

## BAB IV IMPLEMENTASI

### 4.1 Implementasi Sistem

Aplikasi yang dibangun merupakan aplikasi *stand alone* yang dapat diinstalasi ke sebuah PC atau *workstation* berbasis sistem operasi *Microsoft Windows Server 2003* atau *Microsoft Windows XP*. File paket *startup* aplikasi data mining ini berukuran 5 Mb sehingga untuk pendistribusiannya menggunakan CD, dimana tingkat kegagalan yang rendah dalam proses instalasi jika dibandingkan dengan menggunakan disket.

Sebelum melakukan instalasi pastikan *Microsoft SQL Server 2000* sudah tersedia pada PC, hal ini dikarenakan aplikasi data mining ini membutuhkan asupan data dalam bentuk tabel dalam melakukan proses analisis regresi. Apabila *Microsoft SQL Server 2000* sudah tersedia maka proses instalasi aplikasi data mining bisa dilakukan. Proses instalasi memakan waktu kurang lebih sekitar satu menit untuk PC menggunakan sistem operasi *Windows Server 2003* dengan spesifikasi *Pentium M740* dengan memori 256 Mb. Apabila proses instalasi berjalan lancar maka program akan membuat *icon* aplikasi data mining di *start menu*, dan aplikasi data mining siap untuk digunakan.

### 4.2 Implementasi Basis Data

Pada bagian ini, akan dijelaskan mengenai implementasi dari basis data yang ada pada aplikasi data mining. Aplikasi ini menggunakan sebuah database dengan nama "db\_analisis" dengan tiga buah tabel yaitu :

- a. Tabel distribusi t, digunakan untuk memberikan nilai  $t_{tabel}$  dalam proses perhitungan analisis regresi.

**Tabel 4.1** Tabel Distribusi t

Nama tabel pada perancangan : Tabel distribusi t				
Nama tabel pada file : tbl_distribusi_t				
No.	Nama Field	Tipe	Ukuran	Keterangan
1.	df	Int	4	Nilai dari n-2
2.	Alpha	Decimal	9	Nilai t tabel

- b. Tabel distribusi F, digunakan untuk memberikan nilai  $F_{\text{tabel}}$  dalam proses perhitungan analisis regresi.

**Tabel 4.2** Tabel distribusi F

Nama tabel pada perancangan : Tabel distribusi F				
Nama tabel pada file : tbl_distribusi_f				
No.	Nama Field	Tipe	Ukuran	Keterangan
1.	df	Int	4	Nilai dari n-2
2.	Numerator	Decimal	9	Nilai F tabel

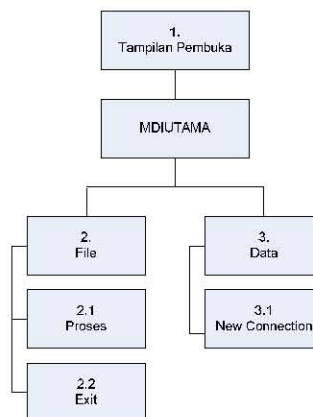
- c. Tabel r pearson, digunakan untuk memberikan nilai  $r_{\text{tabel}}$  dalam proses perhitungan analisis regresi.

**Tabel 4.3** Tabel r Pearson

Nama tabel pada perancangan : Tabel r pearson				
Nama tabel pada file : tbl_r_pearson				
No.	Nama Field	Tipe	Ukuran	Keterangan
1.	df	Int	4	Nilai dari n-2
2.	Nilai_r	Decimal	9	Nilai r tabel pearson

### 4.3 Struktur Program

Pada bagian ini, akan dijelaskan mengenai implementasi dari beberapa modul yang terintegrasi pada aplikasi data mining. Tapi sebelumnya gambar 4.2 di bawah ini akan menjelaskan mengenai struktur program antarmuka secara keseluruhan.



**Gambar 4.1** Struktur Program Antarmuka Keseluruhan

Selanjutnya akan dijelaskan mengenai struktur program antarmuka keseluruhan dengan tabel struktur program dibawah ini :

**Tabel 4.4** Tabel Struktur Program Antarmuka Keseluruhan

No. <i>Form</i>	Nama <i>Form/Menu</i>	DFD	Nama File	Keterangan
1.	Tampilan Pembuka		Frm_awal.cs	Tampilan pembukaan sebelum masuk ke dalam form utama
	MDIUTAMA		Frm_utama.cs	Menu utama yang menyediakan berbagai menu pilihan
2.	File		Frm_utama.cs	Salah satu sub menu yang tersedia di menu utama
2.1	Proses	Proses 1.2	Frm_utama.cs	Proses untuk melakukan perhitungan analisis regresi
2.2	Exit		Frm_utama.cs	Keluar dari aplikasi
3.	Data		Frm_utama.cs	Salah satu sub menu yang tersedia yang berhubungan dengan data
3.1	New Connection	Proses 1.1	Frm_connection.cs	Proses untuk melakukan koneksi dengan data/tabel yang akan dianalisis

#### 4.4 Petunjuk Penggunaan Aplikasi

Pada bagian ini, akan dijelaskan mengenai petunjuk dari penggunaan aplikasi data mining (DaMi-E). Petunjuk penggunaan dari aplikasi ini adalah sebagai berikut :

##### a. Form awal



**Gambar 4.2** Form Awal

Pada form awal ini, menjelaskan dari nama dan versi dari aplikasi data mining. Aplikasi data mining ini bernama DaMi-E dengan *versi release 1* dan menggunakan teknik regresi. Untuk melanjutkan aplikasi user bisa melanjutkan dengan memilih tombol In atau user bisa keluar dari aplikasi dengan memilih tombol Quit. Apabila user memilih tombol In, maka aplikasi akan menampilkan form utama dimana user bisa melakukan proses analisis.

##### b. Form utama

**Gambar 4.3** Form Utama

Gambar diatas merupakan tampilan dari form utama yang bisa diakses oleh user, pada form awal ini terdapat beberapa menu dan tombol yang bisa diakses oleh user. Tombol yang tersedia yaitu tombol New Connection dan tombol Proses, tombol New Connection digunakan untuk melakukan koneksi terhadap database dan tabel yang akan dianalisis. Sedangkan tombol Proses digunakan untuk memulai analisis regresi terhadap tabel yang telah terkoneksi. Untuk menu yang tersedia pada form ini akan dijelaskan pada tabel dibawah ini.

**Tabel 4.5** Tabel Menu dan Sub Menu Pada Form Utama

Menu	Sub Menu	Keterangan
File	Proses	Melakukan proses analisis regresi
	Exit	Keluar dari aplikasi
Data	New Connection	Melakukan koneksi terhadap database dan tabel yang akan dianalisis.

c. Form *connection*

**Gambar 4.4** Form *Connection*

Form *Connection* diatas merupakan form untuk melakukan proses koneksi terhadap database dan tabel yang akan dianalisis. Sebelumnya user harus mengisikan *text box* yang telah tersedia, yang harus diisikan user berupa nama server, *user name*, *password*, nama database, dan nama tabel yang akan dianalisis. Pada form ini tersedia beberapa tombol yang bisa diakses oleh user, tombol yang tersedia yaitu tombol Load, Connection, OK dan Reset. Tombol Load digunakan untuk mengisi *combo box* dengan nama database yang tersedia dari server yang telah dipilih dan tombol Connection digunakan untuk memulai koneksi dengan tabel yang telah dipilih. Setelah kita melakukan koneksi user bisa memilih tombol OK untuk melanjutkan proses analisis dengan kembali ke form utama atau user bisa melakukan koneksi ulang terhadap database dan tabel yang lain dengan menggunakan tombol reset.

