

Kajian *Partial Least Squares* (Studi Kasus: Regresi Cox-PLS)

[A Study of Partial Least Squares (Case Study: Cox-PLS Regression)]

Retno Subekti dan Rosita Kusumawati

*Juridik Matematika, FMIPA, Universitas Negeri Yogyakarta (UNY),
Kampus Karangmalang, Sleman, DI Yogyakarta 55281
retnosubekti@uny.ac.id*

diterima 2 Desember 2013, disetujui 3 Februari 2014

Abstrak

Indikasi multikolinieritas pada analisis regresi akan menyebabkan salah penafsiran saat menginterpretasikan hasil. Salah satu penanganan terhadap kasus multikolinieritas adalah dengan menggunakan PLS (*partial least squares*). Tujuan penelitian ini adalah mengkaji pendekatan PLS untuk memberikan gambaran PLS secara umum dan penerapannya pada regresi Cox.

Secara sederhana PLS menggunakan pendekatan ordinary least square (OLS) seperti pada regresi linear. Perkembangan PLS dibagi menjadi dua bagian yaitu PLS-R (*regression*) untuk regresi dan PLS-PM (*path modelling*) untuk analisis jalur.

Hasil penelitian ini adalah kajian PLS secara umum beserta klasifikasinya baik konsep dan metodologinya serta penerapannya. PLS secara umum dibagi menjadi dua cabang yaitu PLS regresi dan PLS analisis jalur. Pada penerapan PLS dalam penelitian ini dipilih kasus PLS untuk regresi Cox. Data yang digunakan adalah data ketahanan hidup penderita TB (*tuber colosis*) di Yogyakarta yang diperoleh dari sebuah rumah sakit swasta di Yogyakarta. Secara umum, data dianalisis menggunakan regresi Cox, tetapi karena terdapat multikolinieritas maka terdapat kesalahan dalam penafsiran signifikansi model. Dengan menggunakan regresi Cox diperoleh satu komponen PLS yang terdiri dari satu variabel bebas yaitu kelas perawatan. Hasil evaluasi diperoleh ketepatan prediksi meningkat menjadi 64% dibandingkan jika digunakan regresi Cox yang semula 56 %.

Kata kunci: PLS, regresi Cox, analisis Jalur

Abstract

Indication of multicollinearity in regression analysis will lead to wrong interpretation when interpreting the results. One of the handling of the case of multicollinearity is to use of PLS (*partial least squares*). The purpose of this study is to provide a general overview of PLS. The results of this study are in general PLS study along both the concept and the classification methodology and its application. PLS is generally divided into two branches, namely PLS regression and path analysis. In the application of PLS, the data used TB patient survival (*tuberculosis*) in Yogyakarta, which is obtained from a private hospital in Yogyakarta. The data were analyzed using Cox regression, but there is multicollinearity so then there is an error in the interpretation of the significance of the model. By using PLS-Cox regression, we obtained one PLS component consisting of one independent variable, namely class care.

Key words: PLS regression, PLS path modelling, Cox regression

Pendahuluan

Beberapa hal yang perlu diperhatikan ketika melakukan analisis regresi linear antara lain adalah pemeriksaan asumsi-asumsi seperti normalitas, linearitas, dan homoskedastisitas dan multikolinieritas. Apabila diantara variabel predictor atau independennya ternyata terdapat korelasi yang cukup tinggi, maka dapat memunculkan salah penafsiran saat menginterpretasikan output yang dihasilkan. Sebagai contoh kesalahan interpretasi output adalah saat pengujian model regresi dinyatakan signifikan secara statistik, tetapi saat dilakukan uji signifikansi parsial atau individu variabel independennya diperoleh kesimpulan tidak ada satupun yang berpengaruh terhadap variabel respon.

PLS muncul pertama kali di tahun 1985 oleh Herman Wold [1]. Metode PLS ini berkembang pada berbagai bidang seperti bidang kimia, ekonomi, sosial, dan bisnis. Metode PLS sendiri dapat dikembangkan tidak hanya untuk regresi linear berganda. Beberapa pengembangan *generalized linear* regresi PLS lainnya adalah pada regresi logistik, regresi Cox seperti yang dijelaskan oleh Bastien et.al. dalam papernya tentang PLS-GLR [2] dan regresi multivariate, serta PLS untuk analisis jalur.

