

**SKRIPSI**

**PREDIKSI FREKUENSI KLAIM ASURANSI KENDARAAN  
BERMOTOR MENGGUNAKAN MODEL LINEAR  
TERGENERALISASI DENGAN PENDEKATAN FREKUENTIS  
DAN BAYESIAN**



**LUTHFI CAPRINANDA PUTRA**

**NPM: 6161901125**

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

**FINAL PROJECT**

**THE CLAIM'S FREQUENCY PREDICTION OF MOTOR  
VEHICLE INSURANCE USING GENERALIZED LINEAR  
MODEL WITH FREQUENTIST AND BAYESIAN APPROACH**



**LUTHFI CAPRINANDA PUTRA**

**NPM: 6161901125**

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

## LEMBAR PENGESAHAN

# PREDIKSI FREKUENSI KLAIM ASURANSI KENDARAAN BERMOTOR MENGGUNAKAN MODEL LINEAR TERGENERALISASI DENGAN PENDEKATAN FREKUENTIS DAN BAYESIAN

**Luthfi Caprinanda Putra**

**NPM: 6161901125**

Telah lulus ujian skripsi pada 18 Januari 2024 dengan penguji:  
Agus Sukmana, M.Sc. dan Iwan Sugiarto, M.Si.

**Bandung, 1 Februari 2024**

**Menyetujui,**

**Pembimbing**

**Felivia Kusnadi, M.Act.Sc.**

**Mengetahui,**

**Ketua Program Studi**

**Jonathan Hoseana, Ph.D.**

## PERNYATAAN

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

### **PREDIKSI FREKUENSI KLAIM ASURANSI KENDARAAN BERMOTOR MENGUNAKAN MODEL LINEAR TERGENERALISASI DENGAN PENDEKATAN FREKUENTIS DAN BAYESIAN**

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,  
1 Februari 2024



Luthfi Caprinanda Putra  
NPM: 6161901125

## ABSTRAK

Asuransi merupakan salah satu bentuk pengendalian risiko, di mana risiko tersebut dialihkan ke perusahaan asuransi. Dalam asuransi, terdapat istilah yang bernama frekuensi klaim. Frekuensi klaim merupakan banyaknya klaim yang diajukan oleh nasabah kepada pihak asuransi dalam kurun waktu tertentu. Informasi frekuensi klaim sangat penting bagi pihak asuransi karena dapat digunakan untuk menentukan tarif premi. Maka dari itu, penting bagi pihak asuransi untuk memodelkan data frekuensi klaim. Salah satu metode yang paling sering digunakan untuk memodelkan data frekuensi klaim adalah Model Linear Tergeneralisasi (GLM). Dalam GLM, terdapat berbagai pilihan distribusi, di mana distribusinya merupakan anggota keluarga eksponensial. Pada skripsi ini, digunakan distribusi Poisson dalam memodelkan data frekuensi klaim karena setelah dibandingkan dengan distribusi Binomial Negatif, distribusi Poisson lebih cocok dengan data frekuensi klaim yang digunakan. Dalam model GLM, terdapat parameter model yang harus diestimasi, di mana terdapat dua pendekatan dalam mengestimasi suatu parameter, yaitu pendekatan frekuentis dan Bayesian. Pada skripsi ini, dibentuk model GLM terbaik dan dilakukan prediksi frekuensi klaim menggunakan model dari masing-masing pendekatan. Model 4 terpilih sebagai model GLM dengan pendekatan frekuentis terbaik, sedangkan untuk pendekatan Bayesian, Model D terpilih sebagai model terbaik berdasarkan metode *k-fold Cross Validation*. Kedua model ini memiliki tingkat keakuratan yang tergolong tinggi dalam memprediksi frekuensi klaim, yaitu sebesar 92,47% untuk Model 4 dan 91,40% untuk Model D. Terakhir, dengan menggunakan metode *k-fold Cross Validation* dan mempertimbangkan *running time*, model GLM dengan pendekatan frekuentis dianggap lebih baik daripada model GLM dengan pendekatan Bayesian dalam memodelkan data frekuensi klaim dari asuransi kendaraan bermotor.

