

## ABSTRAK

Jika diberikan data pengamatan independen  $\{(X_i, Y_i)\}_{i=1}^n$  yang diterapkan pada model  $Y_i = g(X_i) + e_i$ ,  $i=1,2,\dots,n$  dengan  $g$  adalah fungsi regresi yang belum diketahui dan  $e_i$  adalah variabel random independen dengan mean 0 dan varian  $\sigma^2$ , maka fungsi  $g$  untuk model rancangan acak dapat diestimasi dengan metode kernel menggunakan Penduga Nadaraya Watson. Pada estimasi fungsi dengan Penduga Nadaraya Watson, yang paling dominan menentukan tingkat kemulusan estimator adalah nilai bandwidth. Nilai bandwidth yang kecil memberikan estimasi fungsi yang sangat tidak mulus, sedangkan nilai bandwidth yang besar memberikan estimasi fungsi yang sangat mulus. Oleh karena itu perlu dipilih nilai bandwidth optimal untuk menentukan estimasi fungsi yang optimal, yakni dengan menggunakan metode cross validasi dengan estimasi leave-one-out. Dengan metode ini dijamin uji signifikansi dengan menggunakan metode bootstrap pivot bersarang terpenuhi, sehingga dapat diketahui pengaruh  $X$  terhadap  $Y$ .

Kata kunci : Metode kernel, penduga nadaraya watson, bootstrap pivot bersarang.

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Dalam analisis statistik, model regresi mempunyai peranan penting, yaitu untuk menggambarkan hubungan antara variabel respon  $Y$  dengan satu atau lebih variabel prediktor  $X$ . Bentuk dari fungsi memberikan penjelasan bahwa nilai pengamatan dari  $Y$  diperoleh dari suatu harga  $X$ . Jika terdapat  $n$  data  $(X_i, Y_i)$  dengan  $i=1,2, \dots, n$  maka model regresi :

$$Y_i = g(X_i) + e_i \quad , i=1,2, \dots, n$$

dengan  $X_i$  adalah variabel prediktor ke- $i$ ,  $Y_i$  merupakan variabel respon ke- $i$  dan  $g$  adalah fungsi regresi. Sementara  $e_i$  diasumsikan sebagai variabel random independen dengan mean 0 dan varian  $\sigma^2$ .

Ada dua pendekatan yang dapat dilakukan dalam mengestimasi fungsi regresi  $g$  yakni pendekatan parametrik dan pendekatan nonparametrik. Pendekatan parametrik dilakukan jika asumsi bentuk  $g$  diketahui tergantung dari suatu parameter, sedangkan pendekatan nonparametrik menghubungkan variabel respon  $Y$  dengan satu atau lebih variabel prediktor  $X$  tanpa diketahui asumsi bentuk dari fungsi  $g$ . Kegunaan paling sederhana dari regresi nonparametrik adalah dalam pemulusan *scatterplot*. Dalam hal ini, ada sebuah respon  $Y$  dan prediktor  $X$  dan digambarkan secara visual hubungan antara dua variabel tersebut dalam sebuah *scatterplot*. Dengan teknik regresi nonparametrik mengasumsikan bahwa fungsi  $g$  termuat dalam fungsi kelas mulus, artinya mempunyai turunan yang kontinyu.

Ada banyak metode nonparametrik untuk mengestimasi fungsi regresi  $g$  dengan asumsi bahwa error adalah independen berdistribusi normal dengan mean 0 dan varian  $\sigma^2$ . Salah satunya adalah Metode Kernel. Pada metode kernel pemilihan nilai parameter pemulus yang kecil menghasilkan estimasi kurang mulus sehingga pemilihan nilai parameter pemulus besar akan menghasilkan estimasi mulus. Parameter pemulus dalam metode kernel disebut *bandwidth*. Untuk mendapatkan bandwidth optimal akan digunakan *Metode Cross Validation* yang kemudian akan dicari estimasi fungsi  $g$ .

Estimasi fungsi  $g$  tergantung distribusi data pengamatan  $X$ . Apabila  $X$  non random maka modelnya adalah model rancangan tetap/ *fixed design* dengan error observasi berdistribusi independen dan identik yakni normal dengan mean 0 dan varian  $\sigma^2$ . Apabila  $X$  random maka modelnya adalah model rancangan acak/ *random design*, dengan  $(X_i, Y_i)$  independen dan  $g(x) = E(Y|X = x)$  dan  $e_i = Y_i - g(X_i)$  (Hardle, 1993).