

BAB I

PENDAHULUAN

Pengumpulan suatu data pengamatan tidak selalu berjalan dengan mulus, adakalanya terjadi bermacam kendala yang mengakibatkan data menjadi tidak lengkap atau memuat beberapa nilai yang hilang, sehingga menyulitkan pada saat melakukan pengolahan data. Permasalahan data tidak lengkap dapat ditangani dengan distribusi Gaussian (normal) bila semua variabelnya kuantitatif (Schimert dan Schafer, 2001)

Di dalam menghadapi permasalahan data tidak lengkap dapat dibagi ke dalam 3 kelompok (Rubin, 1976; Paul Kofman dan Ian Sharpe, 2000), andaikan X variabel yang seluruhnya teramati untuk variabel tidak lengkap Y, maka nilai hilang pada variabel Y,

- MCAR (*Missing Completely at Random*), untuk nilai hilang pada variabel Y tidak bergantung atas variabel X dan Y, maupun sumber informasi lainnya.
- MAR (*Missing at Random*), untuk nilai hilang pada variabel Y bergantung pada variabel X dan tidak pada Y.
- NI (*Non Ignorable*), Untuk nilai hilang pada variabel Y bergantung hanya pada nilai hilang itu sendiri.

Untuk nilai hilang MCAR, metode yang biasa digunakan adalah *Listwise Deletion* dan *Pairwise Deletion*, di mana kasus yang memuat nilai hilang dihapuskan sedangkan pengolahannya hanya menggunakan kasus dengan nilai yang teramati saja. Sedangkan untuk nilai hilang MAR, metode yang digunakan adalah EM (*Expectation Maximization*) *algorithm* dan *Regression*, dengan pengolahan tidak menghapuskan nilai yang hilang. Keempat metode tersebut ada di dalam program aplikasi SPSS 10.0.

Dalam penulisan tugas akhir ini akan dibahas tentang algoritma EM untuk mencari nilai parameter rata-rata dan variansi pada data tidak lengkap normal bivariat dengan satu variabel teramati penuh dan satu variabel memuat nilai hilang.

Algoritma EM adalah suatu proses dua langkah untuk menaksir parameter suatu model data tidak lengkap. Tindakan awalnya dengan memecah data ke dalam *missing* dan *nonmissing*, kemudian mengestimasi nilai hilang melalui regresi linier sehingga data menjadi lengkap. Di mana regresi linier awal yang digunakan diambil dari regresi linier dalam interval kenormalan pada data yang teramati saja, dengan syarat dapat meningkatkan nilai parameter awal. Pada iterasi selanjutnya estimasi nilai hilang diperoleh dari persamaan regresi linier data lengkap pada iterasi sebelumnya. Langkah akan terus berjalan sampai diperoleh kekonvergenan dari data yang hilang tersebut, sehingga didapatkan nilai parameter yang maksimal. Dasar dari algoritma ini terdiri atas 2 langkah yaitu :

1. Langkah E (*Expectation*), menghitung ekspektasi dari model menggunakan nilai awal parameter.
2. Langkah M (*Maximization*), menggunakan nilai dari ekspektasi pada langkah pertama untuk memaksimalkan fungsi *loglikelihood* dalam memperoleh estimasi parameter baru.

Kedua proses tersebut akan terus berulang sampai diperoleh kekonvergenan.

Tujuan dan kegunaan yang ingin diperoleh di dalam penulisan tugas akhir ini adalah untuk :

1. Mengetahui penerapan metode algoritma EM untuk data tidak lengkap.
2. Mengetahui model yang digunakan pada metode algoritma EM.
3. Memperoleh nilai parameter dengan metode algoritma EM

4. Dapat mengolah data tidak lengkap dengan program aplikasi SPSS 10.0 dan mengambil kesimpulan dari output pengolahannya.

Sistematika penulisan dalam penulisan tugas akhir ini disusun dengan urutan sebagai berikut :

BAB I PENDAHULUAN

Dalam pendahuluan ini berisi latar belakang masalah, perumusan masalah, tujuan dan kegunaan penulisan, serta sistematika penulisan

BAB II DASAR TEORI

Berisi landasan teori dan penelitian terdahulu yang meliputi: barisan dan kekonvergenan; distribusi bivariat, marginal dan ekspektasi bersyarat; distribusi normal; pertidaksamaan Jensen; metode maksimum *likelihood*; dan teorema Bayesian.

BAB III ALGORITMA EM DAN PENERAPANNYA

Berisi tentang permasalahan data tidak lengkap; teori dan penerapan algoritma EM; contoh kasus data tidak lengkap dan analisisnya.

BAB IV KESIMPULAN

Berisi tentang kesimpulan dan saran.