

SKRIPSI

PEMODELAN FREKUENSI KLAIM ASURANSI DENGAN  
*GENERALIZED LINEAR MODEL (GLM)*



EVAN FELIX

NPM: 6161901023

PROGRAM STUDI MATEMATIKA  
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS  
UNIVERSITAS KATOLIK PARAHYANGAN  
2024

**FINAL PROJECT**

**MODELING INSURANCE CLAIMS FREQUENCY WITH  
GENERALIZED LINEAR MODEL (GLM)**



**EVAN FELIX**

**NPM: 6161901023**

**DEPARTMENT OF MATHEMATICS  
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES  
PARAHYANGAN CATHOLIC UNIVERSITY  
2024**

# LEMBAR PENGESAHAN

## PEMODELAN FREKUENSI KLAIM ASURANSI DENGAN *GENERALIZED LINEAR MODEL (GLM)*

**Evan Felix**

**NPM: 6161901023**

Telah lulus ujian skripsi pada 24 Januari 2024 dengan penguji:  
Dr. Erwinna Chendra dan Melania Eva Wulanningtyas, M.Pd.

**Bandung, 2 Februari 2024**

**Menyetujui,**

**Pembimbing 1**

**Pembimbing 2**

**Benny Yong, Ph.D.**

**Robyn Irawan, M.Sc.**

**Mengetahui,**

**Ketua Program Studi**

**Jonathan Hoseana, Ph.D.**

## PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

### **PEMODELAN FREKUENSI KLAIM ASURANSI DENGAN *GENERALIZED LINEAR MODEL (GLM)***

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,  
2 Februari 2024



Evan Felix  
NPM: 6161901023

## ABSTRAK

Perkembangan bisnis asuransi yang pesat menunjukkan meningkatnya kesadaran masyarakat akan pentingnya perlindungan terhadap risiko. Perusahaan asuransi memiliki tanggung jawab untuk membayar klaim yang diajukan oleh pemegang polis. Dalam konteks ini, keakuratan dalam mengestimasi frekuensi klaim menjadi krusial bagi perusahaan agar dapat menghindari kerugian besar atau risiko kebangkrutan. Analisis regresi linear terbatas dalam mengestimasi frekuensi klaim asuransi karena keterbatasan dalam menangani banyak variabel bebas dan risiko estimasi bernilai negatif. Untuk mengatasi kendala tersebut, penelitian ini menerapkan *Generalized Linear Model* sebagai metode yang lebih tepat dalam memodelkan frekuensi klaim asuransi. Dibandingkan dengan regresi linear, *Generalized Linear Model* mampu memberikan estimasi frekuensi klaim yang lebih akurat. Data yang digunakan adalah data klaim asuransi mobil yang terdiri dari 10.302 observasi dan 26 variabel. Dengan memanfaatkan perangkat lunak R, data tersebut diolah untuk mengidentifikasi variabel-variabel yang berperan penting dalam mengestimasi frekuensi klaim. Hasil penelitian menunjukkan bahwa *Generalized Linear Model* lebih baik dalam memodelkan frekuensi klaim asuransi mobil, yang terbukti dari nilai Akaike Information Criterion (AIC) *Generalized Linear Model* yang lebih rendah.

**Kata-kata kunci:** Asuransi; Frekuensi Klaim; Regresi Linear; *Generalized Linear Model*; *Akaike Information Criterion*.

## ABSTRACT

The rapid growth in the insurance business signifies an increasing awareness among the public regarding the importance of risk protection. Insurance companies bear the responsibility of paying claims filed by policyholders. In this context, accuracy in estimating claim frequency becomes crucial for these companies to avoid significant losses or the risk of bankruptcy. Linear regression analysis is limited in estimating insurance claim frequency due to constraints in handling numerous independent variables and the risk of negative estimation. To overcome these limitations, this research applies the Generalized Linear Model as a more appropriate method for modeling insurance claim frequency. Compared to linear regression, the Generalized Linear Model can provide more accurate estimates of claim frequency. The data used consists of 10,302 observations and 26 variables related to car insurance claims. Using the R software, this data was processed to identify variables that play a crucial role in estimating claim frequency. The research findings indicate that the Generalized Linear Model performs better in modeling car insurance claim frequency, as evidenced by the lower Akaike Information Criterion (AIC) value of the Generalized Linear Model.

**Keywords:** Insurance; Claim Frequency; Linear Regression; Generalized Linear Regression; Akaike Information Criterion.

