

SKRIPSI

**PREDIKSI CADANGAN KLAIM DENGAN DATA
INDIVIDUAL MENGGUNAKAN METODE XGBOOST DAN
GENERALIZED LINEAR MODEL**



MOHAMMAD ADMIRAL IBROCHIM

NPM: 6161901043

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

FINAL PROJECT

**PREDICTION OF CLAIMS RESERVES WITH INDIVIDUAL
DATA USING GENERALIZED LINEAR MODEL AND
XGBOOST METHOD**



MOHAMMAD ADMIRAL IBROCHIM

NPM: 6161901043

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

LEMBAR PENGESAHAN

PREDIKSI CADANGAN KLAIM DENGAN DATA INDIVIDUAL MENGUNAKAN METODE XGBOOST DAN GENERALIZED LINEAR MODEL

Mohammad Admiral Ibrochim

NPM: 6161901043

Telah lulus ujian skripsi pada 25 Januari 2024 dengan penguji:
Felivia Kusnadi, M.Act.Sc. dan Liem Chin, M.Si.

Bandung, 2 Februari 2024

Menyetujui,

Pembimbing

Maria Anestasia, M.Si., 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 CADANGAN KLAIM DENGAN DATA INDIVIDUAL
MENGUNAKAN METODE *GENERALIZED LINEAR MODEL* 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,



Mohammad Admiral Ibrochim
NPM: 6161901043

ABSTRAK

Perusahaan asuransi membantu pengalihan risiko finansial yang diderita oleh pemegang polis dengan cara memberikan ganti rugi. Untuk mengurangi risiko di pihak perusahaan asuransi dilakukan perhitungan cadangan klaim. Umumnya perhitungan tersebut dilakukan menggunakan data yang sudah diagregatkan. Metode yang paling umum untuk digunakan dalam menghitung cadangan klaim menggunakan data agregat adalah metode *Chain Ladder*. Namun, bentuk data agregat memiliki beberapa kekurangan, yaitu banyak informasi yang hilang dalam proses agregasi dan jenis klaim *Incurred but Not Reported* (IBNR) serta *Reported but Not Settled* (RBNS) tidak dapat dibedakan. Karena alasan tersebut, pada penelitian ini perhitungan cadangan klaim akan dilakukan menggunakan data klaim individual. Metode yang digunakan untuk melakukan perhitungan cadangan klaim adalah XGBoost dan *Generalized Linear Model* (GLM) berdistribusi Poisson dan Quasi Poisson. Metode GLM dipilih karena memberikan interpretasi yang relatif lebih langsung terhadap hubungan antara variabel respon dan variabel prediktor, sedangkan metode XGBoost dipilih karena metode ini mampu menangani data kompleks. Selain itu, sebagai tolak ukur perbandingan model, akan digunakan juga metode *Chain Ladder* untuk menghitung total uang yang harus dibayarkan untuk klaim RBNS yang diamati. Hasil yang diperoleh menggunakan *Generalized Linear Model* dan XGBoost akan dibandingkan dengan hasil dari metode *Chain Ladder*. Model yang memberikan hasil cadangan klaim terdekat dengan hasil metode *Chain Ladder* akan dipilih menjadi model terbaik. Hasil yang diperoleh pada penelitian ini adalah model C, yang merupakan penggabungan metode XGBoost dan GLM, yang merupakan model terbaik karena model ini memberikan hasil prediksi cadangan klaim yang paling mendekati tolak ukur.

Kata-kata kunci: *Generalized Linear Model*; *Chain Ladder*; XGBoost; estimasi cadangan klaim.

