BAB I

PENDAHULUAN

Analisa regresi merupakan suatu teknik statistika untuk menyelidiki dan membuat model hubungan antara variabel tidak bebas (response) dengan satu atau lebih variabel bebas (predictor). Dan selanjutnya dengan model regresi yang didapatkan, dapat digunakan untuk memprediksi $\hat{Y}$. Tetapi sebelum model regresi digunakan, perlu untuk diketahui seberapa baik model tersebut. Untuk itu diperlukan pemilihan variabel dalam model agar diperoleh model regresi yang terbaik.

Cross validation merupakan metode untuk mengevaluasi model-model regresi dengan suatu ukuran kemampuan prediksi dan memilih satu model yang terbaik. Ukuran kemampuan prediksi dalam metode cross validation dicari dengan cara membagi data menjadi dua bagian. Bagian pertama digunakan untuk membuat model regresi, dan bagian yang kedua digunakan untuk memvalidasi, agar diketahui seberapa baik kemampuan prediksi dari model tersebut. Hal ini dilakukan sampai beberapa kali.

Jika terdapat suatu himpunan data, misalkan (y_i, x_i), $(i = 1, 2, 3, \ldots, n)$ maka dengan metode cross validation diatas, data dibagi menjadi m bagian pertama untuk pembuatan model regresi, dan n-m bagian sisanya untuk memvalidasi model tersebut. Cara lain untuk melakukan

pemecahan data yaitu dengan cara mengambil n-1 data dari n data untuk membuat model regresi, sedangkan satu data sisanya untuk memvalidasi. Hal ini dilakukan sampai n kali.

Dalam analisa regresi bila terdapat jumlah variabel bebas misalnya k, maka dapat dibuat semua kemungkinan regresi yang meliputi satu variabel, dua variabel, tiga variabel, dan seterusnya sehingga terdapat k^2-1 model regresi yang berbeda (dalam hal ini model $Y = \beta_0$ tidak termasuk). Metode cross validation dilakukan pada semua kemungkinan regresi, untuk mendapatkan model regresi yang terbaik.

Dalam skripsi ini, akan dibahas metode cross validation pada persamaan regresi linear berganda. Tujuannya adalah untuk mendapatkan persamaan regresi linear yang terbaik.

Sistematika penulisan skripsi ini sebagai berikut:

Bab I berisi pendahuluan yang menyajikan latarbelakang dari permasalahan, pembatasan masalah, serta maksud dan tujuan dari penulisan.

Bab II mengemukakan teori penunjang yang merupakan dasar-dasar pembuatan model regresi linear, baik regresi linear sederhana maupun regresi linear berganda.

Bab III akan menjelaskan tentang metode cross validation pada persamaan regresi linear, baik regresi linear sederhana maupun regresi linear berganda, serta implementasinya disajikan dalam algoritma. Dan yang

terakhir ditutup dengan Bab IV yang memuat tentang kesimpulan - kesimpulan yang diambil sebagai hasil rangkuman penulisan secara keseluruhan.