**Kata-kata kunci:** frekuensi klaim; asuransi kendaraan bermotor; model linear tergeneralisasi; pendekatan frekuentis; pendekatan Bayesian.

## ABSTRACT

Insurance is a form of risk control, where the risk is transferred to the insurance company. In insurance, there is a term called claim frequency. Claim frequency is the number of claims submitted by customers to the insurance company within a certain period of time. Claim frequency information is very important for the insurance company because it can be used to determine premium rates. Therefore, it is important for the insurance company to modeling the claim frequency data. One of the most commonly used methods to modeling claim frequency data is the Generalized Linear Model (GLM). In GLM, there are various distribution options, where the distribution is a member of the exponential family. In this thesis, the Poisson distribution is used in modeling claim frequency data because after being compared with the Negative Binomial distribution, the Poisson distribution is more suitable for the claim frequency data used. In the GLM model, there are model parameters that must be estimated, where there are two approaches in estimating a parameter, namely the frequency and Bayesian approaches. In this thesis, the best GLM model is formed and prediction of claim frequency is done using the model from each approach. Model 4 was selected as the best GLM model with frequency approach, while for Bayesian approach, Model D was selected as the best model based on the k-fold Cross Validation method. Both models have a relatively high level of accuracy in predicting claim frequency, which is 92,47% for Model 4 and 91,40% for Model D. Finally, using the *k-fold Cross Validation* method and considering *running time*, the GLM model with the frequency approach is considered better than the GLM model with the Bayesian approach in modeling claim frequency data from motor vehicle insurance.

**Keywords:** claim frequency; motor vehicle insurance; Generalized Linear Model; frequentist approach; Bayesian approach.

*Untuk Ayah dan Ibu*



## KATA PENGANTAR

Segala puji dan syukur kepada Tuhan Yang Maha Esa karena atas berkat, rahmat, dan karunia-Nya yang diberikan kepada penulis, penulis dapat menyelesaikan skripsi yang berjudul "**Prediksi Frekuensi Klaim Asuransi Kendaraan Bermotor Menggunakan Model Linear Tergeneralisasi Dengan Pendekatan Frekuentis dan Bayesian**". Setelah melalui proses selama beberapa bulan, akhirnya penulis dapat mengajukan skripsi ini sebagai syarat tugas akhir dan syarat memperoleh gelar sarjana (S1) Program Studi Matematika di Fakultas Teknologi Informasi dan Sains, Universitas Katolik Parahyangan.

Selama masa perkuliahan hingga penyelesaian skripsi ini, banyak tantangan dan hambatan yang dialami oleh penulis. Maka dari itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Ayah, Ibu, Mas Arbi, dan Shalma, terima kasih karena selalu memberikan dukungan, memberikan doa, memberikan nasihat, dan memenuhi kebutuhan penulis selama proses penyelesaian skripsi ini sehingga penulis dapat menyelesaikan skripsinya dengan baik.
2. Ibu Felivia Kurnadi, M.Act.Sc., ASAI, selaku pembimbing tunggal dan dosen wali, terima kasih sudah membimbing penulis kurang lebih selama setahun, sudah menyediakan waktu, pikiran, tenaga, dan kesabaran selama proses penyelesaian skripsi ini sehingga penulis dapat menyelesaikan skripsinya dengan baik.
3. Bapak Agus Sukmana, M.Sc dan Bapak Iwan Sugiarto, M.Si selaku dosen penguji, terima kasih sudah meluangkan waktu untuk memberikan kritik dan saran sehingga penulis dapat memperbaiki skripsinya menjadi lebih baik lagi.
4. Seluruh dosen dan Tata Usaha FTIS yang sudah memberikan ilmu, waktu, dan bantuan selama penulis menjalankan perkuliahan,
5. Sahabat seperjuangan skripsi, Dennis, Admiral, Egha, dan Adit. Terima kasih sudah menjadi teman diskusi dan sudah memberikan dukungan selama proses penyelesaian skripsi ini. Terima kasih atas waktu dan canda tawanya selama masa perkuliahan sampai penyelesaian skripsi ini. Terima kasih sudah menjadi sahabat penulis selama masa perkuliahan. Semoga tali persahabatan kita tidak akan putus sampai kedepannya.
6. Teman-teman Ansos Mat, Kevin, Biqy, Michael, Billy, dan Timi. Terima kasih atas waktu, ilmu, canda tawa, hiburan, bantuan, dan dukungan kepada penulis selama masa perkuliahan sehingga penulis dapat menyelesaikan skripsi ini dengan baik.
7. Sahabat seperbandungan, Faiq, Alif, Nuno, dan Dimas. Terima kasih sudah menemani penulis selama masa perkuliahan. Terima kasih atas dukungan yang diberikan selama proses penyelesaian skripsi ini. Terima kasih sudah menjadi sahabat penulis sejak SMA hingga saat ini. Semoga tali persahabatan kita tidak akan putus sampai kedepannya.
8. Teman-teman RESIST 2019, terima kasih sudah menghibur penulis selama mengerjakan skripsi ini. Terima kasih atas dukungan yang diberikan selama proses penyelesaian skripsi ini.