ABSTRACT

Insurance companies aid in transferring the financial risks experienced by policyholders by providing compensation. To mitigate risks on the insurer's side, claim reserve calculations are typically conducted, often using aggregated data. The most commonly employed method for calculating claim reserves with aggregated data is the Chain Ladder method. However, aggregated data has several drawbacks, such as the loss of information during the aggregation process, and the inability to distinguish between types of claims such as Incurred but Not Reported (IBNR) and Reported but Not Settled (RBNS). Therefore, this study conducts claim reserve calculations using individual claim data. The methods utilized for this calculation include XGBoost and the Generalized Linear Model (GLM) with Poisson and Quasi-Poisson distributions. GLM is chosen for its relatively direct interpretation of the relationship between response and predictor variables, while XGBoost is selected for its ability to handle complex data. Additionally, the Chain Ladder method is used as a benchmark for comparing model results, specifically in calculating the total amount payable for observed RBNS claims. The results obtained using the Generalized Linear Model and XGBoost are compared with those from the Chain Ladder method. The model that provides claim reserve predictions closest to the Chain Ladder benchmark is selected as the best model. The findings from this research indicate that model C, a combination of XGBoost and GLM, is the optimal model, as it yields claim reserve predictions that closely align with the benchmark.

Keywords: Generalized Linear Model; Chain Ladder; XGBoost; claim reserve estimation.

KATA PENGANTAR

Penulis ingin mengucapkan terima kasih banyak kepada Tuhan Yang Maha Esa atas segala berkat dan rahmat-Nya yang tiada hentinya mengalir, sehingga penulis dapat menyelesaikan skripsi yang berjudul “**Prediksi Cadangan Klaim dengan Data Individual Menggunakan Metode XGBoost dan Generalized Linear Model**”. Dalam penyusunan skripsi ini, tentunya penulis mendapatkan dukungan dan bimbingan dari berbagai pihak, sehingga skripsi ini dapat diselesaikan dengan baik. Oleh karena itu, penulis juga ingin menyampaikan rasa terima kasih kepada:

1. Bapak, Ibu, Adik, Kakak, dan Kinan selaku keluarga tercinta yang selalu memberikan doa, dukungan, dan semangat sehingga penulis dapat menyelesaikan skripsi ini.
2. Ibu Maria Anastasia, M.Si., M.Act.Sc. selaku pembimbing yang selalu memberi dorongan, arahan, dan bersedia membimbing dengan segala keterbatasan penulis. Keberhasilan penyusunan skripsi ini tidak terlepas dari bantuan dan dukungan yang luar biasa dari pembimbing, yang senantiasa memberikan pandangan kritis, masukan konstruktif, dan kesabaran dalam mengarahkan penulis.
3. Ibu Felivia Kusnadi, M.Act.Sc. dan Bapak Liem Chin, M.Si. selaku penguji, dan Bapak Dr. Daniel Salim selaku koordinator skripsi yang telah memberikan masukan untuk menyempurnakan skripsi ini.
4. Teman-teman penghuni kost The Maple Place Residence yang selalu menyempatkan waktu untuk memberikan distraksi ketika penulis sedang dalam kebuntuan. Terima kasih atas kebaikan dan kehangatan yang selalu dihadirkan, menjadikan setiap kesulitan terasa lebih ringan.
5. Teman-teman “Ansof Mat” (Aditya, Billy, Biqytofa, Dennis, Egha, Kevin, Luthfi, Michael, dan Timothy) yang saling memberikan arahan, masukan, dan dukungan selama proses penyusunan skripsi ini.
6. Ariliya Rosa, yang telah memberikan semangat, dukungan, dan waktu. Terima kasih telah menemani penulis selama proses penyelesaian skripsi ini.
7. Semua pihak yang tidak dapat disebutkan satu per satu dalam ucapan terima kasih ini, yang juga telah memberikan dukungan dan doa kepada penulis sehingga dapat menyelesaikan kewajibannya

