



Catatan Kelas

Pemilihan Model Terbaik

Pemilihan Model Terbaik

Model regresi terbaik adalah model yang dapat menjelaskan perilaku variabel terikat dengan sebaik-baiknya dengan memilih variabel-variabel bebas dari sekian banyak variabel bebas yang tersedia dalam data. Untuk menentukan variabel bebas mana yang akan dimasukkan ke dalam model regresi, menurut Draper ada dua kriteria yang saling bertentangan yaitu :

1. Agar persamaannya bermanfaat untuk peramalan, biasanya ingin dimasukkan sebanyak mungkin peubah sehingga diperoleh nilai ramalan yang andal
2. Karena untuk memperoleh informasi dari banyak peubah serta pemantauannya seringkali diperlukan biaya yang tinggi, maka diinginkan persamaan regresi yang mencakup sesedikit mungkin peubah. Kompromi diantara kedua kriteria itulah yang disebut pemilihan model regresi terbaik

Masalah yang sering muncul dalam regresi berganda adalah adanya hubungan linear antara peubah bebas. Kondisi ini disebut masalah multikolinearitas. Adanya masalah multikolinearitas ditunjukkan oleh nilai Variance Inflation Factors (VIF) yang tinggi. Jika ada masalah multikolinearitas maka kesimpulan

yang didapat dari hasil pengujian untuk model regresi maupun untuk masing-masing peubah yang ada dalam model, seringkali tidak tepat. Oleh sebab itu, masalah multikolinearitas harus dihindari. Untuk pemilihan model regresi terbaik ada beberapa metode yang biasa digunakan dalam mencari model terbaik.

Stepwise

Regresi Stepwise merupakan sebuah metode yang dapat digunakan dalam menentukan model regresi berganda. Prosedur yang dilakukan yaitu dengan menyusupkan satu demi satu variabel bebas sampai diperoleh persamaan regresi yang memuaskan. Urutan penyisipan ditentukan dari koefisien korelasi (r) (Draper & Smith, 1992). Jadi, variabel bebas yang mempunyai korelasi terbesar dengan variabel terikat akan masuk pertama kali dalam model. Tahap selanjutnya adalah memasukkan variabel bebas yang memiliki korelasi terbesar kedua dengan variabel terikat. Tahap ini terus dilakukan sampai semua variabel masuk dalam model. Namun setelah diperoleh satu variabel bebas yang masuk dalam model perlu dilakukan pengujian keberartian variabel dalam model.

Hipotesis :

$H_0 : \beta_i = 0$, dimana variabel yang masuk dalam model tidak berarti

$H_1 : \beta_i \neq 0$, dimana variabel yang masuk dalam model berarti.

Untuk suatu $i = 1, 2, 3, \dots$

Pengujian dilakukan dengan uji-t atau melihat P-Value dengan daerah penolakan $t_{hitung} > t_{tabel} (t(\alpha, n-k-1))$ atau $P - Value < \alpha$

dimana α merupakan taraf nyata yang digunakan ($\alpha = 0.05$).

Apabila hasil pengujian ternyata menolak H_0 maka variabel tersebut tetap berada dalam model dan apabila sebaliknya maka variabel harus keluar dari model. Dalam metode Stepwise, apabila ada dua variabel bebas yang saling berkorelasi, maka hanya ada satu variabel yang masuk dalam model. Pemilihan variabel yang masuk didasarkan pada koefisien korelasi terbesar terhadap variabel terikat (Iriawan, 2006).

Akses data : <https://bit.ly/dataipmdunia>

Syntax R

```
library(readxl)
```

```
library(olsrr)
```

```

data_ipm_ <- read_excel("C:/Users/user/Downloads/data_ipm .xlsx")
View(data_ipm_)
model<-lm(Y~X1+X2+X3+X4+X5,data=data_ipm_)
summary(model)
cor(data_ipm_[3:8],data_ipm_$Y)

#melakukan stepwise
sw.hc=ols_step_both_p(model,details = TRUE)
print(sw.hc)

```

Best Subset Regression

Pada setiap Best Subset Regression ditampilkan berdasarkan beberapa kriteria yaitu :

- a. Nilai R-square (ukuran akurasi model) terbesar
- b. Nilai Adjusted R-square (ukuran akurasi model yang disesuaikan dengan prediktor signifikan) terbesar
- c. Nilai Cp Mallows (besarnya bias) terkecil
- d. Nilai AIC (estimator kesalahan prediksi) terkecil
- e. Nilai S (sisa regresi) terkecil

Metode Best Subset Regression ini merupakan salah satu cara untuk mengatasi multikolinearitas yaitu dengan memilih variabel-variabel bebas yang dapat digunakan dalam model agar diperoleh persamaan regresi terbaik yang mengandung sebagian atau seluruh peubah bebas.

Syntax R :

```

#melakukan subset
bs.hc=ols_step_best_subset(model)
bs.hc

```

Forward Selection

Forward selection adalah metode pembentukan model regresi terbaik berdasarkan pendekatan dengan cara melakukan pemilihan variabel dengan memasukkan variabel bebas yang memiliki nilai korelasi terbesar. Setelah itu, dilakukan pemeriksaan nilai F model yang terbentuk. Jika signifikan, maka variabel dengan nilai korelasi terbesar selanjutnya masuk ke dalam model. Perulangan selesai jika semua variabel prediktor

sudah masuk ke dalam model atau nilai F model yang terbentuk sudah tidak signifikan (Kurniawan dan Yuniarto, 2016). Kelebihan dari metode forward selection adalah sebagai alternatif untuk mengurangi kemungkinan adanya multikolinieritas dalam model yang terbentuk. Di sisi lain, metode ini juga memiliki kekurangan yaitu terdapat kemungkinan adanya lebih banyak variabel prediktor yang tidak begitu signifikan.

Syntax:

```
foward.selection=ols_step_forward_p(model)
```

```
foward.selection
```

Backward Selection

Backward elimination adalah salah satu cara dalam mendapatkan model regresi terbaik. Dalam metode backward elimination, seluruh variabel prediktor dimasukkan ke dalam model lalu secara bertahap variabel-variabel yang tidak memenuhi syarat kelayakan dieliminasi dari model hingga terbentuk model terbaik dengan variabel-variabel yang telah lolos uji di dalamnya (Kurniawan dan Yuniarto, 2016). Kelebihan dari metode backward selection adalah lebih hemat waktu karena di dalam backward elimination hanya perlu memperhatikan satu variabel saja, yaitu nilai dari F parsial terkecil dalam model (Kurniawan dan Yuniarto, 2016). Kelemahan dari metode backward selection adalah ketika suatu variabel prediktor dikeluarkan dari dalam model, maka variabel tersebut tidak dapat dipertimbangkan atau digunakan kembali.

Syntax:

```
backward.selection=ols_step_backward_p(model)
```

```
backward.selection
```