*Numerous problems, a lot of drama, but in the end, you find yourself reading this.*



## KATA PENGANTAR

Penulis ingin menyampaikan rasa syukur kepada Tuhan Yang Maha Esa karena telah memberikan berkat dan kemudahan dalam menyelesaikan penulisan skripsi ini. Skripsi berjudul "Pemodelan Frekuensi Klaim Asuransi dengan *Generalized Linear Model (GLM)*" disusun sebagai salah satu persyaratan dalam menyelesaikan Program Studi Matematika. Di kesempatan ini, penulis ingin mengungkapkan terima kasih yang mendalam atas segala dukungan yang diberikan oleh berbagai pihak, yakni:

- Orang tua dan keluarga penulis yang selalu memberi dukungan, kasih, dan doa yang tulus selama proses penyelesaian skripsi ini.
- Bapak Benny Yong, Ph.D. dan Bapak Robyn Irawan, M.Sc. selaku dosen pembimbing yang selalu meluangkan waktu dan dengan sabar memberikan kritik serta saran yang membangun sehingga penulis mampu menyelesaikan skripsi ini dengan baik.
- Ibu Dr. Erwinna Chendra dan Ibu Melania Eva Wulanningtyas, M.Pd. selaku dosen penguji yang telah memberikan kritik, saran, dan kontribusinya dalam penyempurnaan skripsi ini.
- Seluruh dosen baik dari dalam Universitas Katolik Parahyangan, yakni dari Program Studi Matematika dan di luarnya, serta yang dari luar Universitas Katolik Parahyangan, yang mengasah ilmu penulis sehingga penulisan skripsi ini dapat diselesaikan dengan lancar.
- Timothy, Felix Alpha Winarto, Jovansen Hiustar, Raymond Susanto, dan Willy Zoe Ardiya Tanako selaku sahabat-sahabat penulis yang selalu memberikan dorongan, dukungan, dan hiburan dalam suka maupun duka selama keberlangsungan proses perkuliahan.
- Billy, Janice, Joice, Vania, dan Anastashia yang telah menjadi mitra penulis baik di dalam maupun di luar dunia perkuliahan.
- Admiral, Dennis, Luthfi, Egha, Aditya dan lain-lain yang telah menghibur dan menemani penulis dalam suka dan duka penulis.
- Seluruh mahasiswa angkatan 2019 atas kebersamaannya dalam bentuk apapun selama studi penulis di Universitas Katolik Parahyangan.

Penulis berharap setiap kenangan yang telah dibangun terukir di dalam hati setiap pribadi dan penulis mendoakan kesuksesan untuk setiap jalan yang sedang, atau akan ditempuh.

Penulis menyadari bahwa masih banyak kekurangan dan kesalahan dalam skripsi ini. Oleh karena itu, penulis menerima kritik dan saran yang membangun dalam bentuk apapun supaya menjadi skripsi ini menjadi lebih baik, berkembang, dan berguna bagi pembaca.

Bandung, 2 Februari 2024

Penulis

# DAFTAR ISI

<b>KATA PENGANTAR</b>	<b>viii</b>
<b>DAFTAR ISI</b>	<b>ix</b>
<b>DAFTAR GAMBAR</b>	<b>xi</b>
<b>DAFTAR TABEL</b>	<b>xii</b>
<b>1 PENDAHULUAN</b>	<b>1</b>
1.1 Latar Belakang . . . . .	1
1.2 Rumusan Masalah . . . . .	2
1.3 Tujuan . . . . .	2
1.4 <i>State of the Art</i> . . . . .	2
<b>2 LANDASAN TEORI</b>	<b>3</b>
2.1 Asuransi dan Frekuensi Klaim . . . . .	3
2.2 Regresi Linear Berganda . . . . .	3
2.2.1 Metode Kuadrat Terkecil . . . . .	4
2.2.2 Masalah dalam Mengestimasi Frekuensi Klaim Menggunakan Regresi Linear Berganda . . . . .	5
2.3 Generalized Linear Model . . . . .	5
2.3.1 Distribusi Keluarga Eksponensial . . . . .	6
2.3.2 Fungsi <i>Link</i> . . . . .	7
2.4 Estimasi Parameter Model Regresi Poisson dengan <i>Maximum Likelihood Estimation</i> (MLE) . . . . .	8
2.5 Fungsi <i>Stepwise</i> . . . . .	10
2.6 Pemilihan Model Terbaik . . . . .	10
2.7 Overdispersi . . . . .	10
<b>3 METODOLOGI PENELITIAN</b>	<b>12</b>
3.1 Regresi Linear Berganda . . . . .	12
3.2 <i>Generalized Linear Model</i> . . . . .	14
<b>4 HASIL DAN PEMBAHASAN</b>	<b>17</b>
4.1 Data . . . . .	17
4.2 Analisis Data Eksploratif . . . . .	18
4.3 Analisis Model . . . . .	19
4.4 Analisis model . . . . .	21
<b>5 KESIMPULAN DAN SARAN</b>	<b>24</b>
5.1 Kesimpulan . . . . .	24
5.2 Saran . . . . .	24
<b>DAFTAR REFERENSI</b>	<b>26</b>