Untuk menunjukkan bagaimana cara kerja PLS, dalam penelitian ini akan diuraikan konsep PLS pada regresi linear berganda. Selanjutnya beberapa contoh PLS pada regresi lain yaitu regresi logistik dan regresi Cox. Peneliti tertarik untuk mengembangkan PLS pada regresi lain seperti regresi Poison, regresi multivariate ataupun pengembangan PLS pada analisis jalur atau SEM tetapi pembahasan akan disesuaikan dengan waktu penelitian. Dengan kegunaan PLS yang sangat variatif sehingga pengembangan dan aplikasinya juga beragam maka topik PLS masih sangat memungkinkan untuk dipilih mahasiswa sebagai topik skripsi baik dari penjabaran modelnya hingga aplikasinya. Dalam *Handbook of Partial Least Squares* [3], terdapat contoh beberapa aplikasi PLS untuk riset marketing seperti indeks kepuasan konsumen dan PLS dengan pengembangan regresi lain juga diajukan oleh Preda dan Saporta [4] tahun 2005 yaitu *clusterwise regression*.

Analisis *survival* merupakan penyelidikan daya tahan hidup suatu unit atau individu pada suatu keadaan tertentu. Uji hidup biasa digunakan dalam bidang teknik, biologi, kedokteran dan lain-lain. Penelitian-penelitian tersebut menggunakan data yang berkaitan dengan waktu hidup dari suatu individu, untuk menganalisis data waktu hidup tersebut disebut analisis *survival*. Analisis *survival* berfokus pada penelitian awal berlanjut pada tahap berikutnya sampai muncul suatu kejadian. Kejadian tersebut dapat berupa perkembangan suatu penyakit, respon terhadap perawatan, kambuhnya suatu penyakit, kematian atau kejadian lain yang ditentukan peneliti. Hal terpenting pada analisis *survival* adalah memodelkan waktu kegagalan yang memiliki korelasi dengan variabel independen. Untuk menentukan besarnya hubungan antara variabel independen dengan variabel dependennya digunakan model regresi Cox. Waktu kegagalan dapat didefinisikan sebagai waktu dari awal observasi hingga terjadinya kejadian, dapat dalam hari, bulan dan tahun [5].

Regresi Cox pada umumnya digunakan untuk regresi data *survival*. Secara umum model regresi Cox dihadapkan pada situasi dimana kemungkinan kegagalan individu pada suatu waktu yang dipengaruhi oleh satu atau lebih variabel independen. Model ini tidak didasarkan pada asumsi-asumsi tentang sifat atau bentuk distribusi yang mendasari *survival* dan model ini sangat populer karena fungsi *baseline hazard* pada model tidak ditentukan, merupakan pengestimasi koefisien regresi yang baik dan kurva *survival* biasa untuk situasi data yang bervariasi dapat diperoleh dengan model ini. Selain itu regresi Cox dikatakan juga sebagai model *robust*, yaitu hasil dari model Cox ini hampir sama dengan hasil menggunakan model parametrik [5]. Regresi Cox merupakan pendekatan model matematika yang digunakan untuk mengestimasi kurva *survival* ketika mempertimbangkan beberapa variabel independen secara serentak [5]. Variabel-variabel ini merupakan kovariat yang dikenal dengan faktor risiko yaitu faktor yang diestimasi mempengaruhi waktu *survival*.

Dalam penelitian ini akan digunakan data mengenai ketahanan hidup penderita Tuberkulosis (TB) di Yogyakarta dengan