## 4.5 Implementasi Terhadap Target Data

### 4.5.1 Implementasi Terhadap Tabel Sampel

Tabel sampel merupakan tabel yang terdiri dari beberapa *record*/tupel hasil sampling dari tabel populasi, metode sampling yang digunakan adalah *systematic sampling*. *Systematic sampling* kadang disebut juga dengan interval sampling yang berarti terdapat gap atau celah diantara data yang diambil, metode ini biasanya digunakan dalam industri. Kelebihan dari metode ini adalah sangat sederhana dalam memilih sampel yang dipilih secara acak dan sebaran datanya sangatlah bagus karena menyebar secara menyeluruh dari awal sampai akhir data. Kelemahan dari metode ini adalah dalam menempatkan data yang ke berapa sebagai awal atau acuan dari pengambilan data secara acak.

Adapun perhitungan interval dari metode ini adalah sebagai berikut :

$$I = \frac{N_{populasi}}{N_{sampel}}$$

### 4.5.2 Implementasi Terhadap Tabel Populasi

Merupakan tabel populasi dengan *record*/tupel atau N populasi sebanyak 1637 data tupel. Hasil pengujian terhadap kedua tabel yaitu tabel sampel dan tabel populasi yang berupa kesimpulan, grafik dan data nilai-nilai digunakan untuk

melakukan pengambilan keputusan oleh pihak yang berkepentingan yaitu unit Akademik Universitas Widyatama.

#### 4.6 Pengujian Program

Untuk mengetahui hubungan antar variabel-variabel penelitian dengan teori yang ada, maka dilakukanlah simulasi dengan melakukan pengujian terhadap aplikasi data mining (DaMi-E) yang telah dibuat dan membandingkan hasil proses analisis regresi dari aplikasi dengan hasil proses analisis regresi secara manual, adapun variabel yang terlibat adalah:

##### 4.6.1 Variabel Input

Variabel input atau masukan dalam hal ini adalah tabel sampel dan tabel populasi yang diumpankan terhadap aplikasi data mining. Tabel sampel merupakan tabel yang terdiri dari 91 *tupel/record* hasil sampling dari data populasi yang telah diurutkan dari kecil ke besar, sedangkan tabel populasi sebanyak 1637 data *tupel*. Dalam melakukan sampling metode yang digunakan adalah *systematic sampling*. Pengambilan 91 data sampel dari tabel populasi adalah sebagai berikut :

$$I = \frac{1637}{91} = 17.98$$
$$I = 18$$

Maka pengambilan 91 data dilakukan dalam interval 18 data dari titik awal atau titik acuan sampel data pertama diambil.

##### 4.6.2 Variabel Proses

Variabel proses yang terlibat disini adalah proses analisis regresi yang diimplementasikan ke dalam sebuah aplikasi data mining atau aplikasi DaMi-E dengan teknik regresi.

##### 4.6.3 Variabel Output

Variabel output atau keluaran dari aplikasi ini adalah sebuah hasil perhitungan dan grafik regresi yang merupakan representasi dari pengetahuan yang berhasil di gali (*mined*) dari tabel sampel dan tabel populasi.

#### 4.6.4 Pengujian Secara Manual Dengan Data Sampel

Tabel 4.6 adalah tabel yang menyajikan tupel data yang diambil secara acak dari tabel populasi. Atribut yang dimiliki dari tabel sampel yaitu nilai USM dan nilai IP.

Tabel 4.6 Tabel Sampel

N	Nilai_USM (X)	Nilai_IP (Y)	N	Nilai_USM (X)	Nilai_IP (Y)
1	9	26	47	38	67
2	15	75	48	39	67
3	19	87	49	39	48
4	21	54	50	39	56
5	24	8	51	39	65
6	25	82	52	40	35
7	26	74	53	40	63
8	27	48	54	40	81
9	28	58	55	41	51
10	29	69	56	41	47
11	30	78	57	41	58
12	31	68	58	41	60
13	31	51	59	42	79
14	31	78	60	42	54
15	31	66	61	42	72
16	31	67	62	43	72
17	32	64	63	44	69
18	32	40	64	44	55
19	32	65	65	44	79
20	32	4	66	44	56
21	33	31	67	45	75
22	33	21	68	45	86
23	33	83	69	45	64
24	33	54	70	46	63
25	33	66	71	46	93
26	34	78	72	48	82
27	34	84	73	48	77
28	34	41	74	48	83
29	34	51	75	49	52
30	35	86	76	49	90
31	35	2	77	50	78
32	35	64	78	51	51
33	35	42	79	52	92
34	36	77	80	53	76

<b>35</b>	36	48	<b>81</b>	53	73
<b>36</b>	36	69	<b>82</b>	54	77
<b>37</b>	36	49	<b>83</b>	55	69
<b>38</b>	36	80	<b>84</b>	56	71
<b>39</b>	37	44	<b>85</b>	58	72
<b>40</b>	37	56	<b>86</b>	59	95
<b>41</b>	37	60	<b>87</b>	61	96
<b>42</b>	37	84	<b>88</b>	64	81
<b>43</b>	37	38	<b>89</b>	66	90
<b>44</b>	38	25	<b>90</b>	70	84
<b>45</b>	38	62	<b>91</b>	79	81
<b>46</b>	38	73			

Dengan data tabel sampel diatas kita akan memulai melakukan pengujian secara manual, sesuai dengan teori-teori analisis regresi yang telah dijelaskan pada bab sebelumnya. Langkah pertama yaitu menghitung satuan-satuan yang diperlukan dan sebaiknya data disusun seperti dalam tabel berikut :