A KODE PROGRAM

27

B DIAGRAM PENCILAN & BOXPLOT

30



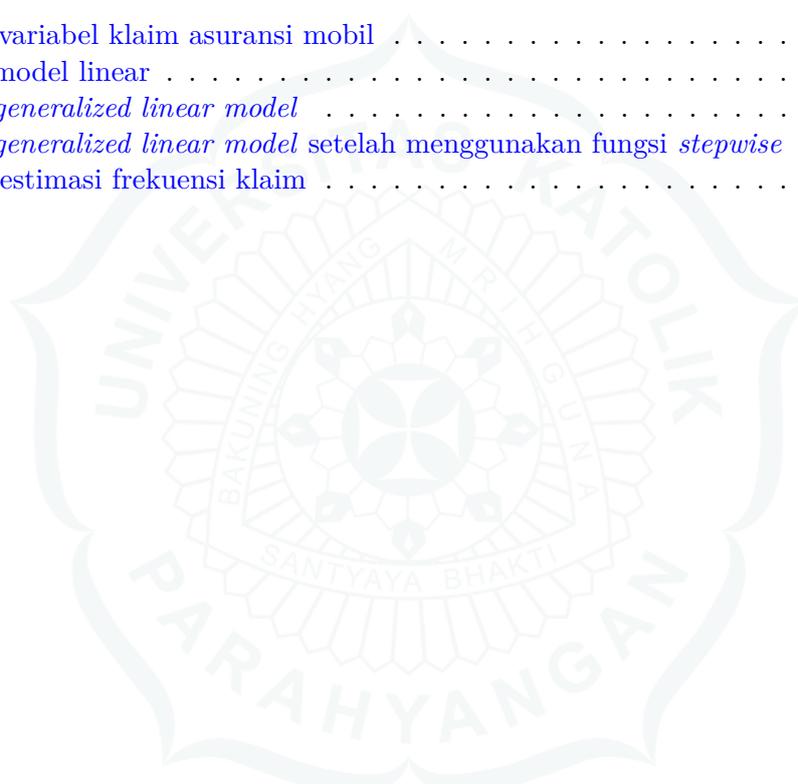
## DAFTAR GAMBAR

4.1	Analisa data eksploratif dengan diagram pencilan dan <i>boxplot</i> . . . . .	19
-----	---	----



## DAFTAR TABEL

2.1	Distribusi keluarga eksponensial dengan parameternya . . . . .	6
2.2	Fungsi-fungsi <i>link</i> yang sering digunakan . . . . .	8
2.3	Contoh kasus . . . . .	10
3.1	Data contoh . . . . .	12
4.1	Variabel-variabel klaim asuransi mobil . . . . .	17
4.2	Analisis model linear . . . . .	20
4.3	Analisis <i>generalized linear model</i> . . . . .	21
4.4	Analisis <i>generalized linear model</i> setelah menggunakan fungsi <i>stepwise</i> . . . . .	22
4.5	Evaluasi estimasi frekuensi klaim . . . . .	23



# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Asuransi merupakan salah satu bentuk manajemen risiko yang sudah ada sejak lama dan salah satu bisnis yang perkembangannya cukup pesat. Berkembangnya bisnis asuransi ini disebabkan oleh meningkatnya kesadaran masyarakat akan pentingnya perlindungan terhadap kemungkinan terjadinya risiko. Banyak produk asuransi yang saat ini ditawarkan oleh penyedia jasa asuransi, seperti asuransi jiwa, asuransi kesehatan, dan asuransi kendaraan.

Perusahaan asuransi memiliki tanggung jawab untuk membayar klaim yang diajukan oleh pemegang polis. Frekuensi klaim merupakan salah satu permasalahan yang sangat serius bagi perusahaan jika tidak ditangani dengan baik. Oleh karena itu, perusahaan asuransi harus memiliki estimasi yang tepat atas frekuensi klaim dari setiap nasabahnya. Hal ini memungkinkan perusahaan asuransi untuk mengatur besarnya premi dengan lebih tepat dan mencegah terjadinya risiko keuangan yang tidak diinginkan, seperti kerugian besar atau bahkan kebangkrutan.