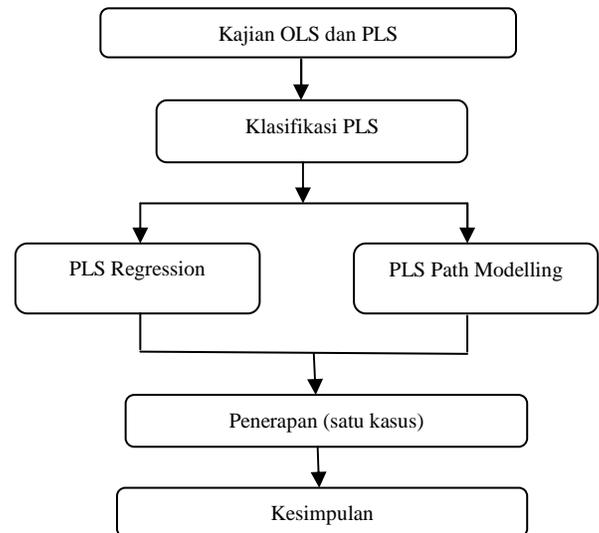
mengambil sampel data dari sebuah rumah sakit di Yogyakarta. Tuberkulosis adalah penyakit infeksi yang disebabkan oleh bakteri *Mycobacterium tuberculosis*. Penyakit ini paling sering menyerang paru-paru walaupun pada sepertiga kasus menyerang organ tubuh lain dan ditularkan orang ke orang. Keterlambatan dalam menegakkan diagnosa dan ketidakpatuhan dalam menjalani pengobatan mempunyai dampak yang besar karena pasien tuberkulosis akan menularkan penyakitnya pada lingkungan, sehingga jumlah penderita semakin bertambah. Pengobatan Tuberkulosis berlangsung cukup lama yaitu setidaknya 6 bulan pengobatan dan selanjutnya dievaluasi oleh dokter apakah perlu dilanjutkan atau berhenti, karena pengobatan yang cukup lama seringkali membuat pasien putus berobat atau menjalankan pengobatan secara tidak teratur. Kedua hal ini fatal akibatnya, yaitu pengobatan tidak berhasil dan kuman menjadi kebal disebut *Multi Drugs Resistance* (MDR)[6].

Metode Penelitian

Metode Penelitian ini terdiri dari dua macam kegiatan, yaitu mengkaji topik PLS yang relevan dengan referensi-referensi yang telah berkembang dan menyajikan contoh penerapan PLS. Penerapan PLS akan disesuaikan dengan kecocokan analisis dengan data yang diperoleh. Pada penelitian ini digunakan data ketahanan hidup penderita TB di RS Yogyakarta, sehingga penerapan PLS digunakan pada analisis regresi Cox data yang diperoleh adalah data survival. Data diambil dari penelitian Dewi Susanti, Mahasiswa S1 Jurusan Pendidikan Matematika-Program Studi Matematika angkatan 2009 yang mengambil data dari sebuah Rumah Sakit Swasta di Yogyakarta.

Langkah-langkah dalam pembentukan regresi PLS pada regresi Cox analog dengan regresi linier. Terlebih dahulu akan ditunjukkan apakah ada pelanggaran asumsi khususnya multikolinieritas pada data. Sehingga pemilihan penerapan PLS dimungkinkan tepat. Hasil akhir model yang diperoleh melalui analisis Regresi

Cox dengan PLS akan dibandingkan dengan bila digunakan regresi Cox biasa.



Gambar 1. Untuk mendapatkan gambaran PLS akan dikaji terlebih dahulu OLS. Selanjutnya ditelusuri PLS klasifikasi regresi dan PLS untuk analisis jalur. Selanjutnya diharapkan dapat diberikan contoh penerapannya.

Hasil dan Diskusi

Karena keterbatasan tim peneliti dan waktu penelitian maka pada penelitian ini difokuskan pada penerapan PLS pada analisis regresi. Peneliti menentukan penerapan pada analisis regresi cox karena disesuaikan dengan data yang diperoleh yaitu data survival. Sehingga hasil akhir, peneliti belum membahas secara detil mengenai PLS untuk SEM.

PLS-R

Beberapa asumsi-asumsi seperti normalitas, linearitas, dan homoskedastisitas multikolinieritas perlu diperhatikan ketika melakukan analisis regresi. Diperlukan penanganan terhadap masalah multikolinieritas sehingga tidak menimbulkan salah penafsiran saat menginterpretasikan output yang dihasilkan.

Beberapa pendekatan untuk mengatasi masalah ini dapat dicoba untuk mendapatkan kesimpulan yang lebih beralasan, misalnya regresi ridge, metode stepwise, PCR (principal component regression) dan PLS (partial least squares).

