac{(\sum Y)^2}{n}$$

$$\sum y^2 = 2812 - \frac{27556}{10}$$

$$\sum y^2 = 56.4$$

d. Menentukan nilai $\sum \boldsymbol{x_1 y}$

$$\sum x_1 y = \sum X_1 Y - \frac{\sum X_1 \sum Y}{n}$$

$$\sum x_1 y = 1546 - \frac{91 \times 166}{10}$$

$$\sum x_1 y = 1546 - \frac{15106}{10}$$

$$\sum x_1 y = 35{,}4$$

e. Menentukan nilai $\sum \boldsymbol{x_2 y}$

$$\sum x_2 y = \sum X_2 Y - \frac{\sum X_2 \sum Y}{n}$$

$$\sum x_2 y = 17327 - \frac{1037 \times 166}{10}$$

$$\sum x_2 y = 17327 - \frac{172142}{10}$$

$$\sum x_2 y = 112{,}79$$

f. Menentukan nilai $\sum \boldsymbol{x_1 x_2}$

$$\sum x_1 x_2 = \sum X_1 X_2 - \frac{\sum X_1 \sum X_2}{n}$$

$$\sum x_1 x_2 = 9514 - \frac{91 \times 1037}{10}$$

$$\sum x_1 x_2 = 9514 - \frac{94367}{10}$$

$$\sum x_1 x_2 = 77{,}3$$

Nilai yang telah didapatkan tersebut digunakan untuk menentukan nilai dari a, b_1, dan b_2 yang dapat dilihat pada penyelesaian matematika berikut:

a. Nilai $\boldsymbol{b_1}$

$$b_1 = \frac{[(\sum x_2^2 \times \sum x_1 y) - (\sum x_2 y \times \sum x_1 x_2)]}{[(\sum x_1^2 \times \sum x_2^2) - (\sum x_1 x_2)^2]}$$

$$b_1 = \frac{[(298{,}1 \times 35.4) - (112.79 \times 77.29)]}{[(24.9 \times 298.1) - (77.29)^2]}$$

$$b_1 = 1{,}26$$

b. Nilai $\boldsymbol{b_2}$

$$b_2 = \frac{[(\sum x_1^2 \times \sum x_2 y) - (\sum x_1 y \times \sum x_1 x_2)]}{[(\sum x_1^2 \times \sum x_2^2) - (\sum x_1 x_2)^2]}$$

$$b_2 = \frac{[(24.9 \times 112.79) - (35{,}4 \times 77.29)]}{[(24.9 \times 298.1) - (77.29)^2]}$$

$$b_2 = 0{,}049$$

c. Nilai $\boldsymbol{a}$

$$a = \frac{(\sum Y) - (b_1 \times \sum X_1) - (b_2 \times \sum X_2)}{n}$$

$$a = \frac{(166) - (1{,}26 \times 91) - (0{,}049 \times 1037)}{10}$$

$$a = -1{,}06$$

Setelah nilai a, b_1, dan b_2 didapatkan maka dapat dibentuk persamaan regresi linier berganda, yaitu:

$$\boldsymbol{Y = -1,06 + 1,26X_1 + 0,049X_2}$$

2. Berikut adalah cara untuk menentukan nilai koefisien determinasi (r^2)

$$r^2 = \frac{(b_1 \sum x_1 y) + (b_2 \sum x_2 y)}{\sum y^2}$$

$$r^2 = \frac{(1{,}26 * 35{,}4) + (0{,}049 * 112{,}8)}{56{,}4}$$

$$r^2 = 0{,}89$$

$$r^2 = 89\ \%$$

Nilai berat badan yang merupakan variabel terikat/*response* (Y) dapat dijelaskan oleh variabel bebas/*predictor* mpasi (X_1) dan tinggi badan (X_2) pada persamaan $\boldsymbol{Y = -1,06 + 1,26X_1 + 0,49X_2}$ dengan persentase 89%. Untuk persentase 11% dijelaskan oleh faktor lain diluar persamaan tersebut.

Contoh 2:

Dalam penjualan sepeda motor *matic second*, sorang penjual mengalami kebingungan untuk memprediksi harga jual sepeda motor tersebut. Variabel-variabel yang mempengaruhi penjualan diantaranya adalah jarak tempuh (KM) dan lama pemakaian (Tahun). Untuk hal tersebut dilakukan penelitian terhadap 10 data penjualan sebagai berikut.

Tabel 4.13
Data Penjuaan Sepeda Motor

Motor ke-	X_1	X_2	Y
1	39	2	11
2	33	7	8
3	18	10	8
4	62	6	6
5	62	8	5
6	54	10	3
7	16	4	15
8	42	7	7
9	41	5	8
10	25	9	9
jumlah	392	68	80

Penyelesaian:

1. Untuk menentukan persamaan regresi linier maka harus dicari nilai konstanta a, nilai b_1 dan b_2.

Tabel 4.14
Perhitungan X_1^2, X_2^2, X_1X_2, X_1Y, X_2Y

Motor ke-	X_1	X_2	Y	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y	Y2
1	39	2	11	1521	4	78	429	22	121
2	33	7	8	1089	49	231	264	56	64
3	18	10	8	324	100	180	144	80	64
4	62	6	6	3844	36	372	372	36	36
5	62	8	5	3844	64	496	310	40	25
6	54	10	3	2916	100	540	162	30	9
7	16	4	15	256	16	64	240	60	225
8	42	7	7	1764	49	294	294	49	49
9	41	5	8	1681	25	205	328	40	64
10	25	9	9	625	81	225	225	81	81
Jumlah	392	68	80	17864	524	2685	2768	494	738