

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Asuransi kendaraan bermotor merupakan salah satu jenis asuransi yang sudah umum diperjualbelikan. Asuransi kendaraan bermotor dapat melindungi pemegang polisnya secara finansial dari bencana tertentu seperti kebakaran, kecelakaan, pencurian, dan hal-hal lain yang terjadi pada kendaraan yang diasuransikan. Setiap perusahaan asuransi yang menjaminkan perlindungan finansial kepada para pemegang polisnya haruslah menentukan ekspektasi total klaim dan besar premium yang tepat. Pada konteks ini, total klaim adalah total kerugian yang akan ditanggung oleh perusahaan asuransi jika suatu bencana yang diasuransikan terjadi. Premium adalah besar nilai yang harus dibayarkan pemegang polis kepada perusahaan asuransi.

Ekspektasi total klaim dan besar premium perlu ditentukan sedemikian sehingga premium yang dibayarkan semua pemegang polis cukup besar untuk menjamin semua klaim yang akan terjadi dapat terbayarkan. Prinsip penentuan klaim ini disebut dengan istilah prinsip ekuivalen [1]. Selanjutnya, perusahaan dapat menentukan besar laba yang diharapkan. Dalam menentukan model, bisa saja menggunakan asumsi bahwa setiap kendaraan yang diasuransikan memiliki peluang yang sama untuk mengklaim. Maka dari itu, model yang digunakan bisa saja berbentuk

$$P = E \left[ \sum_{k=1}^n Y_k \right]. \quad (1.1)$$

Pada model di Persamaan (1.1), besar premium ( $P$ ) dimodelkan dengan menghitung ekspektasi dari jumlah total klaim yang dibayarkan untuk setiap kejadian ( $Y_k$ ) sebanyak  $n$  kejadian.  $Y_k$  bisa disebut sebagai total klaim annual. Model pada Persamaan (1.1) dapat kita sebut sebagai model naif. Model naif mengasumsikan setiap pemegang polis memiliki peluang yang sama untuk mengalami bencana. Jika model naif diterapkan pada semua asuransi, maka akan timbul suatu masalah.

Meskipun tujuan utama seseorang membeli asuransi adalah untuk mengantisipasi risiko finansial yang bisa terjadi, pada dasarnya tidak ada orang yang ingin mengalami bencana. Meskipun begitu, masing-masing pembeli

asuransi memiliki tingkat risikonya masing-masing. Pembeli asuransi yang lebih waspada dalam menghindari bencana akan dirugikan. Mereka harus membayar premium lebih mahal dari yang seharusnya. Untuk mengatasi masalah tersebut, akan dibuat model Bayesian yaitu

$$P(\theta_i) = E \left[ \sum_{k=1}^n Y_k \mid \theta_i \right]. \quad (1.2)$$

Untuk menggunakan model di Persamaan (1.2), perusahaan asuransi perlu melakukan suatu proses yang disebut sebagai *underwriting*. *Underwriting* adalah proses pengumpulan data pemegang polis / objek yang dijaminan [1]. Data-data yang dikumpulkan tersebut disebut sebagai *rating factor* ( $\theta$ ). *Rating factor* pemegang polis ke- $i$  dinotasikan sebagai  $\theta_i$ . *Rating factor* berpotensi untuk berkorelasi dengan frekuensi dan besar klaim masing-masing pemegang polis. Pada penelitian ini, *rating factor* yang akan digunakan adalah gender pengemudi, model kendaraan, daerah kendaraan dikemudikan, usia pengemudi, dan usia kendaraan.

Model pada Persamaan (1.2) dapat memberikan nilai premium yang lebih adil bagi setiap pemegang polis. Untuk mempermudah pemodelan, dapat diasumsikan bahwa  $Y_k$  untuk setiap  $k$  saling independen. Asumsi tersebut menyebabkan model di Persamaan (1.2) dapat dituliskan sebagai

$$P(\theta_i) = \sum_{k=1}^n E[Y_k \mid \theta_i]. \quad (1.3)$$

Selanjutnya, masing-masing nilai ekspektasi bisa dimodelkan dengan menggunakan model apapun. Salah satu model yang bisa dijadikan pilihan adalah *Artificial Neural Network* (ANN) dengan regularisasi. Keuntungan dari menggunakan ANN adalah model yang diciptakan lebih fleksibel karena terhindar dari asumsi distribusi. Selanjutnya, diterapkan juga regularisasi untuk mencegah terjadinya *overfitting*. Dengan melakukan pemodelan diatas, perusahaan asuransi dapat menetapkan premium yang tepat.

