

**SKRIPSI**

**PREDIKSI FREKUENSI KLAIM UNTUK ASURANSI  
KENDARAAN BERMOTOR DENGAN MENGGUNAKAN  
MODEL LINEAR TERGENERALISASI DAN *XGBOOST***



**Laurentco**

**NPM: 6161801019**

**PROGRAM STUDI MATEMATIKA  
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS  
UNIVERSITAS KATOLIK PARAHYANGAN  
2022**

**FINAL PROJECT**

**CLAIM FREQUENCY PREDICTION FOR MOTOR  
INSURANCE USING GENERALIZED LINEAR MODEL AND  
XGBOOST**



**Laurentco**

**NPM: 6161801019**

**DEPARTMENT OF MATHEMATICS  
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES  
PARAHYANGAN CATHOLIC UNIVERSITY  
2022**

# LEMBAR PENGESAHAN

## PREDIKSI FREKUENSI KLAIM UNTUK ASURANSI KENDARAAN BERMOTOR DENGAN MENGGUNAKAN MODEL LINEAR TERGENERALISASI DAN *XGBOOST*

Laurentco

NPM: 6161801019

Bandung, 24 Januari 2022

Menyetujui,

Pembimbing



Maria Anestasia, M.Si., M.Act.Sc.

Ketua Tim Penguji



Prof. Dr. Julius Dharma Lesmono

Anggota Tim Penguji



Felivia Kusnadi, M.Act.Sc.

Mengetahui,

Ketua Program Studi



Dr. Livia Owen

## PERNYATAAN

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

**PREDIKSI FREKUENSI KLAIM UNTUK ASURANSI KENDARAAN  
BERMOTOR DENGAN MENGGUNAKAN MODEL LINEAR  
TERGENERALISASI DAN *XGBOOST***

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,  
Tanggal 24 Januari 2022



Laurentco

NPM: 6161801019

## ABSTRAK

Transportasi merupakan salah satu sarana yang menunjang kehidupan masyarakat di Indonesia. Setiap harinya terjadi peningkatan mobilitas dan kebutuhan masyarakat terhadap kendaraan bermotor. Hal ini menyebabkan tingginya risiko yang harus dihadapi oleh pengguna kendaraan. Salah satu cara untuk mengurangi risiko tersebut adalah menggunakan asuransi kendaraan bermotor. Dalam asuransi kendaraan bermotor, banyaknya klaim yang diajukan oleh pemegang polis memengaruhi harga polis tersebut, sehingga penting bagi perusahaan untuk memprediksi frekuensi banyaknya klaim, sesuai dengan karakteristik masing-masing pemegang polis. Pada skripsi ini, akan dilakukan pemodelan untuk memprediksi frekuensi klaim dengan menggunakan data asuransi kendaraan bermotor di Perancis. Data tersebut memiliki variabel-variabel yang lengkap dan identik dengan pengaplikasiannya di Indonesia. Terdapat dua metode yang digunakan dalam skripsi ini, yaitu metode linear tergeneralisasi (GLM) berdistribusi Poisson dan *XGBoost*. Kedua metode ini banyak digunakan dalam model regresi untuk memprediksi nilai dengan *output* berbentuk numerik. Variabel yang dipertimbangkan dalam pemodelan di antaranya adalah variabel usia kendaraan, usia pengemudi, tenaga kendaraan, tipe bahan bakar, merek kendaraan, *area*, *region*, dan kepadatan penduduk. Selanjutnya, hasil pemodelan akan dianalisis untuk memperoleh prediksi frekuensi klaim. Terakhir, akan dilakukan perbandingan antara kedua model tersebut dengan menggunakan ukuran perbandingan yaitu deviansi Poisson dan *root mean squared error* (RMSE). Perbandingan dilakukan untuk menentukan model yang lebih cocok dalam memprediksi frekuensi klaim pada data. Hasil percobaan yang telah dilakukan menunjukkan bahwa model *XGBoost* lebih baik dalam memodelkan dan memprediksi frekuensi klaim pada asuransi kendaraan bermotor dibandingkan dengan model linear tergeneralisasi (GLM).