Bandung, 2 Februari 2024

Penulis

DAFTAR ISI

KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
1.4 <i>State of the Art</i>	3
1.5 Batasan Masalah	3
2 LANDASAN TEORI	4
2.1 Cadangan Klaim	4
2.2 Metode <i>Chain-Ladder</i>	5
2.3 <i>Generalized Linear Model</i>	11
2.4 <i>Overdispersed Poisson</i>	14
2.5 <i>Mean Squared Error</i>	14
2.6 Pohon Keputusan	14
2.7 <i>Gradient Boosting</i>	15
3 METODE PENELITIAN	17
3.1 Karakteristik Data	17
3.2 Analisis Data Eksploratif	18
3.3 Model GLM Berdistribusi Poisson	21
3.4 XGBoost	23
4 PREDIKSI CADANGAN KLAIM	25
4.1 Prediksi Menggunakan <i>Chain Ladder</i>	25
4.2 Prediksi Menggunakan GLM	27
4.3 Prediksi Menggunakan XGBoost	31
4.4 Perbandingan Hasil Prediksi Model GLM dan XGBoost	31
5 KESIMPULAN DAN SARAN	34
5.1 Kesimpulan	34
5.2 Saran	34
DAFTAR REFERENSI	36
A HASIL EKSPERIMEN	38



DAFTAR GAMBAR

2.1	Proses penyelesaian klaim	4
2.2	Diagram pohon keputusan	15
3.1	Diagram pencar waktu terjadinya kecelakaan dalam satuan mingguan vs. waktu penundaan pelaporan dalam satuan mingguan untuk tipe klaim 5	19
3.2	Besar klaim dan waktu terjadinya kejadian tak terduga dalam satuan bulanan untuk klaim 5	19
3.3	<i>Boxplot</i> perbandingan total besar klaim yang dibayarkan berdasarkan masing-masing tipe klaim	20
3.4	Analisis penundaan pelaporan berdasarkan tipe klaim dan analisis penundaan pelaporan berdasarkan total besar klaim	21
3.5	Analisis penundaan penyelesaian berdasarkan tipe klaim dan analisis penundaan penyelesaian berdasarkan total besar klaim	21
A.1	Grafik waktu terjadinya kecelakaan dalam satuan mingguan dan waktu penundaan pelaporan dalam satuan mingguan	38
A.2	Besar klaim dan waktu terjadinya kejadian tak terduga dalam satuan bulanan	39

DAFTAR TABEL

2.1	Segitiga <i>run-off</i> inkremental	6
2.2	Contoh segitiga <i>run-off</i> inkremental	6
2.3	Segitiga <i>run-off</i> kumulatif	7
2.4	Contoh segitiga <i>run-off</i> kumulatif	7
2.5	Faktor penundaan individual segitiga <i>run-off</i>	8
2.6	Contoh faktor penundaan individual	8
2.7	Contoh periode penundaan	9
2.8	Estimasi segitiga <i>run-off</i> kumulatif	9
2.9	Contoh estimasi segitiga <i>run-off</i> kumulatif	10
2.10	Segitiga <i>run-off</i> penuh	10
2.11	Contoh segitiga <i>run-off</i> penuh	11
2.12	Fungsi <i>link</i> berdasarkan distribusinya	13
3.1	Keterangan variabel yang digunakan	17
3.2	Contoh beberapa baris data	18
4.1	Data segitiga <i>run-off</i> (dalam ribuan)	25
4.2	Data segitiga <i>run-off</i> kumulatif (dalam ribuan)	26
4.3	Hasil prediksi segitiga <i>run-off</i> kumulatif (dalam ribuan)	26
4.4	Hasil prediksi segitiga <i>run-off</i> inkremental (dalam ribuan)	27
4.5	Estimasi parameter GLM berdistribusi Poisson	28
4.6	Estimasi parameter GLM berdistribusi Quasi Poisson	29
4.7	Hasil prediksi menggunakan GLM	32
4.8	Hasil prediksi menggunakan GLM dengan fokus tipe klaim 2	32
4.9	Hasil prediksi cadangan klaim menggunakan XGBoost	33
B.1	Estimasi parameter GLM berdistribusi Poisson dengan fokus pada tipe klaim 2	40
B.2	Estimasi parameter GLM berdistribusi Quasi Poisson	40