**Tabel 4.7** Tabel Penolong Data Sampel

N	X	Y	X <sup>2</sup>	Y <sup>2</sup>	XY	N	X	Y	X <sup>2</sup>	Y <sup>2</sup>	XY
<b>1</b>	9	26	81	676	234	<b>47</b>	38	67	1444	4489	2546
<b>2</b>	15	75	225	5625	1125	<b>48</b>	39	67	1521	4489	2613
<b>3</b>	19	87	361	7569	1653	<b>49</b>	39	48	1521	2304	1872
<b>4</b>	21	54	441	2916	1134	<b>50</b>	39	56	1521	3136	2184
<b>5</b>	24	8	576	64	192	<b>51</b>	39	65	1521	4225	2535
<b>6</b>	25	82	625	6724	2050	<b>52</b>	40	35	1600	1225	1400
<b>7</b>	26	74	676	5476	1924	<b>53</b>	40	63	1600	3969	2520
<b>8</b>	27	48	729	2304	1296	<b>54</b>	40	81	1600	6561	3240
<b>9</b>	28	58	784	3364	1624	<b>55</b>	41	51	1681	2601	2091
<b>10</b>	29	69	841	4761	2001	<b>56</b>	41	47	1681	2209	1927
<b>11</b>	30	78	900	6084	2340	<b>57</b>	41	58	1681	3364	2378
<b>12</b>	31	68	961	4624	2108	<b>58</b>	41	60	1681	3600	2460
<b>13</b>	31	51	961	2601	1581	<b>59</b>	42	79	1764	6241	3318
<b>14</b>	31	78	961	6084	2418	<b>60</b>	42	54	1764	2916	2268
<b>15</b>	31	66	961	4356	2046	<b>61</b>	42	72	1764	5184	3024
<b>16</b>	31	67	961	4489	2077	<b>62</b>	43	72	1849	5184	3096
<b>17</b>	32	64	1024	4096	2048	<b>63</b>	44	69	1936	4761	3036
<b>18</b>	32	40	1024	1600	1280	<b>64</b>	44	55	1936	3025	2420
<b>19</b>	32	65	1024	4225	2080	<b>65</b>	44	79	1936	6241	3476
<b>20</b>	32	4	1024	16	128	<b>66</b>	44	56	1936	3136	2464
<b>21</b>	33	31	1089	961	1023	<b>67</b>	45	75	2025	5625	3375
<b>22</b>	33	21	1089	441	693	<b>68</b>	45	86	2025	7396	3870

23	33	83	1089	6889	2739	69	45	64	2025	4096	2880
24	33	54	1089	2916	1782	70	46	63	2116	3969	2898
25	33	66	1089	4356	2178	71	46	93	2116	8649	4278
26	34	78	1156	6084	2652	72	48	82	2304	6724	3936
27	34	84	1156	7056	2856	73	48	77	2304	5929	3696
28	34	41	1156	1681	1394	74	48	83	2304	6889	3984
29	34	51	1156	2601	1734	75	49	52	2401	2704	2548
30	35	86	1225	7396	3010	76	49	90	2401	8100	4410
31	35	2	1225	4	70	77	50	78	2500	6084	3900
32	35	64	1225	4096	2240	78	51	51	2601	2601	2601
33	35	42	1225	1764	1470	79	52	92	2704	8464	4784
34	36	77	1296	5929	2772	80	53	76	2809	5776	4028
35	36	48	1296	2304	1728	81	53	73	2809	5329	3869
36	36	69	1296	4761	2484	82	54	77	2916	5929	4158
37	36	49	1296	2401	1764	83	55	69	3025	4761	3795
38	36	80	1296	6400	2880	84	56	71	3136	5041	3976
39	37	44	1369	1936	1628	85	58	72	3364	5184	4176
40	37	56	1369	3136	2072	86	59	95	3481	9025	5605
41	37	60	1369	3600	2220	87	61	96	3721	9216	5856
42	37	84	1369	7056	3108	88	64	81	4096	6561	5184
43	37	38	1369	1444	1406	89	66	90	4356	8100	5940
44	38	25	1444	625	950	90	70	84	4900	7056	5880
45	38	62	1444	3844	2356	91	79	81	6241	6561	6399
46	38	73	1444	5329	2774	N	ΣX	ΣY	ΣX <sup>2</sup>	ΣY <sup>2</sup>	ΣXY
						91	3629	5815	156383	407293	240216

Selanjutnya masukan angka-angka statistik dari tabel 4.7 pada persamaan rumus 2.2 untuk mencari  $b_0$  dan  $b_1$ , sehingga kita peroleh harga-harga :

$$b_0 = \frac{(\sum y_i)(\sum x_i^2) - (\sum x_i)(\sum x_i y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$

$$b_0 = \frac{(5815)(156383) - (3629)(240216)}{91(156383) - (3629)^2} = 35.4531$$

dan

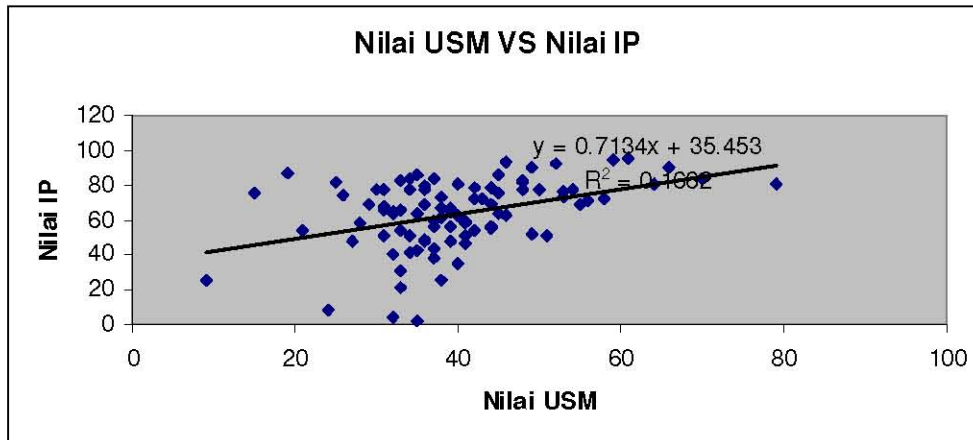
$$b_1 = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$

$$b_1 = \frac{91(240216) - (3629)(5815)}{91(156383) - (3629)^2} = 0.7133$$

Dengan demikian, persamaan regresi linear y atas x untuk data tabel sampel adalah :

$$\bar{y} = 35.4531 + 0.7133x$$

Grafik regresi dari persamaan regresi linear diatas dapat dilihat dalam gambar 4.5, grafik regresi ini didapat dengan menggunakan bantuan *software Microsoft Excel 2003*.



**Gambar 4.5** Grafik Regresi Data Tabel Sampel

Setelah kita mendapatkan persamaan regresi linear dan grafik regresi, maka langkah selanjutnya kita melakukan uji signifikan dengan langkah-langkah berikut :

- Menghitung jumlah kuadrat regresi atau *Sum of Square (SS)*, sesuai dengan rumus 2.4 :

$$SS_y = \sum y^2 - \frac{(\sum y)^2}{n} = 407293 - \frac{(5815)^2}{91} = 35708.11$$

$$SS_x = \sum x^2 - \frac{(\sum x)^2}{n} = 156383 - \frac{(3629)^2}{91} = 11661.67$$

$$SS_{xy} = \sum xy - \frac{(\sum X)(\sum Y)}{N} = 240216 - \frac{(3629)(5815)}{91} = 8318.91$$

- Menghitung jumlah kuadrat residu atau *Sum of Square for Error (SSE)*, sesuai dengan rumus 2.6 :

$$SSE = SS_y - b_1 SS_{xy} = 35708.11 - (0.71335)(8318.912) = 29773.77$$

- Menghitung rata-rata kuadrat residu atau *Mean Square Error (MSE)*, sesuai dengan rumus 2.5 :

$$MSE = \frac{SSE}{n-2} = \frac{29773.77}{89} = 334.5367$$

- Menghitung *Standar Error of Estimate* ( $s$ ), sesuai dengan rumus 2.11 :

$$s = \sqrt{334.5367} = 18.29034$$

- Menghitung *Standar Error* terhadap  $b_0$ , sesuai dengan rumus 2.9 :

$$s(b_0) = \frac{s \sqrt{\sum x^2}}{\sqrt{nSS_x}} = \frac{18.2903 \sqrt{156383}}{\sqrt{91 * 11661.67}} = 7.0212$$