**Kata-kata kunci:** model linear tergeneralisasi, asuransi kendaraan bermotor, *XGBoost*, frekuensi klaim

## ABSTRACT

Transportation is a tool that support the daily activities of Indonesian people. As days go by, there is an increase in mobility and people's need for vehicle. This causes a high risk that must be faced by driver, where this risk not only interferes with the safety of vehicle users, but also results in damages of the vehicle. One way to reduce this risk is to use motor vehicle insurance. In motor vehicle insurance, the number of claims that submitted by policyholders can affect the price of the policy, so it is important for insurance companies to predict the frequency of the number of claims that will be submitted by policyholders, each of which has different characteristics. In this final project, the model will be carried out to predict the claim frequencies using motor vehicle insurance data based in France. The data has complete variables and related to its application in Indonesia. There are two methods used in this final project, which is Generalized Linear Model (GLM) and XGBoost. Both of these methods are widely used in regression models and are used to predict values with numeric output. The variables considered in the model include the age of the vehicle, the age of the driver, the power of the vehicle, the type of fuel, the brand of the vehicle, area, region, and population density. Furthermore, the modeling results will be analyzed to get a prediction of the frequency of claims. Finally, a comparison will be made between the two models using comparison measure, namely the Poisson deviation and root mean squared error (RMSE). Comparisons were made to determine the most suitable model for predicting the frequency of claims in the data. The experiment showed that the XGBoost model is better at modeling and predicting the claim frequencies on motor vehicle insurance compared to the Generalized Linear Model (GLM)

**Keywords:** Generalized Linear Model, motor vehicle insurance, XGBoost, claim frequency

## KATA PENGANTAR

Puji dan syukur kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, penulis dapat menyelesaikan skripsi yang berjudul "**Prediksi Frekuensi Klaim Untuk Asuransi Kendaraan Bermotor Dengan Menggunakan Model Linear Tergeneralisasi Dan XGBoost**". Penulisan skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan studi Strata-1 di Program Studi Matematika, Fakultas Teknologi Informasi dan Sains, Universitas Katolik Parahyangan, Bandung.

Pada kesempatan ini, penulis ingin menyampaikan ucapan terima kasih kepada pihak-pihak yang telah membantu penulis selama proses perkuliahan dan penyusunan skripsi ini, diantaranya adalah :

1. Mami, Papi, dan Cece yang senantiasa memberikan motivasi, dukungan, dan doa sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Terima kasih karena telah hadir di dunia ini dan memberikan kesempatan kepada penulis untuk dapat mengenyam pendidikan hingga jenjang perkuliahan.
2. Ibu Maria Anastasia, M.Si., M.Act.Sc. selaku dosen pembimbing yang telah membantu, mengarahkan, dan memberikan saran serta masukan kepada penulis selama proses penyusunan skripsi ini. Terima kasih karena telah membimbing penulis dengan penuh kesabaran, mulai dari proses bimbingan pertama hingga terakhir, sehingga penulis dapat menyelesaikan skripsi ini dengan baik.
3. Bapak Prof. Dr. Julius Dharma Lesmono selaku Ketua Penguji dan Ibu Felivia, MActSc. selaku Anggota Penguji. Terima kasih karena telah meluangkan waktu serta memberikan saran, masukan, dan kritik yang sangat membantu penulis agar skripsi ini menjadi lebih baik lagi.
4. Bapak Liem Chin, M.Si. selaku Koordinator Skripsi serta Dosen Wali yang telah memberikan saran dan masukan selama proses perkuliahan maupun selama proses penyusunan skripsi ini.
5. Seluruh dosen, tata usaha, dan prakarya FTIS yang telah memberikan ilmu, nasihat, serta bimbingan selama proses perkuliahan.
6. Gabrielle Octarina yang selalu menemani dan membantu penulis dari masa SMA hingga akhir perkuliahan ini. Terima kasih karena telah mengisi hari-hari penulis dengan penuh kebahagiaan, serta memotivasi penulis untuk terus berubah menjadi individu yang lebih baik lagi.
7. Ci Rachael dan Ko Jeffry yang selalu memberikan dukungan dan masukan selama proses penyusunan skripsi ini. Terima kasih karena telah meyakinkan penulis bahwa tidak ada hal yang mustahil untuk dicapai jika dilakukan dengan tekad dan niat.
8. Teman-teman SIMONYONG: Andrew, Patrick, Elbert, Kevin, dan Santo. Terima kasih karena telah menemani penulis selama proses perkuliahan dan menjadi tempat untuk bertukar ilmu sehingga penulis menjadi pribadi yang lebih baik.
9. Teman-teman Recis Community: Aldrich, Aldo, Ricky, Nicky, Atin, Jessen, Jason, Verdy, Tio, KP, Ojan, Bule, Darren, dan Xander. Terima kasih karena telah menghibur dan menyemangati penulis selama proses perkuliahan dan skripsi.
10. Teman-teman JKT18: Anya, Vellina, Audrey, Nadia, Bryan, Chihan, Mikhael, Felix, Santo, Andrew, Patrick, Kevin, Elbert, dan Aji. Terima kasih atas kenangan dan kebersamaannya selama proses perkuliahan.
11. Teman-teman KMBP: Billie, Ivan, Andi, Chihan, Deddy, Devon, Felix, Desmond, dan Rhandy.