9. Teman-teman magang Indonesia Re, terima kasih karena senantiasa menyemangati dan bersedia untuk *take over* pekerjaan disaat penulis sedang fokus dalam pengerjaan skripsi.

Bandung, 1 Februari 2024

Penulis



# DAFTAR ISI

<b>KATA PENGANTAR</b>	<b>viii</b>
<b>DAFTAR ISI</b>	<b>x</b>
<b>DAFTAR GAMBAR</b>	<b>xii</b>
<b>DAFTAR TABEL</b>	<b>xiii</b>
<b>1 PENDAHULUAN</b>	<b>1</b>
1.1 Latar Belakang . . . . .	1
1.2 Rumusan Masalah . . . . .	3
1.3 Tujuan . . . . .	3
1.4 <i>State of the Art</i> . . . . .	3
<b>2 LANDASAN TEORI</b>	<b>5</b>
2.1 Regresi Linear . . . . .	5
2.1.1 Regresi Linear Sederhana . . . . .	5
2.1.2 Regresi Linear Berganda . . . . .	6
2.2 Estimasi Parameter . . . . .	7
2.2.1 <i>Maximum Likelihood Estimator</i> . . . . .	8
2.2.2 Estimasi Bayesian . . . . .	9
2.3 Proses <i>Stepwise</i> . . . . .	11
2.3.1 <i>Forward Selection</i> . . . . .	11
2.3.2 <i>Backward Elimination</i> . . . . .	11
2.4 Evaluasi Model . . . . .	11
2.4.1 <i>Mean Squared Error</i> . . . . .	12
2.4.2 <i>k-Fold Cross Validation</i> . . . . .	12
2.5 Teorema Bayes . . . . .	12
2.6 Distribusi Peluang . . . . .	13
2.6.1 Distribusi Poisson . . . . .	13
2.6.2 Distribusi Binomial Negatif . . . . .	15
2.6.3 Distribusi Seragam . . . . .	17
<b>3 MODEL LINEAR TERGENERALISASI DENGAN PENDEKATAN FREKUENTIS DAN BAYESIAN</b>	<b>19</b>
3.1 Model Linear Tergeneralisasi . . . . .	19
3.1.1 <i>Exponential Dispersion Model (EDM)</i> . . . . .	19
3.1.2 Komponen Sistematis . . . . .	22
3.1.3 Fungsi <i>Link (Link Function)</i> . . . . .	22
3.1.4 Model GLM untuk Frekuensi Klaim . . . . .	23
3.2 Estimasi Parameter Pada Model Linear Tergeneralisasi . . . . .	24
3.2.1 <i>Maximum Likelihood Estimator</i> Dalam Model Linear Tergeneralisasi (GLM) . . . . .	24
3.2.2 <i>Power Prior</i> Dalam Model Linear Tergeneralisasi (GLM) . . . . .	25

