

## **PREDIKSI WAKTU KETAHANAN HIDUP DENGAN METODE *PARTIAL LEAST SQUARE***

**PANDE PUTU BUDI KUSUMA<sup>1</sup>, I GUSTI AYU MADE SRINADI<sup>2</sup>,**

<sup>1,2</sup>Jurusan Matematika FMIPA Universitas Udayana, Bukit Jimbaran-Bali,  
e-mail: <sup>1</sup>pande.mathboy@gmail.com, <sup>2</sup>srinadiigustiayumade@yahoo.co.id

### ***Abstract***

*Coronary heart disease is caused due to an accumulation of fat on the inside walls of blood vessels of the heart (coronary arteries). The factors that had led to the occurrence of coronary heart disease is dominated by unhealthy lifestyle of patients, and the survival times of different patients. This research objective is to predict the survival time of patients with coronary heart disease by taking into account the explanatory variables were analyzed by the method of Partial Least Square (PLS). PLS method is used to resolve the multiple regression analysis when the specific problems of multicollinearity and microarray data. The purpose of the PLS method is to predict the explanatory variables with multiple response variables so as to produce a more accurate predictive value. The results of this research showed that the prediction of survival for the three samples of patients with coronary heart disease had an average of 13 days, with a RMSEP value (error value) was 1.526 which means that the results of this study are not much different from the predicted results in the field of medicine. This is consistent with the fact that the medical field suggests that the average survival for patients with coronary heart disease by 13 days.*

**Keywords:** *Survival, Coronary Heart Disease, Multicollinearity, Microarray Data, Partial Least Square*

### **1. Pendahuluan**

Analisis survival merupakan alat yang digunakan untuk menganalisis data yang berhubungan dengan kondisi kegagalan atau pada waktu dari kondisi awal (*time origin*) hingga berakhirnya kejadian atau titik akhir, Allison [1]. Dalam penelitian kesehatan, data pada saat terjadinya kejadian tertentu, seperti kematian pasien, seringkali ditemui. Hal yang diutamakan dalam penelitian bukan hanya hasil kejadian, namun lebih kepada waktu sampai pada keadaan tertentu. Dalam menentukan waktu survival, terdapat tiga faktor yang dibutuhkan, yaitu: 1) Waktu awal suatu kejadian, 2) Waktu kejadian akhir, dan 3) Skala pengukuran sebagai bagian dari waktu harus jelas.

Secara umum, penyakit diklasifikasikan menjadi tiga yaitu penyakit menular yang disebabkan oleh kuman, penyakit tidak menular yang disebabkan

<sup>1</sup> Alumni Mahasiswa Jurusan Matematika FMIPA Universitas Udayana

<sup>2</sup> Staf Pengajar Jurusan Matematika FMIPA Universitas Udayana

karena problem fisiologis maupun metabolisme tubuh dan penyakit kronis yang merupakan penyakit yang diderita sangat lama dan bisa menyebabkan kematian, Bustan [2]. Salah satu penyakit kronis yang menyerang organ tubuh manusia khususnya pada jantung adalah jantung koroner.

Penyakit jantung koroner merupakan penyakit yang pada mulanya disebabkan karena adanya penumpukan lemak pada dinding dalam pembuluh darah jantung (pembuluh koroner), lama kelamaan diikuti berbagai proses, seperti penimbunan jaringan ikat, pengapuran, atau pembekuan darah, yang nantinya akan berakibat mempersempit atau menyumbat pembuluh darah. Hal ini mengakibatkan otot jantung di daerah tersebut mengalami kekurangan aliran darah dan dapat menimbulkan berbagai akibat yang cukup serius, seperti nyeri dada sampai infark jantung, yang dalam masyarakat dikenal dengan serangan jantung yang dapat menyebabkan kematian mendadak.

Penelitian dalam bidang kedokteran menyatakan bahwa faktor yang selama ini menyebabkan timbulnya penyakit jantung koroner didominasi oleh pola hidup yang tidak sehat dari pasien itu sendiri, seperti minum-minuman keras, merokok dan pola hidup tidak sehat lainnya. Jika pasien itu sendiri memiliki faktor keturunan ditambah dengan pola hidup yang tidak sehat maka resiko untuk terkena penyakit jantung koroner lebih besar. Hal inilah yang menyebabkan waktu ketahanan hidup pasien penyakit jantung koroner berbeda-beda sehingga tidak diketahui secara pasti waktu tahan hidupnya (tersensor).

Metode *Partial Least Square* (PLS) digunakan ketika terjadi permasalahan spesifik pada data, seperti ukuran sampel yang akan diteliti kecil (*microarray data*), adanya data yang hilang (*missing value*), terjadinya korelasi antar peubah penjelas (multikolinearitas), terjadinya pendugaan yang bias (*overfitting*) dan data yang digunakan merupakan data yang waktu ketahanan hidupnya tidak diketahui secara pasti (data tersensor), Esposito et all [3], Somnath dan Susmita [5].