Terima kasih atas kebersamaannya selama proses perkuliahan.

12. Ko Robyn yang telah membantu dan memberikan masukan, saran, serta motivasi sehingga penulis dapat menyelesaikan skripsi ini.
13. Teman-teman jurusan Matematika angkatan 2016, 2017, 2018, 2019, dan 2020 yang tidak dapat disebutkan namanya satu per satu. Terima kasih atas saran dan ilmu yang diberikan selama proses perkuliahan.
14. Semua pihak yang telah memberikan bantuan dan dukungan kepada penulis selama proses perkuliahan dan penyusunan skripsi.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, penulis terbuka terhadap seluruh kritik dan saran dari para pembaca agar skripsi ini dapat menjadi lebih baik lagi. Penulis juga berharap agar skripsi ini dapat berguna bagi semua pembaca.

Bandung, Januari 2022

Penulis



# DAFTAR ISI

<b>KATA PENGANTAR</b>	<b>xv</b>
<b>DAFTAR ISI</b>	<b>xvii</b>
<b>DAFTAR GAMBAR</b>	<b>xix</b>
<b>DAFTAR TABEL</b>	<b>xxi</b>
<b>1 PENDAHULUAN</b>	<b>1</b>
1.1 Latar Belakang . . . . .	1
1.2 Rumusan Masalah . . . . .	2
1.3 Tujuan . . . . .	2
1.4 Batasan Masalah . . . . .	3
1.5 Sistematika Pembahasan . . . . .	3
<b>2 LANDASAN TEORI</b>	<b>5</b>
2.1 Variabel Acak Diskret . . . . .	5
2.2 Distribusi Poisson . . . . .	6
2.3 Regresi Linear . . . . .	8
2.3.1 Regresi Linear Sederhana . . . . .	8
2.3.2 Regresi Linear Majemuk . . . . .	9
2.4 Estimasi Parameter . . . . .	9
2.4.1 <i>Least Squares Estimation</i> . . . . .	9
2.4.2 <i>Maximum Likelihood</i> . . . . .	10
2.5 <i>Machine Learning</i> . . . . .	10
2.6 <i>Decision Tree</i> . . . . .	13
2.7 <i>Gradient Boosting</i> . . . . .	15
2.8 Deret Taylor . . . . .	16
<b>3 METODE PENELITIAN</b>	<b>17</b>
3.1 Model Linear Tergeneralisasi . . . . .	17
3.1.1 <i>Exponential Dispersion Model</i> . . . . .	17
3.1.2 Variabel Yang Digunakan . . . . .	19
3.1.3 Komponen GLM . . . . .	19
3.1.4 Estimasi Parameter GLM . . . . .	20
3.1.5 Frekuensi Klaim Dalam Model GLM . . . . .	21
3.1.6 Algoritma Penerapan GLM . . . . .	21
3.2 <i>XGBoost</i> . . . . .	21
3.2.1 Penerapan <i>XGBoost</i> . . . . .	22
3.2.2 <i>XGBoost</i> Secara Matematis . . . . .	22
3.2.3 Algoritma Pencarian Titik <i>Split</i> . . . . .	24
3.2.4 Algoritma Penerapan <i>XGBoost</i> pada Sistem Komputer . . . . .	25
3.3 Uji Perbandingan dan Kecocokan Model . . . . .	26