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Asuransi adalah perjanjian atau kontrak antara pemegang polis dan perusahaan asuransi¹. Pemegang polis (tertanggung) mengalihkan risiko finansial atas suatu kejadian tak terduga dan sebagai imbalannya, tertanggung menerima manfaat (*benefit*) sesuai dengan kontrak yang berlaku. Sebagai kompensasi dari pengalihan risiko finansial tersebut, perusahaan asuransi (penanggung) meminta pembayaran berupa premi dari tertanggung. Asuransi sendiri dibagi menjadi dua, yaitu: asuransi jiwa atau *life insurance*, dan asuransi umum atau *non-life insurance*. Pada asuransi jiwa, perusahaan asuransi memberikan ganti rugi kepada ahli waris apabila pemegang polis meninggal dalam jangka waktu yang telah ditentukan dalam perjanjian. Pada asuransi umum, perusahaan asuransi memberikan jasa pertanggungan risiko berupa penggantian kerugian jika terjadi kerusakan, kehilangan keuntungan, atau biaya yang timbul yang mungkin diderita oleh pemegang polis atas suatu kejadian tak terduga.

Berdasarkan Johny et al. [1], klaim asuransi adalah suatu tuntutan yang dilakukan oleh pihak pemegang polis kepada perusahaan asuransi atas adanya kontrak perjanjian asuransi yang mengikat antar pihak dalam menjamin pembayaran ganti rugi apabila terjadi kejadian yang tak terduga yang dialami oleh pihak pemegang polis, di mana klaim dapat diajukan apabila premi telah dibayarkan oleh pihak pemegang polis. Agar tidak terjadi gagal bayar dari pihak perusahaan, perusahaan asuransi perlu melakukan perhitungan cadangan. Cadangan adalah sejumlah dana yang dapat digunakan untuk memenuhi kewajiban perusahaan asuransi kepada pemegang polis apabila pemegang polis sudah melakukan pelaporan klaim atas sebuah musibah. Salah satu jenis cadangan adalah cadangan klaim. Cadangan klaim adalah sejumlah uang yang harus disiapkan oleh perusahaan asuransi untuk membayar klaim yang akan datang.

Cadangan klaim dapat diestimasi menggunakan beberapa metode, salah satunya adalah menggunakan *Generalized Linear Model* (GLM). GLM merupakan salah satu metode yang sering digunakan untuk mengestimasi cadangan klaim. Terlebih lagi, bentuk data yang paling banyak digunakan untuk mengestimasi cadangan klaim adalah bentuk data agregat. Berdasarkan penelitian yang dilakukan Rhesa Mahardika [2], peneliti membandingkan nilai estimasi cadangan klaim dari data *worker's compensation* tahun 2009-2017 menggunakan metode GLM dan *bootstrap*. Pada penelitian tersebut disimpulkan bahwa untuk mengestimasi nilai cadangan klaim metode GLM lebih cocok untuk digunakan karena nilai *prediction error* yang dihasilkan dari metode GLM lebih kecil dibandingkan dengan metode *bootstrap*. Pada penelitian yang dilakukan oleh Jun Zhou dan José Garrido

¹<https://www.ojk.go.id/id/kanal/iknb/Pages/Asuransi.aspx> diakses pada 18 Oktober 2023

[3], peneliti membandingkan model pencadangan klaim menggunakan metode *Generalized Linear Model*, metode *Chain Ladder*, dan metode *Bornhuetter-Ferguson*. Pada penelitian ini, disimpulkan bahwa model pencadangan klaim dengan metode GLM memberikan estimasi cadangan klaim yang lebih stabil dan akurat dibandingkan dengan metode *Chain Ladder* dan *Bornhuetter-Ferguson*.