3.3	Algoritma Penerapan Model Linear Tergeneralisasi . . . . .	27
3.4	Sistematika Pembentukan Model Linear Tergeneralisasi . . . . .	27
<b>4</b>	<b>HASIL DAN ANALISIS</b>	<b>28</b>
4.1	Data . . . . .	28
4.1.1	Deskripsi Data . . . . .	28
4.1.2	Pengolahan Data Awal . . . . .	30
4.1.3	Analisis Data Eksplorasi . . . . .	31
4.2	Perbandingan Model GLM Poisson Dengan GLM Binomial Negatif . . . . .	34
4.3	Pembentukan Model GLM Poisson Dengan Pendekatan Frekuentis . . . . .	35
4.3.1	Pengolahan Data . . . . .	35
4.3.2	Pembentukan Model . . . . .	36
4.3.3	Pemilihan Model Terbaik . . . . .	41
4.4	Pembentukan Model GLM Poisson Dengan Pendekatan Bayesian . . . . .	42
4.4.1	Pengolahan Data . . . . .	42
4.4.2	Pembentukan Model . . . . .	42
4.4.3	Pemilihan Model Terbaik . . . . .	48
4.5	Interpretasi Model . . . . .	48
4.5.1	Interpretasi Model GLM Poisson Dengan Pendekatan Frekuentis . . . . .	48
4.5.2	Interpretasi Model GLM Poisson Dengan Pendekatan Bayesian . . . . .	49
4.6	Prediksi Frekuensi Klaim . . . . .	51
4.6.1	Prediksi Frekuensi Klaim Menggunakan Model GLM Poisson Dengan Pendekatan Frekuentis . . . . .	52
4.6.2	Prediksi Frekuensi Klaim Menggunakan Model GLM Poisson Dengan Pendekatan Bayesian . . . . .	52
4.7	Perbandingan Model GLM Poisson Dengan Pendekatan Frekuentis dan Pendekatan Bayesian . . . . .	53
<b>5</b>	<b>KESIMPULAN DAN SARAN</b>	<b>55</b>
5.1	Kesimpulan . . . . .	55
5.2	Saran . . . . .	55
	<b>DAFTAR REFERENSI</b>	<b>56</b>

## DAFTAR GAMBAR

2.1	Grafik Fungsi Massa Peluang Distribusi Poisson . . . . .	14
4.1	Histogram Variabel ClaimNb . . . . .	31
4.2	Histogram Variabel DrivAge . . . . .	32
4.3	Histogram Variabel VehAge . . . . .	32
4.4	Histogram Variabel Density Berdasarkan Area . . . . .	33
4.5	Histogram Variabel BonusMalus . . . . .	33
4.6	Histogram Variabel VehBrand Berdasarkan Banyaknya Klaim . . . . .	34
4.7	Histogram Variabel VehPower Berdasarkan Banyaknya Klaim . . . . .	34
4.8	Histogram Variabel Area Berdasarkan Banyaknya Klaim . . . . .	34
4.9	Histogram Variabel VehGas Berdasarkan Banyaknya Klaim . . . . .	34

## DAFTAR TABEL

3.1	Jenis-jenis fungsi <i>link</i> . . . . .	23
4.1	Deskripsi Variabel . . . . .	29
4.2	Persebaran variabel ClaimNb . . . . .	31
4.3	Nilai MSE dari masing-masing model . . . . .	35
4.4	Kriteria pengelompokkan variabel VehBrandCat . . . . .	36
4.5	Kriteria pengelompokkan variabel VehPowerCat . . . . .	36
4.6	Kriteria pengelompokkan variabel DrivAgeCat . . . . .	36
4.7	<i>Summary</i> Model 1 . . . . .	37
4.8	Nilai MSE dari masing-masing bagian untuk Model 1 . . . . .	37
4.9	<i>Summary</i> Model 2 . . . . .	38
4.10	Nilai MSE dari masing-masing bagian untuk Model 2 . . . . .	38
4.11	<i>Summary</i> Model 3 . . . . .	39
4.12	Nilai MSE dari masing-masing bagian untuk Model 3 . . . . .	40
4.13	<i>Summary</i> Model 4 . . . . .	40
4.14	Nilai MSE dari masing-masing bagian untuk Model 4 . . . . .	41
4.15	Nilai <i>Cross Validation</i> untuk masing-masing model GLM dengan pendekatan frekuentis . . . . .	41
4.16	<i>Summary</i> Model A . . . . .	43
4.17	Nilai MSE dari masing-masing bagian untuk Model A . . . . .	43
4.18	<i>Summary</i> Model B . . . . .	44
4.19	Nilai MSE dari masing-masing bagian untuk Model B . . . . .	44
4.20	<i>Summary</i> Model C . . . . .	45
4.21	Nilai MSE dari masing-masing bagian untuk Model C . . . . .	46
4.22	<i>Summary</i> Model D . . . . .	47
4.23	Nilai MSE dari masing-masing bagian untuk Model D . . . . .	47
4.24	Nilai <i>Cross Validation</i> untuk masing-masing model GLM dengan pendekatan Bayesian . . . . .	48
4.25	Rata-rata nilai MSE . . . . .	53