3.3.1	<i>Mean Squared Error</i> dan <i>Root Mean Squared Error</i> . . . . .	26
3.3.2	Deviansi Poisson . . . . .	26
3.3.3	<i>Akaike's Information Criterion (AIC)</i> . . . . .	27
<b>4</b>	<b>DATA DAN ANALISIS</b>	<b>29</b>
4.1	Penjelasan Data . . . . .	29
4.2	Pengolahan Data Awal . . . . .	30
4.3	Eksplorasi Data . . . . .	32
4.4	Hasil Eksperimen GLM Poisson . . . . .	35
4.4.1	Pengolahan data pada GLM . . . . .	35
4.4.2	Pemilihan Model GLM Poisson Terbaik . . . . .	35
4.4.3	Pemodelan Frekuensi Klaim dengan GLM Poisson 2 . . . . .	38
4.4.4	Prediksi Frekuensi Klaim Dengan Model GLM Poisson 1 . . . . .	38
4.5	Hasil Eksperimen <i>XGBoost</i> . . . . .	40
4.5.1	Pengolahan Data Pada <i>XGBoost</i> . . . . .	40
4.5.2	Pemodelan Frekuensi Klaim Dengan <i>XGBoost</i> . . . . .	41
4.5.3	Prediksi Frekuensi Klaim Dengan Model <i>XGBoost</i> . . . . .	41
4.6	Perbandingan Model Dalam Prediksi Frekuensi Klaim . . . . .	43
<b>5</b>	<b>KESIMPULAN DAN SARAN</b>	<b>45</b>
5.1	Kesimpulan . . . . .	45
5.2	Saran . . . . .	45
	<b>DAFTAR REFERENSI</b>	<b>47</b>

## DAFTAR GAMBAR

2.1	Perbandingan <i>Traditional Programming</i> Dengan <i>Machine Learning</i> Dalam Pemrograman . . . . .	11
2.2	Hasil Pemrograman dengan Metode <i>Traditonal Programming</i> . . . . .	12
2.3	Hasil Pemrograman dengan Metode <i>Machine Learning</i> . . . . .	12
2.4	Contoh Data Yang Terdapat Pada Ruang Fitur Dua Dimensi . . . . .	13
2.5	Contoh Pembagian Partisi Data Ke Dalam Empat <i>Region</i> . . . . .	14
2.6	Contoh <i>Tree</i> Yang Telah Didapat Dari Hasil <i>Split</i> Pada Partisi Data Pada Gambar 2.5 . . . . .	15
3.1	Contoh Penerapan <i>Default</i> Pada Proses <i>Sparsity Aware Split Finding</i> . . . . .	25
4.1	Gambar mengenai <i>Data Head</i> Dengan Menggunakan Bantuan Program <i>R</i> . . . . .	30
4.2	<i>Boxplot</i> Penyebaran <i>exposure</i> Pada Data . . . . .	31
4.3	Frekuensi Banyaknya Klaim . . . . .	32
4.4	Frekuensi <i>exposure</i> . . . . .	32
4.5	Frekuensi Variabel <i>Area</i> . . . . .	32
4.6	Frekuensi Variabel Usia Pengemudi . . . . .	32
4.7	Frekuensi Variabel Usia Kendaraan . . . . .	33
4.8	Frekuensi Variabel Kepadatan Penduduk . . . . .	33
4.9	Frekuensi Banyaknya Klaim Berdasarkan Tipe Bahan Bakar . . . . .	33
4.10	Frekuensi Banyaknya Klaim Berdasarkan Tenaga Kendaraan . . . . .	33
4.11	Pengujian Kalibrasi Variabel Prediktor Pada Model GLM . . . . .	40
4.12	Hasil <i>Tree</i> Yang Diperoleh Dengan Model <i>XGBoost</i> . . . . .	42
4.13	Variabel yang termasuk dalam <i>Feature Importance</i> Untuk Model <i>XGBoost</i> . . . . .	43

