

## Regresi Logistik pada Data *Rare Event*

<sup>1</sup>Rundy Rumi Ari Wistara, <sup>2</sup>Suliadi, <sup>3</sup>Abdul Kudus

<sup>1,2,3</sup>*Statistika, Fakultas MIPA, Universitas Islam Bandung,  
Jl. Ranga Malela No. 1 Bandung 40116*

e-mail: <sup>1</sup>rundy\_ra@ymail.com, <sup>2</sup>suliadi@gmail.com, <sup>3</sup>akudus69@ahoo.com

**Abstrak.** Regresi logistik merupakan salah satu metode statistika yang digunakan untuk menganalisis hubungan beberapa faktor dengan sebuah variabel respon. Pada regresi logistik variabel respon terdiri dari dua kategori yaitu “sukses” dan “gagal” yang dinotasikan dengan  $y = 1$  (sukses) dan  $y = 0$  (gagal). Regresi logistik baik digunakan jika persentase  $y = 0$  dan  $y = 1$  tidak jauh berbeda. Dalam kasus kredit macet dapat dijumpai kondisi di mana persentase kredit macet jauh lebih kecil dibandingkan dengan persentase kredit lancar pada variabel respon. Kondisi seperti itu disebut dengan *rare event*. Pada data *rare event* akan menyebabkan  $\Pr(Y = 1)$  *underestimates* sedangkan untuk  $\Pr(Y = 0)$  *overestimates*. Masalah sampel terbatas (*finite sample*) dapat menyebabkan (i) model yang terbentuk akan menghasilkan penaksir parameter yang bias; (ii) kesalahan baku bagi penaksir yang lebih kecil (*underestimates*) dan (iii) dapat menyebabkan  $P(Y = 1)$  *underestimates*. Skripsi ini membahas bagaimana mengoreksi penduga parameter yang bias dan koreksi peluang pada regresi logistik jika data respon jarang terjadi (*rare event*). Model regresi logistik *rare event* akan diterapkan pada data kasus kredit bank di Amerika dengan  $Y = 1$  jika nasabah mengalami kredit macet lebih dari 90 hari. Hasil koreksi bias terhadap koefisien regresi adalah bahwa bias pada  $\hat{\beta}_0$  lebih besar di bandingkan dengan yang lainnya. Namun untuk hasil bias pada keseluruhan taksiran parameter kecil, hal ini karena sampel yang cukup besar yaitu sebanyak 12013. Hasil kesalahan baku penaksir terlihat bahwa kesalahan baku penaksir parameter terkoreksi lebih kecil dibandingkan dengan kesalahan baku pada penaksir parameter regresi logistik. Oleh karena itu regresi logistik pada data *rare event* lebih baik digunakan daripada regresi logistik. Sedangkan untuk koreksi taksiran peluang ( $\hat{\pi}$ ) lebih kecil dibandingkan dengan taksiran peluang jika tidak menggunakan koreksi pada regresi logistik.

**Kata kunci:** Regresi Logistik, *Rare Event*, Bias, Peluang.

### A. Pendahuluan

Dalam kehidupan sehari-hari semua orang pasti memiliki kebutuhan. Kebutuhan ada yang bersifat mendesak dan ada yang tidak. Kebutuhan yang mendesak menuntut untuk segera dipenuhi. Namun pemenuhan tersebut tidak terlepas dari masalah biaya atau dana. Dana yang diperlukan biasanya tidak sedikit jumlahnya, sementara dana yang tersedia acapkali tidak mencukupi.

Kebanyakan orang dalam menghadapi kekurangan dana salah satu jalan keluar yang dapat dilakukan adalah dengan berutang kepada pihak bank. Para nasabah yang telah memperoleh fasilitas kredit dari bank tidak seluruhnya dapat mengembalikan utangnya dengan lancar sesuai dengan waktu yang telah diperjanjikan. Akibat nasabah tidak dapat membayar lunas utangnya, maka akan tergambar perjalanan kredit menjadi macet atau terhenti.

Salah satu metode yang dapat dipergunakan untuk memetakan nasabah ke dalam kategori kredit macet dan lancar yaitu metode regresi logistik. Regresi logistik merupakan salah satu metode statistika yang digunakan untuk menganalisis hubungan antara satu variabel respon ( $Y$ ) dengan satu atau lebih variabel bebas ( $X_i$ ). Dimana variabel respon terdiri dari dua kategori yaitu “sukses” dan “gagal” yang dinotasikan dengan  $Y = 1$  (sukses) dan  $Y = 0$  (gagal). Sebagai contoh pada kasus kartu kredit,  $Y = 0$