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Transportasi merupakan perpindahan manusia atau barang dari satu tempat ke tempat lainnya menggunakan sebuah kendaraan. Dengan menggunakan transportasi, manusia dapat lebih mudah untuk melakukan aktivitas sehari-hari. Terdapat berbagai alternatif transportasi, yakni melalui darat, laut, dan udara. Dalam skripsi ini, kita akan lebih memfokuskan pada transportasi darat. Contoh kendaraan dari transportasi darat adalah mobil, bus, dan motor. Dari ketiga kendaraan tersebut, motor memiliki tingkat risiko paling tinggi untuk mengalami kecelakaan yang di mana hal ini divalidasi berdasarkan *Integrated Road Safety Management System* (IRSMS). IRSMS menunjukkan bahwa di Indonesia, pada jangka waktu 27 Juni 2022 sampai 3 Juli 2022, terjadi kecelakaan lalu lintas sebanyak 1.653 kecelakaan dengan 74% di antaranya melibatkan sepeda motor<sup>1</sup>. Hal ini masuk akal karena motor memiliki ukuran lebih kecil dibandingkan kendaraan lainnya dan memiliki kecepatan yang cukup tinggi sehingga cenderung lebih rentan terhadap kecelakaan lalu lintas. Jika kecelakaan terjadi, maka sang pengendara akan mengalami berbagai kerugian (seperti jatuh sakit dan harus memperbaiki motornya yang rusak), di mana hal tersebut merupakan risiko dalam berkendara. Oleh sebab itu, pengendara tentunya harus bisa meminimalisir kerugian-kerugian tersebut ketika kecelakaan terjadi.

Terdapat berbagai cara untuk meminimalisir kerugian yang ada, salah satunya adalah membeli produk asuransi. Asuransi sendiri merupakan salah satu bentuk pengendalian risiko, di mana risiko tersebut dialihkan dari satu pihak ke pihak lain dan dalam hal ini adalah perusahaan asuransi. Asuransi terbagi menjadi dua, yaitu asuransi jiwa dan asuransi umum. Asuransi kendaraan bermotor termasuk ke dalam kategori asuransi umum. Dengan membeli produk asuransi kendaraan bermotor, maka kerugian yang dialami ketika kecelakaan terjadi dapat teralihkan dengan menerima manfaat yang dijanjikan oleh produk asuransi tersebut. Untuk mendapatkan manfaat tersebut, pemegang polis harus membayarkan sejumlah uang kepada pihak asuransi terkait yang disebut sebagai premi asuransi. Besar premi yang dibayarkan dari masing-masing pemegang polis pastinya berbeda, bergantung pada risiko yang ditanggung oleh pihak asuransi. Semakin tinggi risiko yang ditanggung, maka semakin besar premi yang dibayarkan. Besar premi asuransi yang harus dibayarkan biasanya sudah tertera pada dokumen polis asuransi.

Berbicara tentang besar premi asuransi, terdapat beberapa faktor penting yang digunakan oleh

---

<sup>1</sup><https://lantas.kalsel.polri.go.id/index.php/2022/07/06/data-kecelakaan-nasional-berdasarkan-data-irsms-integrated-road-safety-management-system-korlantas-polri-periode-27-juni-2022-sampai-dengan-03-juli-2022/> diakses pada 4 September 2023.