## DAFTAR TABEL

3.1	Tabel Fungsi Variansi . . . . .	18
3.2	Tabel Fungsi Link . . . . .	20
4.1	Banyak Klaim Pada Data Awal . . . . .	31
4.2	Hasil <i>Output</i> dari Model GLM Poisson . . . . .	34
4.3	Tabel Perbandingan Antara Model GLM Poisson 1 dan GLM Poisson 2 . . . . .	36
4.4	Hasil <i>Output</i> dari Model GLM Poisson 2 . . . . .	37
4.5	Hasil Proses <i>One-Hot Encoding</i> Pada Data Asuransi Kendaraan Bermotor . . . . .	41
4.6	Model Perbandingan GLM Poisson dengan <i>XGBoost</i> Untuk Data <i>Training</i> . . . . .	43
4.7	Model Perbandingan GLM Poisson dengan <i>XGBoost</i> Untuk Data <i>Testing</i> . . . . .	43





# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Transportasi menjadi salah satu sarana yang penting dalam kehidupan sehari-hari, terutama di wilayah Indonesia yang memiliki tingkat mobilitas yang tinggi. Tingginya mobilitas masyarakat dan kebutuhan akan kendaraan bermotor dapat memicu timbulnya risiko yang harus dihadapi oleh pengendara terutama dalam kehidupan sehari-hari, seperti kecelakaan lalu lintas, kerusakan pada bagian mesin akibat kemacetan, serta risiko lainnya. Survei dari *Integrated Road Safety Management System* (IRSMS) mencatat bahwa pada tahun 2020 terdapat sebanyak 100.028 kecelakaan terjadi di Indonesia dengan jumlah korban sebanyak 147.798 dan kerugian yang ditimbulkan mencapai sekitar 198,4 milyar rupiah<sup>1</sup>. Risiko tersebut tidak dapat dihindari oleh setiap pengendara, namun terdapat hal preventif yang dapat dilakukan, contohnya adalah dengan meminimalisir dan mengelola risiko yang mungkin dapat terjadi di kemudian hari. Salah satu upaya untuk menghindari berbagai risiko untuk permasalahan tersebut adalah asuransi.

Asuransi merupakan sebuah bentuk perlindungan secara finansial dengan mendistribusikan aset dari tertanggung (pemegang polis) untuk menutupi kerugian atas risiko yang terjadi. Jika terjadi kerugian, maka perusahaan akan menanggung biaya klaim asuransi tersebut. Salah satu produk asuransi yang paling banyak diklaim adalah asuransi kendaraan bermotor, seperti kerusakan badan atau fisik pada kendaraan secara menyeluruh, tabrakan, dan lainnya. Dalam asuransi kendaraan bermotor, besar biaya polis asuransi akan berbeda-beda pada setiap tertanggung sesuai dengan banyaknya klaim yang diajukan. Setiap klaim yang diajukan oleh tertanggung akan dihitung dalam suatu ukuran yang disebut dengan frekuensi klaim. Frekuensi klaim merupakan banyaknya klaim yang terjadi dalam suatu periode waktu tertentu yang dapat diperoleh dengan membagi jumlah klaim dengan periode exposurnya. Dengan demikian, frekuensi klaim adalah ukuran berbentuk *rate* yang merupakan perbandingan antara dua unit, yaitu banyaknya klaim dan eksposur. Dalam implementasi dunia asuransi, prediksi frekuensi klaim merupakan proses yang penting agar perusahaan asuransi dapat menentukan besar premi yang harus dibayar oleh masing-masing pemegang polis. Salah satu metode untuk memprediksi frekuensi klaim adalah dengan menggunakan *Machine Learning*.