- Menghitung *Standar Error* terhadap  $b_1$ , sesuai dengan rumus 2.10 :

$$s(b_1) = \frac{s}{\sqrt{SS_x}} = \frac{18.29034}{\sqrt{11661.67}} = 0.169371$$

- Menghitung interval kepercayaan untuk  $\beta_0$ , sesuai dengan rumus 2.7 :

$$\text{Dengan } t(\alpha/2, n-2) = t(5/2, 25-2) = t(0.25, 23) = 2.069$$

(nilai  $t$  diperoleh dari tabel nilai kritik sebaran  $t$ )

$$b_0 \pm t_{(\alpha/2, n-2)} s(b_0)$$

$$35.45312 \pm 2.069(7.0212)$$

1- $\alpha$	(1- $\alpha$ ) C.I for $\beta_0$
95%	35.4531 + or - 13.97

- Menghitung interval kepercayaan untuk  $\beta_1$ , sesuai dengan rumus 2.8 :

$$\text{Dengan } t(\alpha/2, n-2) = t(5/2, 25-2) = t(0.25, 23) = 2.069$$

(nilai  $t$  diperoleh dari tabel nilai kritik sebaran  $t$ )

$$b_1 \pm t_{(\alpha/2, n-2)} s(b_1)$$

$$0.71335 \pm 2.069(0.16937)$$

1- $\alpha$	(1- $\alpha$ ) C.I for $\beta_1$
95%	0.71335 + or - 0.3371

- Menghitung rata-rata kuadrat regresi atau *Mean Square Regression* (MSR), sesuai dengan rumus 2.13 :

$$SSR = b_1 SS_{xy} = 0.71335 * 8318.912 = 5934.2958$$

$$MSR = \frac{SSR}{1} = \frac{5934.2958}{1} = 5934.2958$$

- Menghitung  $F_{hitung}$ , sesuai dengan rumus 2.12 :

$$F_{hitung} = \frac{MSR}{MSE} = \frac{5934.2958}{334.53674} = 17.738$$

- Menghitung  $F_{\text{tabel}}$  :

Menggunakan  $\alpha = 0.01$

$$F_{\text{tabel}} = F(1, n-2)$$

$$F_{\text{tabel}} = F(1, 89) = 6.9330$$

- Menghitung korelasi pearson ( $r$ ), sesuai dengan rumus 2.15 :

$$r = \frac{SS_{xy}}{\sqrt{SS_x SS_y}} = \frac{8318.912}{\sqrt{11661.67 * 35708.11}} = 0.40766$$

- Menghitung  $r_{\text{tabel}}$  :

Menggunakan  $\alpha = 0.01$

$$r_{\text{tabel}} = r(n-2, 0.01)$$

$$r_{\text{tabel}} = r(89, 0.01) = 0.2686$$

- Menghitung Koefisien Determinasi (KD), sesuai dengan rumus 2.16 :

$$KD = r^2 * 100\%$$

$$KD = 0.40766^2 * 100\% = 16.61\%$$

Setelah melakukan proses perhitungan analisis regresi, maka nilai-nilai yang diperoleh bisa ditampilkan dalam tabel 4.8 dan selain itu bisa diambil beberapa kesimpulan seperti yang ditampilkan dalam tabel 4.9.

**Tabel 4.8** Tabel Hasil Perhitungan Data Sampel

$b_0$	$b_1$	$r_{\text{hitung}}$	KD	
35.45312	0.71335	0.40766	16.61 %	
Df	s	$s(b_0)$	$s(b_1)$	
89	18.29034	7.0212	0.169371	
Tabel ANOVA				
SV	SS	DF	MS	Fratio
Regression	5934.2958	1	5934.2958	17.738
Error	29773.77	89	334.53674	
Total	35708.11	91		

Tabel 4.9 Tabel Kesimpulan Dari Data Sampel

$F_{hitung}$	$F_{tabel}$	Kesimpulan
17.738	6.9330	$F_{hitung} \geq F_{tabel}$ , maka tolak $H_0$ atau <b>signifikan</b> . Sehingga hipotesis awal ( $H_a$ ) bisa diterima
$r_{hitung}$	$r_{tabel}$	Kesimpulan
0.40766	0.2686	$r_{hitung} \geq r_{tabel}$ , maka tolak $H_0$ atau <b>signifikan</b> . Sehingga hipotesis awal ( $H_a$ ) bisa diterima
$r_{hitung}$	Keratan Hubungan ( <i>Guilford</i> )	
0.40766	Menurut tafsiran kriteria Guilford, maka nilai $r_{hitung}$ yang didapat memiliki <b>hubungan yang cukup</b> .	
Koefisien Determinasi (KD)	Kesimpulan	
16.61 %	Dengan demikian besaran pengaruh antara variabel $x$ (nilai USM) terhadap variabel $y$ (nilai IP) yaitu sebesar <b>28.21 %</b> .	

#### 4.6.5 Pengujian Secara Manual Dengan Data Populasi

Tabel 4.10 adalah tabel yang menyajikan tupel data yang ada pada tabel populasi, data yang ditampilkan tidak keseluruhan data dikarenakan banyaknya data yang tersedia pada tabel populasi. Atribut yang dimiliki dari tabel sampel yaitu nilai USM dan nilai IP.

Tabel 4.10 Tabel Populasi

N	Nilai USM	Nilai IP	N	Nilai USM	Nilai IP
1	35.6	67.5	19	35.13	62.5
2	35.58	53.25	20	36.7	37.5
3	70.63	48.75	40	40.4	59.5
4	44.5	89.25	50	41.5	75
5	43.47	81	60	33.33	74.5
6	46.2	93.25	70	34.05	58.25
7	39.17	94	80	43.02	60
8	25.5	53.25	100	73.4	78
9	34.86	53.25	200	39.62	26.5
10	54.97	90.5	300	37.38	87.25

11	37.92	70.75	400	46.47	76.25
12	37.25	69.75	600	47.86	80.25
13	35.35	53	800	62.57	14
14	34.12	41	1000	46.76	79
15	44.12	96.25	1200	40	63
16	42.42	78	1400	34.73	53.25
17	42.97	51.25	1600	26.23	68.25
18	46.02	81	1637	13.2	53.25

Dengan data tabel populasi diatas kita akan memulai melakukan pengujian secara manual, sesuai dengan teori-teori analisis regresi yang telah dijelaskan pada bab sebelumnya. Langkah pertama yaitu menghitung satuan-satuan yang diperlukan dan sebaiknya data disusun seperti dalam tabel berikut :