Metode regresi linear berganda memiliki persamaan dengan *partial least square* dalam hal memprediksi data. Namun, metode regresi linear berganda tidak bisa digunakan untuk mengatasi multikolinearitas, *overfitting*, maupun *microarray data*, karena jika diselesaikan dengan regresi linear berganda maka nilai *p-value* dan signifikansinya tidak akan diketahui. Selain itu, *Partial Least Square* juga dapat menghasilkan model yang cukup akurat dengan data yang sedikit dan peubah penjelas yang banyak.

Rumusan masalah dari penelitian ini adalah bagaimana memprediksi waktu ketahanan hidup pasien penyakit jantung koroner dengan memperhatikan peubah gen, jenis kelamin, usia, stress, kadar gula dalam darah, tekanan darah, jumlah batang rokok dihisap tiap hari, kolesterol, obesitas, olahraga dan rentang waktu saat pertama kali sakit hingga saat data diambil dalam penelitian (April 2011) dengan menggunakan metode *Partial Least Square* (PLS)?

Penelitian ini bertujuan memprediksi waktu ketahanan hidup pasien penyakit jantung koroner dengan memperhatikan peubah-peubah penjelas, menggunakan metode *Partial Least Square* (PLS).

## 2. Metode Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang diambil dari Rekam Medik, Rumah Sakit Umum Tabanan. Data tersebut adalah data pasien rawat inap yang terkena penyakit jantung koroner bulan April 2011 dengan satu peubah respon (Y) dan 11 peubah penjelas (X). Peubah penjelas dalam penelitian ini meliputi : Keturunan (X1), Jenis Kelamin (X2), Usia/Umur (X3), Stres (X4), Kadar Gula Darah (X5), Tekanan Darah (X6), Jumlah Batang Rokok/Hari (X7), Kolesterol (X8), Obesitas (X9), Olahraga (X10), Waktu Pertama Kali Sakit Sampai Diteliti (X11).

Adapun langkah-langkah untuk menganalisis data menggunakan metode *partial least square* (PLS) adalah sebagai berikut:

1. Memasukkan data pengamatan dalam bentuk matriks pada program R.
2. Menentukan pembagian data menjadi dua kelompok, yaitu kelompok I untuk penentuan komponen yang optimal dan kelompok II untuk validasi.
3. Pada kelompok data I, dilakukan proses penentuan jumlah peubah laten dimana jumlah peubah laten yang dipilih adalah jumlah yang memiliki CV/RMSEP terkecil
4. Mempergunakan kelompok data II untuk melakukan validasi

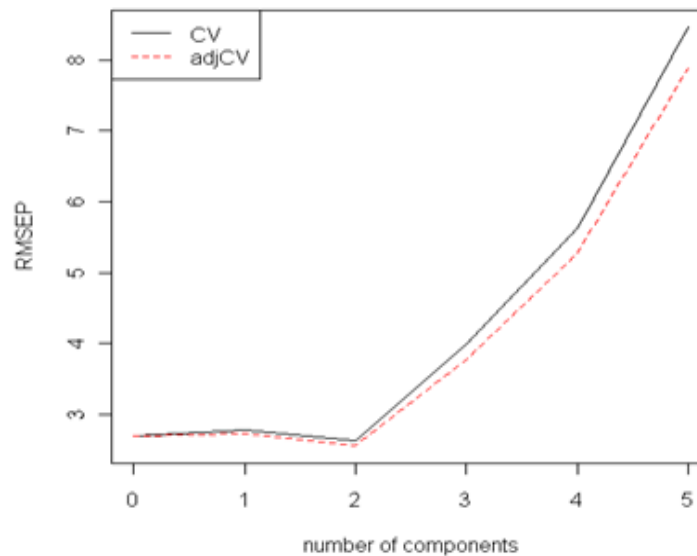
## 3. Hasil dan Pembahasan

Data sepuluh pasien penyakit jantung koroner dalam penelitian ini yang dapat dideskripsikan adalah dominasi umur pasien penyakit jantung koroner berkisar antara 60 sampai 86 tahun, laki-laki lebih banyak terkena penyakit jantung koroner daripada perempuan. Faktor keturunan tidak dominan pada penyakit jantung koroner, melainkan karena pola hidup tidak sehat dari pasien itu sendiri yang menyebabkan pasien terkena penyakit jantung koroner.