Multikolinieritas adalah problem yang sering dijumpai saat melakukan regresi, sehingga perlu dilakukan pendekatan lain agar tidak menghasilkan interpretasi model ataupun koefisien regresi yang tidak tepat dan mungkin saja kesalahan pengambilan keputusan. Karena umumnya diambil tidakan membuang variabel yang saling berkorelasi cukup tinggi, padahal kenyataannya variabel tersebut cukup berpengaruh terhadap variabel responnya. Selain karena adanya korelasi yang cukup tinggi antar variabel independennya, multikolinieritas dapat juga disebabkan karena jumlah observasi yang relatif kecil dengan variabel independen yang cukup banyak.

PLS regresi sering digunakan pada bidang *chemometrics*, *bioinformatics*, *sensometrics*, *neuroscience*, dan *anthropology*. Analisis regresi adalah analisis hubungan antara dua variabel yaitu variabel dependen dan variabel independen. Regresi linier klasik yang dikenal secara umum adalah regresi dengan variabel dependen berupa data kontinu sedangkan variabel independen dapat berupa data kuantitatif maupun kualitatif. PLS yang diterapkan untuk regresi ini dikenal sebagai PLS-R. Sedangkan PLS yang diterapkan untuk variabel dependen yang berupa data kategorik dikenal sebagai PLS-DA (discriminant analysis). Sedangkan dari [2] dikenal PLS-GLR untuk semua PLS yang dikembangkan berdasarkan cara kerja GLR. Secara ringkasnya algoritma PLS-GLR:

1. Komputasi m komponen PLS t_h ($h = 1, 2, \dots, m$)
2. GLR dari y pada m komponen PLS yang digunakan
3. Transformasi komponen PLS ke variable aslinya.

Jika terdapat sejumlah p variabel independen dan sebuah variabel dependen/respon, dalam proses PLS, diasumsikan semua variabel sudah dalam bentuk baku/standard.

Model regresi PLS dengan m komponen dirumuskan sebagai :

$$Y = \sum_{h=1}^m c_h \left(\sum_{j=1}^p w_h^* x_j \right) + \text{sisal} \quad (1)$$

Perhitungan komponen PLS pertama, $t_1 = Xw_1^*$ didefinisikan sebagai

$$t_1 = \frac{1}{\sqrt{\sum_{j=1}^p \text{cov}(y, x_j)^2}} \sum_{j=1}^p \text{cov}(y, x_j)^2 x_j \quad (2)$$

variabel x_j ini dipilih yang berkorelasi tinggi dengan y dan cukup kuat variabilitasnya. Selanjutnya untuk koefisien regresi a_{1j} dapat digunakan untuk menaksir seberapa penting variabel x_j dalam pembentukan t_1 . Regresi sederhana y terhadap x_j dirumuskan:

$$Y = a_{1j} x_j + \text{sisal} \quad (3)$$

Jika a_{1j} tidak signifikan atau tidak berbeda nyata dengan 0 maka dapat diabaikan hubungan variabel independennya.

Perhitungan komponen PLS kedua, t_2 Komponen pls kedua, t_2 didefinisikan sebagai

$$t_2 = \frac{1}{\sqrt{\sum_{j=1}^p \text{cov}(y_1, x_{1j})^2}} \sum_{j=1}^p \text{cov}(y_1, x_{1j}) x_{1j} \quad (4)$$

dimana sebelumnya dilakukan dua hal yaitu :

1. regresi sederhana y terhadap setiap x_j
2. regresi x_j terhadap t_1

$$Y = c_1 x_1 + y_1 \quad (5)$$

$$X_j = \rho_{1j} t_1 + x_{1j} \quad (6)$$

komponen PLS kedua dapat juga dituliskan sebagai

$$t_2 = \frac{1}{\sqrt{\sum_{j=1}^p \text{cov}(y_1, x_j | t_1)^2}} \sum_{j=1}^p \text{cov}(y_1, x_j | t_1) x_{1j} \quad (7)$$

karena korelasi parsial antara y dan x_j jika diketahui t_1 didefinisikan sebagai korelasi antara residual y_1 dan x_{1j} maka kovariansi parsial antara y dan x_j diketahui t_1 juga didefinisikan sebagai kovariansi antara residu y_1 dan x_{1j} .