**Tabel 4.11** Tabel Penolong Data Populasi

N	Nilai USM (X)	Nilai IP (Y)	X <sup>2</sup>	Y <sup>2</sup>	XY
1	35.6	67.5	1267.36	4556.25	2403
2	35.58	53.25	1265.9364	2835.5625	1894.635
3	70.63	48.75	4988.5969	2376.5625	3443.2125
4	44.5	89.25	1980.25	7965.5625	3971.625
5	43.47	81	1889.6409	6561	3521.07
6	46.2	93.25	2134.44	8695.5625	4308.15
7	39.17	94	1534.2889	8836	3681.98
8	25.5	53.25	650.25	2835.5625	1357.875
9	34.86	53.25	1215.2196	2835.5625	1856.295
10	54.97	90.5	3021.7009	8190.25	4974.785
11	37.92	70.75	1437.9264	5005.5625	2682.84
12	37.25	69.75	1387.5625	4865.0625	2598.1875
13	35.35	53	1249.6225	2809	1873.55
14	34.12	41	1164.1744	1681	1398.92
15	44.12	96.25	1946.5744	9264.0625	4246.55
16	42.42	78	1799.4564	6084	3308.76
17	42.97	51.25	1846.4209	2626.5625	2202.2125
18	46.02	81	2117.8404	6561	3727.62
19	35.13	62.5	1234.1169	3906.25	2195.625
20	36.77	37.5	1352.0329	1406.25	1378.875
40	40.4	59.5	1632.16	3540.25	2403.8
50	41.5	75	1722.25	5625	3112.5
60	33.33	74.5	1110.8889	5550.25	2483.085
70	34.05	58.25	1159.4025	3393.0625	1983.4125
80	43.02	60	1850.7204	3600	2581.2

100	73.4	78	5387.56	6084	5725.2
200	39.62	26.5	1569.7444	702.25	1049.93
300	37.38	87.25	1397.2644	7612.5625	3261.405
400	46.47	76.25	2159.4609	5814.0625	3543.3375
...	...	...	...	...	...
1637	13.2	53.25	174.24	2835.5625	702.9
N	ΣX	ΣY	ΣX <sup>2</sup>	ΣY <sup>2</sup>	ΣXY
1637	66438.90	101266.25	2873041.90	6826486.30	4198361.60

Selanjutnya masukan angka-angka statistik dari tabel 4.11 pada persamaan

rumus 2.2 untuk mencari b<sub>0</sub> dan b<sub>1</sub>, sehingga kita peroleh harga-harga :

$$b_0 = \frac{(\sum y_i)(\sum x_i^2) - (\sum x_i)(\sum x_i y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$

$$b_0 = \frac{(101266.25)(2873041.9) - (66438.9)(4198361.6)}{1637(2873041.9) - (66438.9)^2} = 41.54291167$$

dan

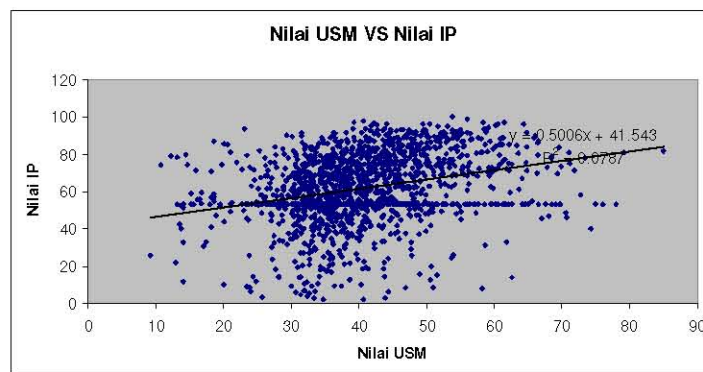
$$b_1 = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$

$$b_1 = \frac{1637(4198361.6) - (66438.9)(101266.25)}{1637(2873041.9) - (66438.9)^2} = 0.500617915$$

Dengan demikian, persamaan regresi linear y atas x untuk data tabel sampel adalah :

$$\bar{y} = 41.543 + 0.5x$$

Grafik regresi dari persamaan regresi linear diatas dapat dilihat dalam gambar 4.6, grafik regresi ini didapat dengan menggunakan bantuan *software Microsoft Excel 2003*.



Gambar 4.6 Grafik Regresi Data Tabel Populasi

Setelah kita mendapatkan persamaan regresi linear dan grafik regresi, maka langkah selanjutnya kita melakukan uji signifikan dengan langkah-langkah berikut :

- Menghitung jumlah kuadrat regresi atau *Sum of Square* (SS), sesuai dengan rumus 2.4 :

$$SS_y = \sum y^2 - \frac{(\sum y)^2}{n} = 6826486.3 - \frac{(101266.25)^2}{1637} = 562067.61$$

$$SS_x = \sum x^2 - \frac{(\sum x)^2}{n} = 2873041.9 - \frac{(66438.9)^2}{1637} = 176568.21$$

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 4198361.6 - \frac{(66438.9)(101266.25)}{1637} = 88393.21$$

- Menghitung jumlah kuadrat residu atau *Sum of Square for Error* (SSE), sesuai dengan rumus 2.6 :

$$SSE = SS_y - b_1 SS_{xy} = 562067.61 - (0.500617915)(88393.21) = 517816.38$$

- Menghitung rata-rata kuadrat residu atau *Mean Square Error* (MSE), sesuai dengan rumus 2.5 :

$$MSE = \frac{SSE}{n-2} = \frac{517816.3855}{1635} = 316.7072694$$

- Menghitung *Standar Error of Estimate* (s), sesuai dengan rumus 2.11 :

$$s = \sqrt{316.7072694} = 17.79627122$$

- Menghitung *Standar Error* terhadap  $b_0$ , sesuai dengan rumus 2.9 :

$$s(b_0) = \frac{s \sqrt{\sum x^2}}{\sqrt{n SS_x}} = \frac{17.79627122 \sqrt{2873041.9}}{\sqrt{1637 * 176568.21}} = 1.774268732$$

- Menghitung *Standar Error* terhadap  $b_1$ , sesuai dengan rumus 2.10 :

$$s(b_1) = \frac{s}{\sqrt{SS_x}} = \frac{17.79627122}{\sqrt{176568.21}} = 0.042351886$$

- Menghitung interval kepercayaan untuk  $\beta_0$ , sesuai dengan rumus 2.7 :

$$\text{Dengan } t(\alpha/2, n-2) = t(5/2, 1637-2) = t(0.25, 1635) = 1.960$$

(nilai  $t$  diperoleh dari tabel nilai kritik sebaran  $t$ )

$$b_0 \pm t_{(\alpha/2, n-2)} s(b_0)$$

$$41.54291167 \pm 1.960(1.774268732)$$

1- $\alpha$	(1- $\alpha$ ) C.I for $\beta_0$
95%	41.54291167 + or - 3.477566714

- Menghitung interval kepercayaan untuk  $\beta_1$ , sesuai dengan rumus 2.8 :

$$\text{Dengan } t(\alpha/2, n-2) = t(5/2, 1637-2) = t(0.25, 1635) = 1.960$$

(nilai  $t$  diperoleh dari tabel nilai kritik sebaran  $t$ )

$$b_1 \pm t_{(\alpha/2, n-2)} s(b_1)$$

$$0.500617915 \pm 1.960(0.042351886)$$

1- $\alpha$	(1- $\alpha$ ) C.I for $\beta_1$
95%	0.500617915 + or - 0.083009697

- Menghitung rata-rata kuadrat regresi atau *Mean Square Regression* (MSR), sesuai dengan rumus 2.13 :

$$SSR = b_1 SS_{xy} = 0.500617915 * 88393.21 = 44251.22449$$

$$MSR = \frac{SSR}{1} = \frac{44251.22449}{1} = 44251.22449$$

- Menghitung  $F_{hitung}$ , sesuai dengan rumus 2.12 :

$$F_{hitung} = \frac{MSR}{MSE} = \frac{44251.22449}{316.7072694} = 139.7227938$$

- Menghitung  $F_{tabel}$  :

Menggunakan  $\alpha = 0.01$

$$F_{tabel} = F(1, 1637-2)$$

$$F_{tabel} = F(1, 1635) = 6.64$$

- Menghitung korelasi pearson ( $r$ ), sesuai dengan rumus 2.15 :

$$r = \frac{SS_{xy}}{\sqrt{SS_x SS_y}} = \frac{88393.21}{\sqrt{176568.21 * 562067.61}} = 0.280587532$$

- Menghitung  $r_{tabel}$  :

$r_{tabel}$  pada data populasi tidak tersedia, dikarenakan jumlah data populasi yang begitu banyak.