Analisis regresi merupakan salah satu teknik statistik yang sering digunakan untuk mengestimasi suatu nilai dengan menganalisis hubungan satu variabel tak bebas (respon) dengan satu atau lebih variabel bebas (prediktor) yang dibentuk dalam suatu persamaan. Analisis regresi dibagi menjadi dua macam yaitu analisis regresi linear sederhana dan analisis regresi linear berganda. Kedua analisis regresi tersebut kurang baik untuk mengestimasi frekuensi klaim asuransi. Frees (2009) menyatakan bahwa regresi linear sederhana hanya dapat menjelaskan hubungan satu variabel prediktor dengan satu variabel respon [1]. Regresi ini tentu tidak dapat digunakan untuk mengestimasi banyaknya klaim yang memiliki banyak variabel prediktor di mana setiap nasabah memiliki banyak karakteristik untuk dimodelkan. Frees, dkk. (2014) menyatakan bahwa regresi linear berganda dapat digunakan, tetapi bukan alat yang terbaik untuk mengestimasi frekuensi klaim [2], salah satu alasannya adalah nilai estimasi dapat bernilai negatif.

Beberapa permasalahan yang timbul dari mengestimasi banyaknya klaim dengan regresi linear dapat diatasi dengan beralih ke *Generalized Linear Model* (GLM). Garrido, dkk. (2016) juga menunjukkan bahwa *Generalized Linear Model* adalah model yang tepat untuk mengestimasi frekuensi klaim [3]. Banyak profesional telah memanfaatkan *Generalized Linear Model* dalam perhitungan frekuensi klaim asuransi, sebagaimana yang terlihat dalam penelitian yang dilakukan oleh Covrig, dkk. [4]. Dalam studi tersebut, Covrig dkk. menggunakan *Generalized Linear Model Poisson* untuk mengestimasi frekuensi klaim asuransi motor. *Generalized Linear Model* dapat digunakan dalam analisis asuransi untuk memodelkan dan memahami frekuensi klaim dengan mempertimbangkan berbagai faktor yang memengaruhinya, seperti jenis kendaraan, usia pengemudi,

lokasi, dan faktor risiko lainnya.

Dalam penelitian ini, *Generalized Linear Model* akan digunakan untuk memodelkan frekuensi klaim yang belum pernah diaplikasikan pada data klaim asuransi mobil yang disediakan oleh Xiaomengsun di Kaggle. Penggunaan *Generalized Linear Model* direncanakan untuk mengatasi kemungkinan nilai negatif yang dapat muncul dalam regresi linear. Selain itu, penelitian ini akan membandingkan performa *Generalized Linear Model* dengan regresi linear dalam konteks pemodelan frekuensi klaim. Perbandingan tersebut diharapkan dapat memberikan pemahaman yang lebih mendalam tentang keunggulan *Generalized Linear Model* dalam memodelkan frekuensi klaim.

## 1.2 Rumusan Masalah

Berikut rumusan masalah-masalah yang akan dibahas dalam skripsi ini, antara lain

1. Bagaimana cara memodelkan frekuensi klaim asuransi mobil dengan model regresi dan *Generalized Linear Model*?
2. Bagaimana perbandingan performa antara model regresi linear dengan *Generalized Linear Model* dalam mengestimasi frekuensi klaim asuransi mobil?

## 1.3 Tujuan

Dengan merujuk kepada rumusan masalah yang telah diuraikan, tujuan yang ingin dicapai dalam skripsi ini adalah

1. Memodelkan frekuensi klaim asuransi mobil dengan model regresi linear dan *Generalized Linear Model*.
2. Menganalisis perbandingan performa antara model regresi linear dengan *Generalized Linear Model* dalam mengestimasi frekuensi klaim asuransi mobil.

## 1.4 *State of the Art*

Dalam skripsi ini, terdapat upaya untuk menganalisis frekuensi klaim asuransi menggunakan *Generalized Linear Model*. Sebagai referensi, terdapat studi terkait yang telah berhasil memodelkan frekuensi klaim asuransi dengan *Generalized Linear Model* yang dilakukan oleh Covrig, dkk. berjudul "*Some Generalized linear Models for the Estimation of the Mean Frequency of Claims in Motor Insurance*" [4]. Pada penelitian tersebut, model yang diterapkan adalah *Generalized Linear Model Poisson* untuk mengestimasi frekuensi klaim pada asuransi motor. Dalam skripsi ini, metode yang serupa, yaitu *Generalized Linear Model Poisson*, akan digunakan untuk mengestimasi frekuensi klaim asuransi. Data yang menjadi fokus penelitian berasal dari dataset asuransi mobil yang disediakan oleh Xiaomengsun di Kaggle.