Berdasarkan [1], *Machine Learning* adalah metode komputasi yang terdiri dari perancangan algoritma yang mampu mempelajari berbagai jenis data dan dapat melakukan prediksi yang akurat, dimana kualitas dan jumlah data sangat penting dalam keakuratan prediksi yang dihasilkan oleh *Machine Learning*. *Machine Learning* dapat mempermudah proses pembentukan model yang variabelnya banyak dan masing-masing variabelnya harus dipertimbangkan. Dalam *Machine Learning*, terdapat dua metode pendekatan dalam memprediksi frekuensi klaim pada asuransi kendaraan bermotor, yaitu model linear tergeneralisasi (*Generalized Linear Model* atau disingkat dengan GLM) dan *XGBoost*. Menurut Goldburd [2], GLM merupakan perluasan dari model regresi linear yang dapat digunakan pada kondisi dimana variabel respon dan variabel prediktornya tidak berhubungan secara linear, serta variabel responnya merupakan variabel dengan jenis distribusi diskret

---

<sup>1</sup>Alza Ahdira, "Kerugian Akibat Kecelakaan di Indonesia Tahun 2020 Capai Rp200 Miliar" <https://www.pikiran-rakyat.com/otomotif/pr-011372808/kerugian-akibat-kecelakaan-di-indonesia-tahun-2020-capai-rp200-miliar> (diakses pada 24 Januari 2022, pukul 19.55)

dan kontinu. Penggunaan GLM dapat dilakukan dengan mengasumsikan variabel prediktornya memiliki efek linear dan variabel responnya merupakan variabel yang memiliki distribusi yang termasuk dalam *Exponential Dispersion Model* (EDM). Distribusi yang dimaksud diantaranya adalah distribusi normal, distribusi Poisson, Gamma, Binomial Negatif, dan Binomial. GLM bertujuan untuk mengetahui hubungan sebab-akibat, serta pengaruh variabel prediktor terhadap variabel respon dan banyak digunakan dalam memodelkan penentuan besar premi asuransi karena mudah diinterpretasikan. Pembentukan model GLM dan penerapan distribusi di dalamnya tercatat pada [3], [4], dan [2]. Sementara itu, menurut Wade [5], *XGBoost* merupakan sebuah tipe pembelajaran mesin yang digunakan untuk berbagai masalah klasifikasi dan pemodelan regresi. *XGBoost* termasuk dalam tipe penerapan *Boosting* yang menggunakan konsep *Decision Tree*, dimana tipe pembelajaran *Boosting* melakukan prediksi dengan menggunakan kesalahan yang dibuat oleh pohon-pohon sebelumnya. *XGBoost* dapat menangani data dalam jumlah yang besar, dan di sisi lain mampu menghasilkan *output* dalam waktu yang relatif cepat berkat pengoptimalan algoritma *XGBoost* dalam mengolah data dan membentuk model.

Untuk dapat memodelkan prediksi frekuensi klaim yang berbentuk *rate*, maka distribusi yang dapat digunakan, baik untuk GLM maupun *XGBoost* adalah distribusi Poisson. Selain itu, menurut Goldburd [2], distribusi Poisson digunakan secara luas dalam ilmu aktuaria sebagai distribusi untuk penghitungan klaim. Berdasarkan alasan tersebut, maka pada skripsi ini akan dibahas mengenai penerapan GLM Poisson dan *XGBoost* dalam memodelkan frekuensi klaim pada data, serta melakukan perbandingan model yang lebih akurat. Dikarenakan adanya keterbatasan data mengenai asuransi kendaraan bermotor di Indonesia, maka produk asuransi yang digunakan pada skripsi ini menggunakan data dari perusahaan asuransi kendaraan bermotor di Perancis. Data tersebut memiliki variabel-variabel yang berhubungan, cocok, dan sesuai untuk diaplikasikan di Indonesia, terutama di pulau Jawa yang padat penduduknya.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dipaparkan dalam skripsi ini, rumusan masalah yang akan dikaji adalah:

1. Bagaimana penerapan GLM Poisson dalam prediksi frekuensi klaim pada data asuransi kendaraan bermotor ?
2. Variabel apa yang memengaruhi prediksi frekuensi klaim pada data asuransi kendaraan bermotor dalam model GLM Poisson ?
3. Bagaimana memprediksi variabel terpenting bagi data asuransi kendaraan bermotor dalam *XGBoost* ?
4. Model manakah yang paling akurat dalam memprediksi frekuensi klaim pada data asuransi kendaraan bermotor ?

## 1.3 Tujuan

Tujuan yang ingin dicapai dalam skripsi ini adalah:

1. Menganalisis dan menentukan prediksi frekuensi klaim dengan model GLM Poisson pada data asuransi kendaraan bermotor.
2. Mengidentifikasi variabel yang memengaruhi prediksi frekuensi klaim pada data asuransi kendaraan bermotor Perancis dalam model GLM Poisson.
3. Menentukan variabel terpenting bagi data asuransi kendaraan bermotor dalam *XGBoost*.
4. Menentukan dan menganalisis model yang paling akurat dalam memprediksi frekuensi klaim pada data asuransi kendaraan bermotor.



## 1.4 Batasan Masalah

Dalam membentuk prediksi model dalam skripsi ini, diperlukan adanya pembatasan masalah agar hasil yang diperoleh dapat lebih efektif. Adapun batasan masalah pada skripsi ini yaitu:

1. Tidak terjadi interaksi antar variabel prediktor yang digunakan.
2. Fungsi *link* yang digunakan pada model GLM Poisson diasumsikan menggunakan *log-link function*.
3. Pemilihan *base level* pada model GLM menggunakan variabel prediktor dengan jumlah klaim terbanyak.

## 1.5 Sistematika Pembahasan

Dalam skripsi ini, akan dijabarkan sistematika pembahasan yang terdiri dari:

### **Bab 1: Pendahuluan**

Bab ini menjelaskan mengenai latar belakang berupa penjelasan awal mengenai komponen yang akan dibahas dalam penulisan skripsi ini. Bab ini terdiri dari lima sub bab, yaitu latar belakang, rumusan masalah, tujuan, batasan masalah, dan sistematika pembahasan.

### **Bab 2: Landasan Teori**

Bab ini membahas teori-teori yang mendukung pengerjaan skripsi ini, seperti Variabel Acak Diskret, Distribusi Poisson, Regresi Linear Sederhana dan Majemuk, Estimasi Parameter, *Machine Learning*, *Decision Tree*, *Gradient Boosting*, dan Deret Taylor.

### **Bab 3: Metode Penelitian**

Bab ini membahas mengenai metode yang digunakan dan pembahasan lebih dalam mengenai hal-hal yang mendukung pengerjaan skripsi ini, diantaranya mengenai model linear tergeneralisasi. penjelasan mengenai algoritma pada *XGBoost*, dan metode dalam uji perbandingan dan kecocokan model.

### **Bab 4: Data dan Analisis**

Bab ini menjelaskan mengenai data asuransi kendaraan bermotor di Perancis yang digunakan secara utuh, penerapan model GLM Poisson, pembahasan dan analisa hasil estimasi frekuensi klaim, penerapan algoritma *XGBoost*, penentuan *variable importance*, serta hasil perbandingan antara model GLM dengan *XGBoost* dengan indikator deviasi Poisson dan RMSE.

### **Bab 5: Kesimpulan dan Saran**

Bab ini berisi kesimpulan yang diperoleh dari hasil di bab-bab sebelumnya, serta saran dan pengembangan untuk penelitian selanjutnya.