$$\text{Cov}(y, x_j | t_1) = \text{cov}(y_1, x_{1j}) \quad (8)$$

untuk melihat kontribusi x_j dalam pembentukan t_2 , dapat diketahui melalui regresi y terhadap t_1 dan x_j .

$$Y = c_{1j} t_1 + a_{2j} x_j + \text{residu} \quad (9)$$

Sedangkan uji koefisien regresi a_{2j} dapat digunakan untuk menaksir seberapa penting variabel x_{1j} dalam pembentukan t_2 . Jika tidak signifikan maka hubungan variabel independennya tidaklah penting dalam pembentukan komponen PLS kedua tersebut.

Perhitungan komponen PLS berikutnya dan aturan penghentiannya. Dengan prosedur yang sama seperti saat mencari t_2 maka dapat dilanjutkan pencarian komponen PLS ke- h ,

$$t_h = XW_h^*$$

Pencarian komponen baru berhenti jika semua kovariansi parsialnya tidak signifikan.

Persamaan regresi partial least square

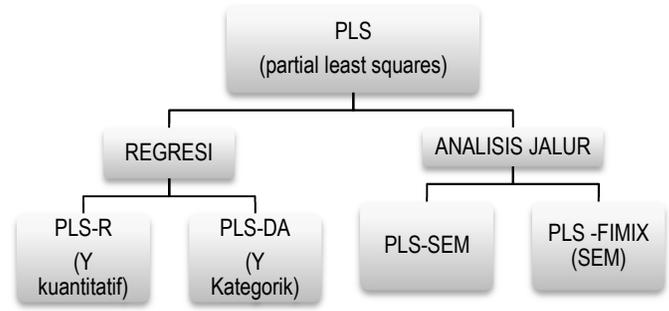
Dalam (1) koefisien c_h diestimasi melalui regresi dari y terhadap komponen PLS t_h . Persamaan regresi estimasinya selanjutnya dapat ditulis ke dalam variabel x_j yang asli.

$$\begin{aligned} \hat{y} &= \sum_{h=1}^m c_h \left(\sum_{j=1}^p w_{hj}^* x_j \right) \\ &= \sum_{j=1}^p \left(\sum_{h=1}^m c_h w_{hj}^* \right) x_j \\ &= \sum_{j=1}^p b_j x_j \end{aligned}$$

Klasifikasi PLS

Berdasarkan pembahasan PLS pada regresi dan analisis jalur, secara umum PLS dapat diklasifikasikan menjadi dua bagian yaitu PLS-R dan PLS-PM. PLS-PM atau PLS SEM masih bisa diklasifikasikan menjadi beberapa pengembangan. Demikian juga untuk PLS-R. menurut [2], PLS-R atau PLS-GLR karena dapat dikembangkan untuk GLR seperti regresi logistik, regresi cox dan keluarga eksponensial dalam model GLM. Dalam hal ini PLS diklasifikasikan menjadi dua kategori yaitu untuk variabel dependen Y kuantitatif dan Y kualitatif. Penerapan

PLS untuk variabel dependen kualitatif/kategorik disebut PLS-DA (PLS discriminant analysis). Berdasarkan bahasan tersebut berikut ini diagram klasifikasi PLS.



Gambar 1. Klasifikasi PLS

Penerapan

Untuk menggambarkan bagaimana PLS diterapkan maka dalam penelitian ini diberikan salah satu contoh kasus. Dalam penelitian ini dipilih regresi cox. Data yang digunakan dalam penelitian ini adalah data pasien rawat inap penderita penyakit tuberkulosis sejak Januari 2012 sampai Desember 2012 yang diambil dari bagian rekam medik Rumah Sakit PKU Muhammadiyah Yogyakarta. Populasi penderita penyakit tuberkulosis yang dirawat inap di Rumah Sakit PKU Muhammadiyah Yogyakarta sejak Januari 2012 sampai Desember 2012 sebanyak 50 orang.