Kadar gula dari pasien rata-rata sebesar  $\pm 120$  mg/dl dan hanya dua orang yang memiliki kadar gula tinggi. Tekanan darah dari pasien tidak ada yang normal (120 mg/dl), lima pasien memiliki tekanan darah rendah dan lima pasien lagi memiliki tekanan darah tinggi. Enam orang pasien memiliki kebiasaan merokok (empat perokok sedang, dua perokok berat), dan empat pasien lainnya tidak merokok. Sebagian besar pasien memiliki kadar kolesterol darah lebih dari 200 mg/dl, selain itu juga pasien memiliki tingkat obesitas yang berkisar antara 18 sampai 30 kg/m<sup>2</sup>, hampir sebagian besar pasien berolahraga namun sedikit sekali yang berolahraga setiap hari. Pasien mengetahui bahwa dirinya terkena penyakit jantung koroner semenjak tiga sampai empat bulan yang lalu.

Berdasarkan pengolahan data dengan menggunakan program R 2.13 diperoleh nilai rata-rata pasien penyakit jantung koroner memiliki waktu ketahanan hidup  $\pm 13$  hari dengan standar deviasi sebesar 2,21 dan variansi sebesar 4,9.

Validasi silang (*Cross Validation/CV*) pada *Partial Least Square (PLS)* untuk memvalidasi antara data pengukuran dengan data prediksi, dan hubungannya dengan nilai *Root Mean Square Error of Prediction (RMSEP)* terlihat dalam Gambar 1.



Gambar 1. Plot Validasi Silang RMSEP dengan Lima Komponen

Plot pada gambar 1 menunjukkan ukuran kinerja prediksi (RMSEP) terhadap banyak komponen. Nilai RMSEP minimum menunjukkan banyak komponen yang optimal dalam model, Mevik [4] dan terhindar dari *overfitting* ., RMSEP terkecil terletak pada banyak komponen bernilai dua, dengan nilai RMSEP sebesar 2,554 sehingga model yang optimal digunakan adalah model dengan dua komponen. Adapun nilai varians (keragaman) yang dapat dijelaskan oleh masing-masing komponen secara tegas dapat dilihat pada tabel 1.

Tabel 1. Nilai Varians pada masing-masing Komponen

Comp 1	Comp 2	Comp 3	Comp 4	Comp 5
71.8688	19.01151	7.323782	1.586353	0.204781

Dengan mengambil dua komponen yang masing-masing memiliki nilai keragaman sebesar 72 % untuk komponen satu dan 19 % untuk komponen dua, maka model dengan dua komponen mampu menerangkan 91% keragaman data yang ada.

Model optimal yang diperoleh digunakan untuk memprediksi nilai respon pada data penelitian baru. Berdasarkan model optimal dengan dua komponen yang diperoleh dari hasil analisis sebelumnya, diperoleh nilai prediksi waktu ketahanan hidup untuk tiga pasien (responden baru) seperti pada tabel 2.

Tabel 2. Nilai Prediksi Waktu Ketahanan Hidup Pasien Penyakit Jantung Koroner

No	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	Y	Predicted
1	1	0	80	4	74	100	13	245	18	4	119	13	13.2615
2	0	0	75	1	105	80	0	180	22	0	76	11	13.4378
3	0	0	42	7	103	100	8	245	29	2	187	13	12.0146

Nilai RMSEP untuk tiga responden baru ini dapat dilihat pada tabel 3.

Tabel 3. Nilai RMSEP untuk Nilai Respon pada Responden Baru

(Intercept)	1 comps	2 comps	3 comps	4 comps	5 comps
1,672	1,042	1,526	2,175	2,150	4,603

Berdasarkan tabel 3, dengan dua komponen optimal dalam model tingkat kesalahan (RMSEP) dalam penelitian ini adalah sebesar 1,526. Nilai ini menunjukkan bahwa nilai ketahanan hidup dari model yang telah ditetapkan tidak jauh berbeda dengan nilai ketahanan hidup dari data sebenarnya.

#### 4. Kesimpulan

Hasil penelitian menunjukkan prediksi waktu ketahanan hidup pasien penyakit jantung koroner untuk tiga responden rata-rata sebesar  $\pm 13$  hari, dengan nilai RMSEP (nilai kesalahan) sebesar 1,526. Hal ini sesuai dengan kenyataan dibidang kedokteran yang menunjukkan bahwa rata-rata ketahanan hidup pasien jantung koroner sebesar 13 hari.

#### Daftar Pustaka

- [1] Allison, P. 1995. *Survival Analysis Using SAS*. SAS institute. United State of America
- [2] Bustan. 2007. *Epidemologi Penyakit Tidak Menular*. Jakarta: PT. Asdi Mahasatya
- [3] Esposito, V., Wynee, Henseler, J., and Wang, Huiwen. 2010. *Handbook of Partial least Squares*. Germany: Springer
- [4] Mevik, H. 2007. "The PLS Package: Principal Component and Partial Least Square Regression in R". *Journal of Statistical Software*, vol 18, 1-24. Norwegian.
- [5] Somnath, D and Susmita, D. 2007. "Predicting Patient Survival from Microarray Data by Accelerated failure Time Modeling using partial Least Squares and LASSO". *Journal of Biometrics*, vol 63, 259-271. United State of America