
Model Distribusi *Zero-inflated* Menggunakan *Poisson Mixture* Dengan Algoritma *Expectation-maximization*

Suryo Adi Gumelar¹, Indwiarti², Hilda Fahlana³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹suryoadig@students.telkomuniversity.ac.id, ²indwiarti@telkomuniversity.ac.id,

³indwiarti@telkomuniversity.ac.id

Abstrak

Count data sering digunakan dalam berbagai aspek kehidupan sehari-hari. Distribusi *Poisson* merupakan metode umum untuk memodelkan *count data*. Namun, distribusi *Poisson* tidak memadai untuk menangani data dengan jumlah nol yang tinggi. Oleh karena itu, diperlukan pendekatan tambahan. Dalam penelitian ini, dilakukan pemodelan dengan distribusi *Zero-inflated* menggunakan campuran *Poisson* dan algoritma *Expectation-maximization*. Hasil penelitian menunjukkan bahwa model *Zero-inflated Poisson* mampu memodelkan data dengan jumlah nol yang tinggi secara akurat

Kata kunci: *Count data*, distribusi *Zero-inflated*, *Poisson Mixture*

Abstract

Count data is frequently used in various aspects of daily life. The Poisson distribution is a common method for modeling count data. However, the Poisson distribution is inadequate for handling data with a high number of zeros. Therefore, an additional approach is required. In this study, a Zero-inflated distribution was modeled using a Poisson mixture and the Expectation-maximization algorithm. The results of the study indicate that the Zero-inflated Poisson model can accurately model data with a high number of zeros

Keywords: Count data, Zero-inflated distribution, Poisson Mixture

1. Pendahuluan

Latar Belakang

Model *Poisson* merupakan model yang paling sering digunakan untuk memodelkan data diskrit yang berhubungan dengan banyak aspek dalam kehidupan sehari – hari. Model ini berasumsi bahwa variabel rata – rata dan variansi adalah sama. Dalam hal ini, data sering mengalami overdispersi. Sehingga, data yang teramati sering diabaikan oleh distribusi *Poisson*. Overdispersi ini terjadi karena parameter θ dari distribusi *Poisson* tidak mampu menangkap data secara utuh. Hal ini disebabkan karena heterogenitas dari populasi yang terbagi lagi menjadi banyak subpopulasi yang homogen. Hal lain yang dapat menyebabkan overdispersi terjadi adalah karena terlalu banyak nilai nol yang ditemukan yang melebihi nilai nol yang dihasilkan dari distribusi *Poisson* yang fit. Hal ini dapat diselesaikan menggunakan model *Zero-inflated*. Model ini berasumsi bahwa data diskrit adalah campuran dari dua data yang terpisah: data yang hanya menghasilkan nol dan data yang lain adalah tipe data *Poisson*. Diantara banyak model *zero-inflated* yang digunakan, model yang paling umum digunakan adalah *Zero-inflated Poisson* (ZIP) [1].

Algoritma EM adalah algoritma yang digunakan untuk mengestimasi parameter dari *maximum likelihood* supaya fungsi *likelihood*-nya tercapai/*convergence*. Algoritma EM mengestimasi parameter dari model melalui pemaksimalan batas bawah dari fungsi *likelihood*, yang dikenal dengan *Q-function*, secara berulang. Terdapat beberapa kegunaan dari pengaplikasian algoritma EM. Secara umum, terdapat dua kategori yaitu: data yang hilang dan variabel yang tersembunyi [2].

Tujuan

Tujuan dilakukannya penelitian ini adalah untuk memodelkan distribusi data *Zero-inflated* menggunakan *Poisson Mixture* dengan algoritma *Expectation-maximization* serta menganalisis performansi yang didapat menggunakan metode tersebut.

Organisasi Tulisan

Setelah bab 1 ini, terdapat beberapa bab yang berisi tentang penjelasan lanjut terkait penelitian ini. Bab 2 membahas tentang studi terkait yang sudah pernah dilakukan peneliti lainnya menggunakan metode yang sama. Bab 3 berisikan rancangan sistem yang akan dibangun pada penelitian ini. Bab 4 berisikan tentang evaluasi dari penelitian ini dimana evaluasi tersebut akan membahas hasil pengujian dan analisis hasil pengujian dan bab 5 berisikan kesimpulan yang didapat dari penelitian ini.