Hasil analisis regresi Cox

Berdasarkan data waktu ketahanan hidup dari 50 pasien penderita tuberkulosis yang pernah dirawat inap di RS PKU Muhammadiyah Yogyakarta pada bulan Januari – Desember 2012, selanjutnya akan dimodelkan dengan menggunakan regresi Cox.

Dengan hipotesis bahwa semua variabel independen berpengaruh terhadap status hidup pasien penderita TB, dengan t sebagai variabel waktu adalah lama perawatan, berikut ini hasil olah data menggunakan PASW Statistics 18.

Variabel-variabel yang diduga mempengaruhi dimasukkan ke dalam model sekaligus dengan menggunakan metode Enter diperoleh nilai -2 Log Likelihood pada model saat semua koefisien $\beta = 0$ adalah 305,982, sedangkan untuk model dengan variabel independen dimasukkan didapatkan nilai -2 Log Likelihood adalah 281,748, dengan nilai $\chi^2 = 24,498$ dan nilai $P=0,079$. Nilai $P>0,05$, hal ini bermakna bahwa model yang dibuat kurang baik. Model yang baik akan mempunyai nilai $P<0,05$.

Tabel 1. Output Signifikansi Model

-2 L Likelihood	Overall (score)		
	Chi-square	df	Sig.
281, 748	24,498	16	,079

Berikut ini merupakan hasil analisis ketahanan dengan model *Cox Proportional Hazard* untuk delapan kovariat.

Tabel 2. Output Koefisien dan signifikansi masing-masing variabel independen.

	B	SE	Wald	Df	Sig.
Gender	-,975	,446	4,765	1	,029
Usia			12,66	4	,013
usia(1)	,702	1,292	,296	1	,587
usia(2)	-,960	1,219	,621	1	,431
usia(3)	-	,776	2,214	1	,137
usia(4)	1,445	,833	3,010	1	,083
Status	-,649	,694	,876	1	,349
pendidikan			3,776	2	,151
pendidikan(1)	1,597	1,081	2,181	1	,140
pendidikan(2)	-,175	,852	,042	1	,837
diagnosis_pasi n	-,429	,884	,235	1	,628
status_penyakit	-,206	,440	,219	1	,640
kelas_rawat			13,43	3	,004
kelas_rawat(1)	1,636	,694	5,555	1	,018
kelas_rawat(2)	2,202	,807	7,442	1	,006

kelas_rawat(3)	1,931	,574	11,30	1	,001
pekerjaan			5,548	3	,136
pekerjaan(1)	1,053	,741	2,018	1	,155
pekerjaan(2)	2,416	1,035	5,444	1	,020
pekerjaan(3)	1,189	,927	1,646	1	,200

Berdasarkan tabel koefisien, variabel jenis kelamin signifikan pada taraf 5% dengan nilai $P < 0,05$ dengan nilai dugaan parameter tersebut negatif. Variabel usia signifikan pada taraf 5% terlihat bahwa nilai $P < 0,05$. Walaupun nilai dugaan parameter untuk kategori usia 1 – 14 tahun positif tetapi tidak berpengaruh nyata. Nilai dugaan parameter untuk kategori usia 15 – 24 tahun negatif tetapi tidak berpengaruh nyata. Nilai dugaan parameter untuk kategori usia 25 – 34 tahun negatif tetapi tidak berpengaruh nyata. Sedangkan nilai dugaan parameter untuk kategori usia 35 – 44 tahun positif tetapi tidak berpengaruh nyata.

Variabel status pernikahan tidak signifikan pada taraf 5% terlihat bahwa nilai $P > 0,05$. Variabel pendidikan tidak signifikan pada taraf 5% terlihat bahwa nilai $P > 0,05$. Variabel diagnosis penyakit pasien dan kondisi penyakit tidak signifikan pada taraf 5%, terlihat dengan nilai $P > 0,05$.