perusahaan asuransi untuk menentukan tarif premi, salah satunya adalah frekuensi klaim. Frekuensi klaim merupakan banyaknya klaim yang diajukan oleh pemegang polis selama periode waktu tertentu. Dengan kata lain, frekuensi klaim dapat mengukur seberapa sering nasabah mengajukan klaim pada polis asuransi mereka. Semakin sering nasabah mengajukan klaim, semakin besar risiko yang dihadapi oleh perusahaan asuransi yang artinya tarif premi yang akan dikenakan pihak asuransi kepada nasabah akan semakin tinggi. Selain itu, frekuensi klaim juga dapat mengantisipasi jumlah klaim yang kemungkinan akan diajukan oleh nasabah selama periode tertentu dan menentukan cadangan dana yang cukup untuk menanggapi klaim tersebut. Jadi, penting bagi perusahaan asuransi untuk melakukan pemodelan data frekuensi klaim untuk melakukan prediksi agar dapat mempersiapkan dan mengelola risiko secara lebih efektif.

Terdapat beberapa metode yang dapat digunakan untuk memodelkan data frekuensi klaim. Berdasarkan penelitian sebelumnya, metode Kredibilitas Buhlmann-Straub pernah digunakan untuk memodelkan data frekuensi klaim [1]. Pemodelan tersebut menggunakan distribusi Binomial Negatif-Lindley karena data yang digunakan memiliki overdispersi yang tinggi. Namun, metode ini memiliki kelemahan, yaitu kurang fleksibel dalam menangani pola yang kompleks atau non-linear dalam data karena pilihan distribusi yang sedikit. Selain metode Kredibilitas Buhlmann-Straub, metode *Random Forest* (RF) juga sering digunakan dalam memodelkan data klaim. RF pernah diterapkan untuk memodelkan data tingkat keparahan klaim [2]. Namun, metode ini memiliki kelemahan, yaitu interpretasi model yang cukup sulit karena ketika tingkat keakuratan model semakin tinggi, model RF akan semakin kompleks.

Metode lain yang sering digunakan adalah *Generalized Linear Model* (GLM). GLM memiliki kelebihan yang tidak dimiliki dua metode yang sudah dijelaskan sebelumnya. Metode GLM dianggap memiliki fleksibilitas yang tinggi karena mempunyai banyak pilihan distribusi untuk menyesuaikan kecocokan dengan data. Selain itu, hasil dari model GLM mudah untuk diinterpretasikan karena GLM memperhitungkan efek dari setiap variabel prediktornya sehingga memungkinkan pengguna untuk mengidentifikasi variabel apa saja yang signifikan untuk melakukan prediksi. GLM juga pernah digunakan untuk pemodelan data frekuensi klaim. Berdasarkan penelitian Uly Saputri dan Dodi Devianto pada tahun 2020 [3], mereka menggunakan GLM untuk memodelkan data frekuensi klaim. Model GLM terbaik dipilih berdasarkan kriteria *Akaike's Information Criterion* (AIC). Dari penelitian tersebut, distribusi Poisson dipilih sebagai distribusi dari data frekuensi klaim sehingga GLM Poisson digunakan untuk memodelkan data tersebut. Selain itu, berdasarkan penelitian Rini Oktavia, dkk. [4], mereka juga menggunakan GLM untuk memodelkan data frekuensi klaim dengan distribusi yang digunakan adalah distribusi Binomial Negatif. Pada penelitian tersebut menunjukkan bahwa model GLM Binomial Negatif memberikan kecocokan pada data frekuensi klaim. Maka dari itu, metode GLM digunakan pada skripsi ini untuk memodelkan data frekuensi klaim dari asuransi kendaraan bermotor dan akan dibandingkan terlebih dahulu GLM Poisson dan Negatif Binomial agar dapat mengetahui distribusi mana yang cocok untuk data frekuensi klaim yang digunakan.