- Menghitung Koefisien Determinasi (KD), sesuai dengan rumus 2.16 :

$$KD = r^2 * 100\%$$

$$KD = 0.280587532^2 * 100 \% = 7.87\%$$

Setelah melakukan proses perhitungan analisis regresi, maka nilai-nilai yang diperoleh bisa ditampilkan dalam tabel 4.12 dan selain itu bisa diambil beberapa kesimpulan seperti yang ditampilkan dalam tabel 4.13.

**Tabel 4.12** Tabel Hasil Perhitungan Data Populasi

$b_0$	$b_1$	$r_{hitung}$	KD	
41.54291167	0.500617915	0.280587532	7.87%	
Df	s	$s(b_0)$	$s(b_1)$	
1635	17.79627122	1.774268732	0.042351886	
Tabel ANOVA				
SV	SS	DF	MS	Fratio
Regression	44251.22449	1	44251.22449	139.7227938
Error	517816.3855	1635	316.7072694	
Total	562067.61	1636		

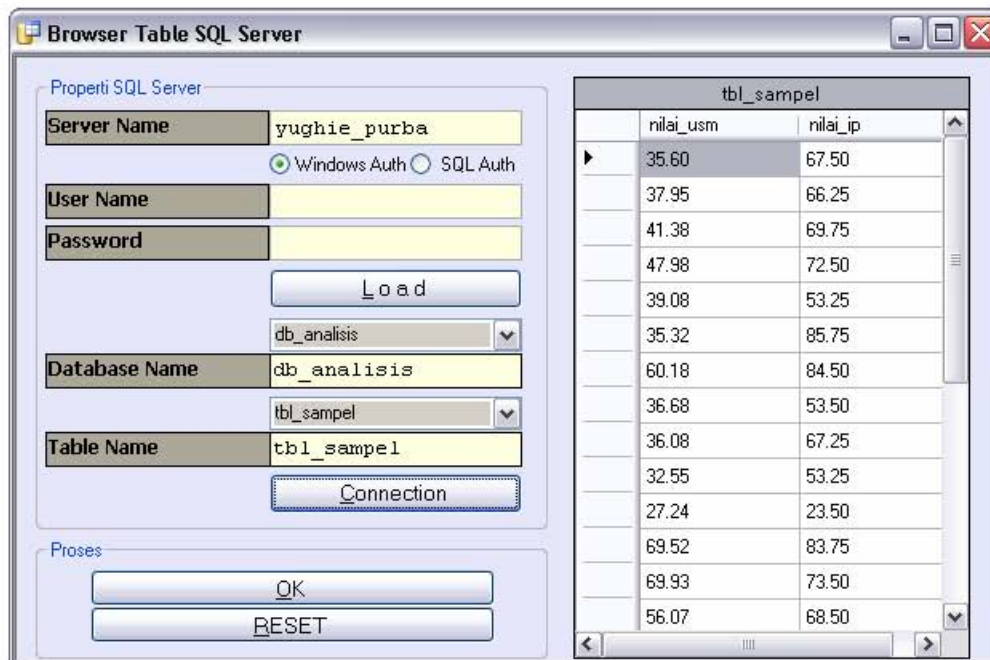
**Tabel 4.13** Tabel Kesimpulan Dari Data Populasi

$F_{hitung}$	$F_{tabel}$	Kesimpulan
139.7227938	6.64	$F_{hitung} \geq F_{tabel}$ , maka tolak $H_0$ atau <b>signifikan</b> . Sehingga hipotesis awal ( $H_a$ ) bisa diterima
$r_{hitung}$	$r_{tabel}$	Kesimpulan
0.280587532	-	Dikarenakan $r_{tabel}$ tidak tersedia maka kesimpulannya tidak cocok.
$r_{hitung}$	Keratan Hubungan ( <i>Guilford</i> )	
0.280587532	Menurut tafsiran kriteria Guilford, maka nilai $r_{hitung}$ yang didapat memiliki <b>hubungan yang sangat kecil (tidak erat)</b> .	
Koefisien Determinasi (KD)	Kesimpulan	
7.87%	Dengan demikian besaran pengaruh antara variabel $x$ (nilai USM) terhadap variabel $y$ (nilai IP) yaitu sebesar <b>7.87%</b> .	

#### 4.6.6 Pengujian Menggunakan Aplikasi DaMi-E Dengan Data Sampel

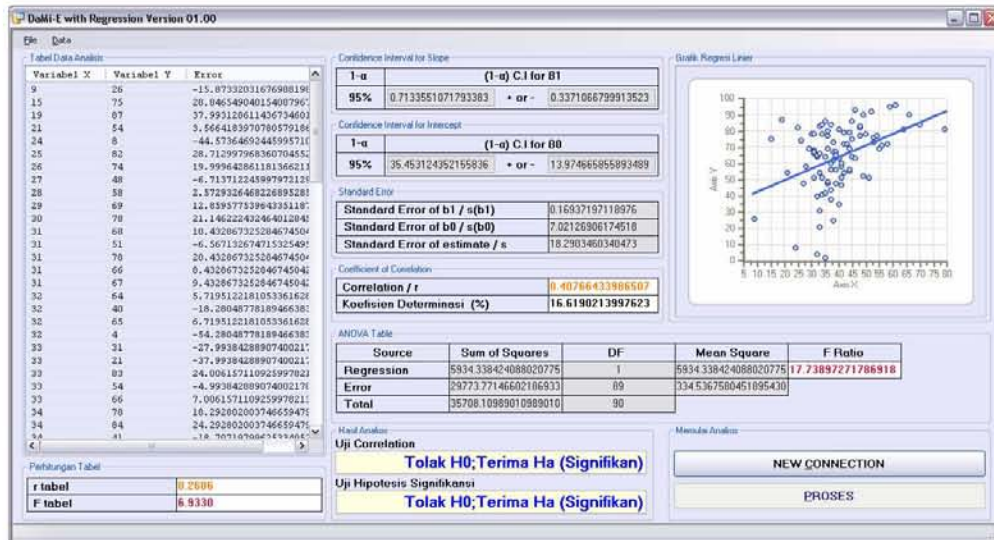
Pada sub bab ini akan dijelaskan hasil dari proses analisis regresi menggunakan aplikasi DaMi-E dengan asupan data berupa tabel sampel. Tabel sampel yang digunakan sudah diperlihatkan pada tabel 4.6.