Perhitungan estimasi cadangan klaim umumnya menggunakan bentuk data yang sudah diagregatkan, namun kenyataannya bentuk data individual merupakan pilihan yang lebih baik untuk mengestimasi cadangan klaim. Beberapa kelebihan data individual, yaitu mempertahankan detail informasi dan jenis klaim IBNR dan RBNS dapat dibedakan [4]. Salah satu metode yang dapat digunakan untuk mengestimasi cadangan klaim menggunakan data individual adalah metode XGBoost. Pada penelitian yang dilakukan oleh Francis Duval dan Mathieu Pigeon [5], peneliti menggunakan data aktual dari suatu perusahaan dan membandingkan model pencadangan klaim menggunakan metode *Chain Ladder* untuk data agregat dengan metode GLM untuk data agregat, dan metode XGBoost dan GLM untuk data individual. Pada penelitian tersebut diperoleh kesimpulan bahwa metode XGBoost lebih baik dalam mengestimasi cadangan klaim karena hasil prediksi ekspektasi cadangan klaim yang dihasilkan oleh XGBoost lebih dekat dengan data yang diobservasi dibanding metode GLM dan *Chain Ladder*. Oleh karena itu, pada skripsi ini prediksi cadangan klaim RBNS akan dihitung menggunakan *Generalized Linear Model* dan XGBoost menggunakan data klaim individual. Pada skripsi ini akan dibuat tiga model, yaitu: model A, model B, dan model C untuk memprediksi cadangan klaim RBNS menggunakan metode XGBoost di mana model A berperan sebagai model acuan, model B digunakan untuk mengukur sejauh mana bias pemilihan, dan model C merupakan model yang dibuat dengan menggabungkan metode GLM dan XGBoost. Selain itu, akan dibandingkan juga nilai RMSE dari metode yang digunakan.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, beberapa masalah yang dikaji dalam makalah ini adalah

1. Bagaimana cara mengestimasi cadangan klaim RBNS dengan data klaim individual menggunakan metode *Chain Ladder*, GLM, dan XGBoost?
2. Model manakah yang paling baik untuk memprediksi cadangan klaim RBNS dengan data individual?

1.3 Tujuan

Tujuan yang ingin dicapai dalam makalah ini adalah

1. Mengestimasi cadangan klaim RBNS dengan data individual menggunakan metode *Chain Ladder*, GLM, dan XGBoost.
2. Menentukan model terbaik untuk memprediksi cadangan klaim RBNS dengan data individual.

1.4 *State of the Art*

Pada skripsi ini, penelitian dilakukan untuk memahami model XGBoost dan GLM yang diterapkan ke data klaim individual yang dibangkitkan, untuk memprediksi cadangan klaim RBNS. Penjelasan mengenai metode XGBoost didasari oleh artikel Tianqi Chen dan Carlos Guestrin [6], di mana pada artikel ini dijelaskan langkah-langkah penerapan metode XGBoost. Basis penggunaan data klaim individual pada penelitian ini didasari oleh artikel Jinlong Huang, Chunjuan Qiu, dan Xianyi Wu [7], di mana pada artikel ini dilakukan perbandingan dalam mengestimasi cadangan klaim menggunakan data klaim individual dan data klaim agregat. Penjelasan mengenai metode GLM didasari oleh artikel Jong-Hwan Yoo [8]. Penerapan metode GLM dan XGBoost untuk memprediksi cadangan klaim RBNS dengan data individual pada skripsi ini didasari oleh artikel yang ditulis oleh Francis Duval dan Mathieu Pigeon [5]. Untuk penerapan metode *Chain Ladder* kepada data yang dibangkitkan, skripsi ini mengacu pada artikel Andrea Gabrielli dan Mario V. Wuthrich [9].

1.5 Batasan Masalah

Dalam melakukan prediksi cadangan klaim, pada penelitian ini dilakukan pembatasan masalah. Pada penelitian ini perhitungan prediksi cadangan klaim menggunakan data klaim individual hanya dilakukan untuk jenis klaim RBNS. Oleh karena itu jenis klaim IBNR tidak dimasukkan dalam perhitungan.