Variabel ruang perawatan pasien signifikan pada taraf 5% dengan nilai $P < 0,05$. Nilai dugaan parameter untuk kategori ruang kelas utama positif namun berpengaruh nyata. Hal ini juga terlihat pada nilai $\text{Exp}(B)$, yang berarti risiko meninggal untuk individu dengan ruang perawatan kelas utama 5,137 kali dari individu dengan ruang perawatan kelas III. Walaupun nilai dugaan parameter untuk kategori ruang kelas I positif namun berpengaruh nyata. Hal ini juga terlihat pada nilai $\text{Exp}(B)$, yang berarti risiko meninggal untuk individu dengan ruang perawatan kelas I 9,046 kali dari individu dengan ruang perawatan kelas III. Nilai dugaan parameter untuk kategori ruang kelas II positif namun berpengaruh nyata. Hal ini juga terlihat pada nilai $\text{Exp}(B)$, yang berarti risiko meninggal untuk individu dengan ruang perawatan kelas II 6,896 kali dari individu dengan ruang perawatan kelas III.

Variabel pekerjaan pasien tidak signifikan pada taraf 5% dengan nilai $P > 0,05$. Nilai dugaan parameter untuk kategori tenaga produksi

atau kasar adalah positif namun berpengaruh nyata. Hal ini juga terlihat pada nilai $\text{Exp}(B)$, yang berarti risiko meninggal untuk individu dengan tenaga produksi atau kasar 11,202 kali dari individu dengan pekerjaan profesional. Nilai dugaan parameter untuk kategori tenaga usaha penjualan dan pertanian positif namun tidak berpengaruh nyata. Hal ini juga terlihat pada nilai $\text{Exp}(B)$, yang berarti risiko meninggal untuk individu dengan pekerjaan tenaga usaha penjualan dan pertanian 3,285 kali dari individu dengan pekerjaan profesional.

Dari output tersebut dapat disimpulkan bahwa variabel-variabel yang berpengaruh nyata terhadap kesembuhan penderita Tuberkulosis yang menjalani perawatan adalah jenis kelamin (X_1), kelas ruang perawatan (X_{7_1} , X_{7_2} dan X_{7_3}) dan pekerjaan (X_{8_2}). Sedangkan variabel-variabel yang pengaruhnya tidak nyata pada taraf 5% adalah usia (X_{2_1} , X_{2_2} , X_{2_3} dan X_{2_4}), status (X_3), pendidikan (X_{4_1} dan X_{4_2}), diagnosis (X_5), kondisi penyakit (X_6) dan pekerjaan (X_{8_1} dan X_{8_3}).

Model yang diperoleh adalah

$$h(t, X) = h_0(t) \exp(-0,975X_1 + 1,636X_{7_1} + 2,202X_{7_2} + 1,931X_{7_3} + 2,416X_{8_2})$$

dengan $h(t, X)$ adalah risiko kematian individu pada t dengan karakteristik X , $h_0(t)$ adalah fungsi *hazard* individu pada keadaan baku $X=0$, tidak tergantung pada karakteristik, X_1 adalah jenis kelamin, X_{7_1} adalah ruang perawatan kelas utama, X_{7_2} ruang perawatan kelas I, X_{7_3} ruang perawatan kelas II dan X_{8_2} adalah pekerjaan tenaga produksi atau kasar.

Pembahasan

Terdapat kontradiksi antara kesimpulan saat uji kecocokan model dengan uji parsial masing-masing variabel independen. Hasil uji hipotesis kecocokan model dinyatakan kurang baik tetapi uji parsial menyimpulkan ada beberapa variabel yang berpengaruh secara signifikan. Perbedaan interpretasi hasil ini mengindikasikan adanya pelanggaran asumsi. Selanjutnya dilakukan pemeriksaan korelasi antar variabel independen. Dari output korelasi di bawah ini menunjukkan adanya korelasi yang cukup tinggi antar beberapa variabel independennya. Sehingga dimungkinkan

terjadinya pelanggaran asumsi multikolinearitas. Oleh karena itu selanjutnya data diolah menggunakan metode PLS untuk pendugaan model regresi cox.