Langkah pertama untuk menggunakan aplikasi DaMi-E ini harus dilakukan koneksi terlebih dahulu terhadap database atau tabel yang akan dianalisis dengan regresi. Untuk melakukan koneksi bisa dilakukan dengan cara memilih tombol *New Connection* lalu akan muncul form *browser table SQL Server*, untuk melakukan koneksi terhadap tabel sampel kita bisa mengisi *textbox-textbox* yang ada di form *browser table SQL Server* seperti gambar dibawah ini.



**Gambar 4.7** Koneksi Terhadap Tabel Sampel

Setelah proses koneksi dilakukan bisa dilanjutkan dengan memilih tombol OK dan akan kembali pada form utama. Setelah itu bisa dimulai proses analisis regresi terhadap tabel sampel dengan memilih tombol proses. Setelah melakukan proses analisis maka tampilan yang akan disajikan aplikasi DaMi-E seperti gambar 4.8 :



Gambar 4.8 Hasil Analisis Terhadap Tabel Sampel

Setelah kita melakukan pengujian secara manual dan dengan aplikasi DaMi-E, maka kita akan membandingkannya dengan menggunakan tabel 4.14.

Tabel 4.14 Tabel Perbandingan Uji Tabel Sampel

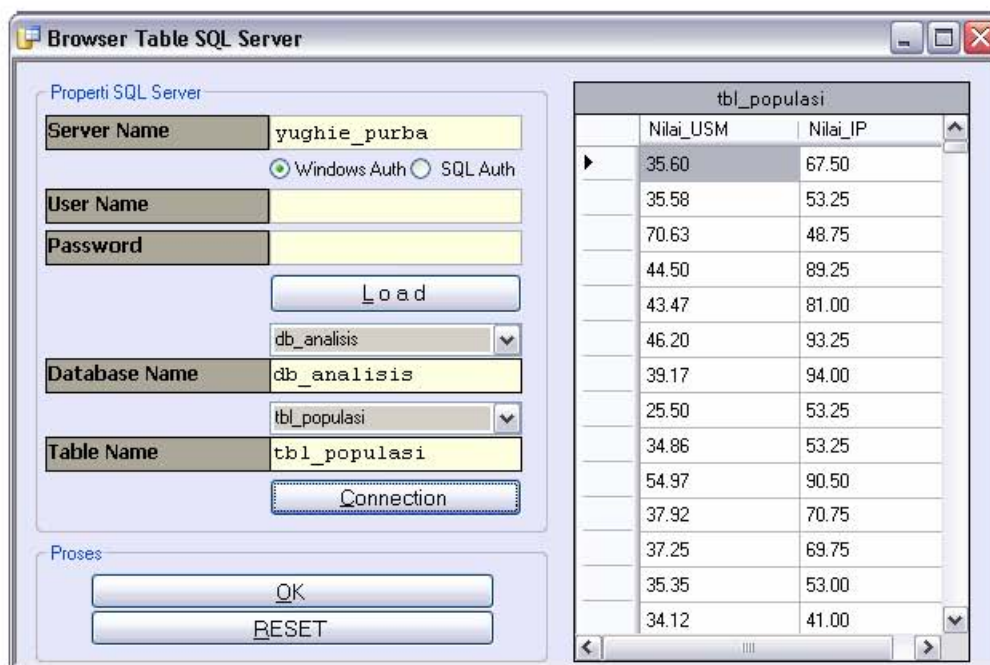
Satuan Nilai	Manual	DaMi-E
$b_0$	35.45312	35.45312
$b_1$	0.713355	0.713355
$s(b_1)$	0.16937	0.16937
$s(b_0)$	7.02126	7.02126
$s$	18.2903	18.2903
$r$	0.40766	0.40766
DK	16.61 %	16.61 %
F	17.738	17.738
Uji Correlation	Signifikan	Signifikan
Uji Hipotesis	Signifikan	Signifikan

Hasil akhir dari proses regresi terhadap tabel sampel yang dilakukan baik secara manual maupun dengan simulasi menggunakan program aplikasi DaMi-E ternyata menghasilkan keluaran atau output yang sama atau menghasilkan keluaran yang sesuai. Dari proses pengujian ini dapat disimpulkan bahwa aplikasi yang dibangun telah sesuai dengan tujuan yang diharapkan.

#### 4.6.7 Pengujian Menggunakan Aplikasi DaMi-E Dengan Data Populasi

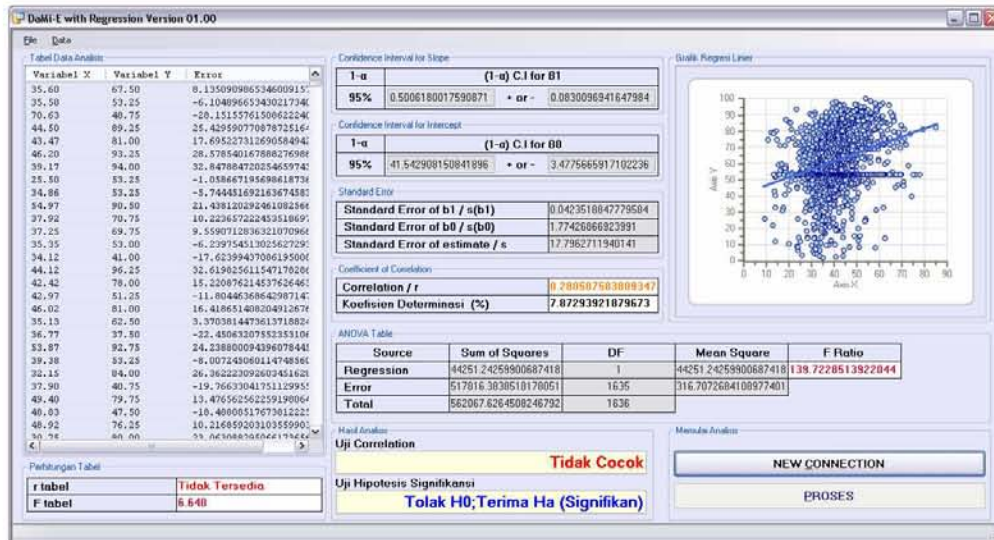
Setelah kita membandingkan hasil uji tabel sampel, maka selanjutnya akan melakukan perbandingan hasil uji tabel populasi. Sebelumnya aplikasi DaMi-E memerlukan asupan data berupa tabel populasi, tabel populasi yang digunakan sudah diperlihatkan pada tabel 4.10.

Langkah pertama untuk menggunakan aplikasi DaMi-E ini harus dilakukan koneksi terlebih dahulu terhadap database atau tabel yang akan dianalisis dengan regresi. Untuk melakukan koneksi bisa dilakukan dengan cara memilih tombol *New Connection* lalu akan muncul form *browser table SQL Server*, untuk melakukan koneksi terhadap tabel sampel kita bisa mengisi *textbox-textbox* yang ada di form *browser table SQL Server* seperti gambar dibawah ini.