Dalam model GLM, terdapat parameter  $\beta$ , di mana parameter ini nilainya akan diestimasi. Terdapat dua pendekatan yang dapat digunakan untuk mengestimasi suatu parameter, yaitu pendekatan frekuentis dan Bayesian. Berdasarkan penelitian sebelumnya, pendekatan frekuentis pernah digunakan untuk mengestimasi model GLM Poisson untuk data klaim asuransi kendaraan

[5]. Penelitian tersebut menggunakan metode *Maximum Likelihood Estimator* (MLE) untuk mengestimasi parameter  $\beta$  pada GLM. Pendekatan Bayesian juga pernah digunakan untuk mengestimasi parameter  $\beta$  pada GLM. Berdasarkan penelitian yang dilakukan oleh Joseph G. Ibrahim dan Purushottam W. Laud pada tahun 1991, penelitian tersebut menggunakan pendekatan Bayesian yaitu *Jeffrey's Prior* untuk mengestimasi parameter  $\beta$  pada GLM. Kemudian, model GLM dengan kedua pendekatan ini juga pernah dibandingkan [6]. Penelitian tersebut menggunakan metode MLE untuk pendekatan frekuentis dan *Power Prior* untuk pendekatan Bayesian dalam mengestimasi parameter  $\beta$  pada GLM. Hasilnya adalah model dengan pendekatan Bayesian merupakan model yang lebih baik karena memiliki galat yang lebih kecil dibandingkan model dengan pendekatan frekuentis berdasarkan ukuran *Mean Squared Prediction Error* (MSPE). Maka dari itu, pada skripsi ini akan dibandingkan model GLM dengan pendekatan frekuentis dan Bayesian menggunakan metode *k-fold Cross Validation*. Metode tersebut memiliki kelebihan yaitu memungkinkan pemanfaatan yang maksimal dari setiap observasi dalam pembuatan dan evaluasi model karena pembuatan dan evaluasi model tersebut dilakukan sebanyak  $k$  kali. Selain itu, akan dilakukan prediksi frekuensi klaim menggunakan model GLM dari masing-masing pendekatan.

## 1.2 Rumusan Masalah

Berikut merupakan rumusan masalah pada skripsi ini.

1. Bagaimana hasil model GLM terbaik dari masing-masing pendekatan (frekuentis dan Bayesian) dalam memodelkan data frekuensi klaim dari asuransi kendaraan bermotor?
2. Seberapa akurat model GLM terbaik dari masing-masing pendekatan (frekuentis dan Bayesian) dalam memprediksi frekuensi klaim?
3. Model dengan pendekatan mana yang lebih baik dalam memodelkan data frekuensi klaim dari asuransi kendaraan bermotor?

## 1.3 Tujuan

Berikut merupakan tujuan penelitian dari skripsi ini.

1. Membentuk model GLM terbaik dari masing-masing pendekatan (frekuentis dan Bayesian) dalam memodelkan data frekuensi klaim dari asuransi kendaraan bermotor.
2. Menganalisis tingkat keakuratan model GLM dari masing-masing pendekatan (frekuentis dan Bayesian) dalam memprediksi frekuensi klaim.
3. Menentukan dan menganalisis model dengan pendekatan terbaik dalam memodelkan data frekuensi klaim dari asuransi kendaraan bermotor.

## 1.4 *State of the Art*

Pada penelitian sebelumnya, model GLM dengan kedua pendekatan (frekuentis dan Bayesian) sudah pernah dibandingkan [6], di mana untuk pendekatan frekuentis menggunakan metode

*Maximum Likelihood Estimator* dan untuk pendekatan Bayesian menggunakan metode *Power Prior*. Pada penelitian tersebut, ukuran yang digunakan dalam membandingkan model adalah *Mean Squared Prediction Error* (MSPE), sedangkan pada skripsi ini, digunakan *k-fold Cross Validation*. Selain itu, walaupun data frekuensi klaim yang digunakan sama, namun pengolahan data pada skripsi ini berbeda dibandingkan penelitian tersebut, di mana perbedaannya seperti pemilihan variabel, pemilihan ukuran sampel, dan pembentukan data historis. Kemudian, untuk *software* yang digunakan, penelitian tersebut menggunakan bantuan *software* SAS dalam membangun model GLM dan mengestimasi parameter model. Pada skripsi ini, akan digunakan *software* R Studio dalam membangun model GLM dan mengestimasi parameter model. Kemudian, model GLM terbaik dari masing-masing pendekatan (frekuentis dan Bayesian) akan digunakan untuk melakukan prediksi frekuensi klaim, di mana pada penelitian sebelumnya, tidak dilakukan prediksi untuk frekuensi klaim.