Tabel 3. Output Correlations

	JK	usia	status	Pendidikan	diag_pasien	status_penyakit	kelas_rawat	pekerjaan
JK	1	-.265	.161	.003	.036	.004	.265	-.223
usia		1	-.741	-.319	-.285	.176	-.183	.327
status			1	.202	.037	-.234	.017	-.198
pendidikan				1	.173	.084	-.295	.348
diag_pasien					1	.054	.071	.058
status_penyakit						1	-.375	.206
kelas_rwt							1	-.435
pekerjaan								1

Langkah –langkah yang dilakukan untuk membentuk komponen PLS yang pertama adalah melakukan regresi cox variabel y dengan satu variabel independen. Karena terdapat 8 kovariat maka dilakukan 8 kali regresi dan diselidiki variabel independen yang berpengaruh secara parsial terhadap variabel y . Ternyata diperoleh hanya satu variabel kelas perawatan yang berpengaruh terhadap y . Kelas perawatan mempunyai 4 kategori level sehingga terdapat 3 dummy variabel. Dari ketiga variabel dummy hanya satu variabel yang signifikan tetapi untuk pembentukan komponen PLS yang pertama akan dibentuk oleh ketiga dummy variabel. Komponen PLS, t_1 adalah

$$t_1 = \frac{-0.24k_1 + 0.292k_2 + 0.852k_3}{0.932}$$

Selanjutnya diselidiki apakah masih ada komponen PLS yang kedua, dengan cara melakukan regresi cox dengan dua variabel independen yaitu t_1 dan masing masing variabel independen. Ternyata diperoleh bahwa kedelapan variabel sudah tidak signifikan berpengaruh terhadap y , sehingga hanya terbentuk satu komponen PLS, yaitu t_1 . Selanjutnya dengan membentuk variabel baru yaitu t_1 yang dibangun dari variabel kelas perawatan sebagai komponen PLS kemudian dilihat signifikansi model regresi dengan melibatkan komponen t_1 tersebut. Berikut ini hasil uji signifikansi modelnya

Tabel 4. Output signifikansi t_1 .

-2 Log Likelihood	Overall (score)		
	Chi-square	Df	Sig.
301,303	5,371	1	,020
	B	SE	
T1	,900	,395	

$$h^*(t, X) = h_0(t) \exp(0.9 t_1)$$

Dari output di atas menunjukkan model regresi baik karena P-value (Sig.) adalah $0.02 < 0.05$. Secara ringkas dapat diinterpretasikan bahwa ketahanan hidup penderita TB yang berobat di sebuah Rumah Sakit akan dipengaruhi oleh faktor kelas perawatan. Pada realitanya ada banyak faktor yang berpengaruh terhadap ketahanan hidup seorang pasien, karena keterbatasan penelitian maka masih banyak faktor yang belum masuk dalam model.

Kesimpulan

Berdasarkan kajian secara umum PLS dapat diklasifikasikan pada dua analisis yaitu regresi dan analisis jalur. PLS pada regresi diklasifikasikan untuk variabel dependen kuantitatif seperti regresi linier berganda dan variabel dependen kualitatif seperti regresi logistik dan regresi cox. PLS pada analisis jalur masih bisa dikategorikan menjadi PLS SEM dan PLS pada SEM yang tersegmentasi yang dikenal sebagai FIMIX PLS.

Hasil Penerapan PLS pada regresi cox diterapkan pada kasus ketahanan hidup data penderita TB dengan diperoleh hanya satu komponen PLS yang terbentuk, yang dibangun dari variabel kelas perawatan.

PUSTAKA

- [1] H. Wold, Partial least squares, Encyclopedia of statistical (1985) 581-591.
- [2] P. Bastien, V. Vinzi, dan M. Tenenhaus. Computational Statistics & Data Analysis (2005) 17-46.
- [3] V. E. Vinzi, Handbook of Partial Least Squares, Springer-Verlag, Berlin 2010.
- [4] C. Preda, G. Saporta, Computational Statistics and Data Analysis 49 (2005) 99–108.
- [5] D. G. Kleinbaum dan M. Klein, Survival Analysis: A Self-Learning Text, 2nd ed., Springer Science Business Medi, Inc., New York, 2005.
- [6] <http://www.klikpdpi.com/konsensus/tb/tb.html> Perhimpunan Dokter Paru Indonesia, 2006, Tuberkulosis, Diambil 21 Maret 2013 pukul 13:17.