**Gambar 4.9** Koneksi Terhadap Tabel Populasi

Setelah proses koneksi dilakukan bisa dilanjutkan dengan memilih tombol OK dan akan kembali pada form utama. Setelah itu bisa dimulai proses analisis regresi terhadap tabel populasi dengan memilih tombol proses. Setelah melakukan proses analisis maka tampilan yang akan disajikan aplikasi DaMi-E seperti gambar 4.10 :



Gambar 4.10 Hasil Analisis Terhadap Tabel Populasi

Setelah kita melakukan pengujian secara manual dan dengan aplikasi DaMi-E, maka kita akan membandingkannya dengan menggunakan tabel 4.15.

Tabel 4.15 Tabel Perbandingan Uji Tabel Populasi

Satuan Nilai	Manual	DaMi-E
$b_0$	41.54291167	41.542908150841896
$b_1$	0.500617915	0.5006180017590871
$s(b_1)$	0.042351886	0.0423518847779584
$s(b_0)$	1.774268732	1.77426866923991
$s$	17.79627122	17.7962711940141
$r$	0.280587532	0.280587583809347
$DK$	7.87 %	7.87293921879673 %
$F$	139.7227938	139.72285139
Uji Correlation	Tidak Cocok	Tidak Cocok
Uji Hipotesis	Signifikan	Signifikan

Hasil akhir dari proses regresi terhadap tabel populasi yang dilakukan baik secara manual maupun dengan simulasi menggunakan program aplikasi DaMi-E ternyata menghasilkan keluaran atau output yang sama atau menghasilkan keluaran yang sesuai. Dari proses pengujian ini dapat disimpulkan bahwa aplikasi yang dibangun telah sesuai dengan tujuan yang diharapkan.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan pelaksanaan penelitian tugas akhir mulai dari tahap pengumpulan data, analisa, perancangan sampai pada tahap implementasi, penulis dapat mengambil kesimpulan sebagai berikut :

1. Aplikasi data mining yang dibangun ini merupakan aplikasi yang membantu untuk menggali dan menemukan pola-pola yang tersembunyi, yang tidak diketahui antara data nilai USM dengan data IP.
2. Tahap data *preprocessing* sangat berpengaruh terhadap proses dan hasil proses analisis regresi itu sendiri, dimana akan menghasilkan suatu pola atau aturan dengan akurasi yang tinggi.
3. Analisis regresi yang dilakukan terhadap data sampel dengan menggunakan aplikasi data mining, mendapatkan pola atau aturan sebagai berikut :

**Tabel 5.1** Kesimpulan Data Sampel

Persamaan Regresi Linear	Kesimpulan
$\bar{y} = 35.4531 + 0.71335x$	Jika nilai USM (x) bertambah, maka rata-rata nilai IP (y) bertambah dengan besaran 35.4531.
Korelasi	Kesimpulan
0.40766	Menurut tafsiran kriteria Guilford, maka nilai USM terhadap nilai IP memiliki hubungan yang cukup
Koefisien Determinasi	Kesimpulan
16.61%	Dengan demikian besar pengaruh antara nilai USM terhadap nilai IP yaitu sebesar 16.61%. Sehingga tujuan penelitian ini untuk mendapatkan besaran pengaruh 70% tidak dapat terpenuhi.

4. Analisis regresi yang dilakukan terhadap data populasi dengan menggunakan aplikasi data mining, mendapatkan pola atau aturan sebagai berikut :

**Tabel 5.2** Kesimpulan Data Populasi

Persamaan Regresi Linear	Kesimpulan
$\bar{y} = 41.5429 + 0.5x$	Jika nilai USM (x) bertambah, maka rata-rata nilai IP (y) bertambah dengan besaran 41.5429.
Korelasi	Kesimpulan
0.280	Menurut tafsiran kriteria Guilford, maka nilai USM terhadap nilai IP memiliki hubungan yang sangat kecil (tidak erat).
Koefisien Determinasi	Kesimpulan
7.87%	Dengan demikian besar pengaruh antara nilai USM terhadap nilai IP yaitu sebesar 7.87%. Sehingga tujuan penelitian ini untuk mendapatkan besaran pengaruh 70% tidak dapat terpenuhi.

## 5.2 Saran

Berikut ini adalah saran yang mungkin dapat digunakan untuk pengembangan penelitian dan aplikasi lebih lanjut :

1. Diharapkan aplikasi data mining ini dapat digunakan oleh Universitas Widyatama sebagai *user* untuk mendukung suatu pengambilan keputusan, selain itu diharapkan juga *user* bisa melakukan pengujian terhadap aplikasi ini, sehingga kekurangan dari aplikasi ini dapat diketahui, yang kemudian bisa dilakukan penyempurnaan terhadap aplikasi ini.
2. Diharapkan aplikasi data mining ini dapat dikembangkan dengan menggunakan teknik-teknik yang ada dalam data mining selain teknik

regresi linier sederhana, sehingga aplikasi data mining ini menjadi aplikasi data mining yang canggih.

3. Diharapkan aplikasi data mining ini dapat dikembangkan agar dapat menganalisis lebih dari 2 variabel data, sehingga aplikasi ini menjadi aplikasi yang dinamis, yang tidak hanya bisa menangani 2 variabel data saja.
4. Diharapkan penelitian ini bisa dilanjutkan atau dikembangkan, dengan mengganti variabel X dan variabel Y dengan jenis data yang lain, yang lebih memiliki hubungan yang erat.
5. Diharapkan dilakukan pemeliharaan terhadap aplikasi data mining ini sehingga aplikasi data mining ini dapat dimanfaatkan seoptimal mungkin.

## DAFTAR PUSTAKA

- [1] Aczel, Amir dan Jayavel Sounderpandian. *Complete Business Statistics Sixth Edition*. Singapore: MC Graw Hill, 2006.
- [2] Fajar, Abdullah, dan Guntari Sekarwangi. *Modul Praktikum Statistika Dasar*. Jurusan Teknik Informatika Universitas Widyatama, 2004.
- [3] Irianto, Agus. *Statistik Konsep Dasar dan Aplikasinya*. Jakarta: Pranada Media, 2004.
- [4] Jaenudin. *Belajar Sendiri .Net dengan Visual C# 2005*. Yogyakarta: ANDI, 2005.
- [5] Martina, Inge. *36 Jam Belajar Komputer Microsoft SQL Server 2000*. Jakarta: Elex Media Komputindo, 2002.
- [6] Riduwan. *Dasar-dasar Statistika*. Bandung: Alfabeta, 2003
- [7] Santosa, Budi. *Data Mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis Teori dan Aplikasi*. Yogyakarta: Graha Ilmu, 2007.
- [8] Santosa, Budi. *Data Mining Terapan dengan MATLAB*. Yogyakarta: Graha Ilmu, 2007.
- [9] Walpole, Ronald. *Pengantar Statistika Edisi ke-3*. Jakarta: Gramedia Pustaka Utama, 1992.
- [10] <http://209.85.175.104/search?q=cache:ReFkdtbvh6QJ:home.unpar.ac.id>
- [11] <http://209.85.175.104/search?q=cache:bplWq0IfDs8J:www.bogor.net>
- [12] <http://ilmukomputer.com/2006/08/29/mengenal-data-mining>.
- [13] <http://www.beritanet.com/Literature/Kamus-Jargon/data-mining.html>.
- [14] [http://en.wikipedia.org/wiki/Data\\_mining](http://en.wikipedia.org/wiki/Data_mining).