Pada penelitian ini, akan dikemukakan dua ide membuat pemodelan total klaim annual asuransi kendaraan bermotor. Ide pertama berupa cara mencari model ANN yang optimal dengan memodifikasi model ANN secara perlahan hingga mendapatkan model dengan biaya yang lebih kecil. Ide kedua berupa cara

memodelkan GLM dengan menggunakan model ANN sehingga berbagai macam model GLM dapat diselesaikan secara paralel.

Ide pertama merupakan hasil pengembangan *greedy algorithm*. *Greedy algorithm* dapat digunakan untuk mencari model yang lebih optimal. Namun, tidak jarang *greedy algorithm* berhenti pada nilai optimal lokal. Solusi untuk mencegah hal ini adalah dengan menambah tingkat kedalaman pada pencarian *greedy algorithm*. Perlu diketahui bahwa meskipun hal ini lebih baik dari segi akurasi, proses pencaharian akan memakan waktu yang lebih lama.

Dengan menerapkan cara pencarian baru ini, diharapkan proses pencarian dapat dilakukan dengan kedalaman yang lebih tinggi tetapi durasi pencarian tidak memakan waktu yang lebih lama. Hal ini dikarenakan algoritma baru tidak akan mengevaluasi seluruh model dalam suatu tingkat kedalaman, tetapi model akan mencari beberapa model saja. Nilai dari tingkat kedalaman pencaharian juga akan berubah secara dinamis.

Ide kedua berasal dari pernyataan bahwa model ANN adalah generalisasi dari model GLM. Maka dari itu, model GLM dapat dinyatakan, diselesaikan, dan dievaluasi layaknya model ANN. Akibat yang paling signifikan adalah model GLM dengan struktur yang serupa dapat diselesaikan secara paralel.

Pada pemodelan ini, data yang digunakan berasal dari SUSEP (Superintendência de Seguros Privados), lembaga yang berwenang meregulasi industri asuransi dan reasuransi di Brazil[2]. Data yang digunakan berjumlah 1.965.355 datum. Diharapkan, pemodelan ini dapat memberikan pandangan baru pada cara pemodelan asuransi. Dengan demikian, perusahaan-perusahaan asuransi kendaraan bermotor dan setiap pemilik kendaraan bermotor dapat mematuhi Hukum No. 6194, 19 Desember 1974 yang mewajibkan pemilik kendaraan bermotor untuk memiliki asuransi kendaraan bermotor[3].

## 1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana cara mengoptimasi model ANN (*Artificial Neural Network*) dalam memodelkan ekspektasi total klaim annual asuransi kendaraan bermotor di Brazil?
2. Bagaimana arsitektur model ANN yang tepat dalam memodelkan ekspektasi total klaim annual asuransi kendaraan bermotor di Brazil?

3. Bagaimana cara mentransformasi GLM sebagai GLMNN (*Generalized Linear Model as Neural Network*) dalam memodelkan ekspektasi total klaim annual asuransi kendaraan bermotor di Brazil?
4. Bagaimana perbandingan APOE (*Absolute Percentage Overall Error*) model ANN terhadap GLMNN pada data validasi?

### 1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengemukakan algoritma EvoNN (*Evolving Neural Network*) untuk mengoptimalkan model ANN dalam memodelkan ekspektasi total klaim annual asuransi kendaraan bermotor di Brazil.
2. Menentukan arsitektur model ANN berupa jumlah *hidden layer*, jumlah neuron, dan besar regularisasi yang tepat dalam memodelkan ekspektasi total klaim annual asuransi kendaraan bermotor di Brazil.
3. Mentransformasi GLM sebagai GLMNN dalam memodelkan ekspektasi total klaim annual asuransi kendaraan bermotor di Brazil.
4. Menganalisis perbandingan APOE dari model ANN terhadap model GLMNN pada data validasi.

### 1.4 Batasan Masalah

Batasan masalah dan asumsi yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini akan dilakukan dengan menggunakan perangkat lunak Visual Studio Code.
2. Pemodelan ANN dan GLMNN akan dilakukan dengan menggunakan bahasa pemrograman Python 3.11.4 dengan bantuan paket Tensorflow 2.14.0.
3. Data yang digunakan berasal dari R *package*, yaitu 'CASdatasets' dengan nama dataset 'brvehins1a', 'brvehins1b', 'brvehins1c', 'brvehins1d', dan 'brvehins1e'.

4. Populasi dari penelitian ini adalah klaim asuransi kendaraan bermotor yang terjadi di Brazil pada tahun 2011.
5. Variabel prediktor kategorikal, yaitu gender pengemudi, model kendaraan, dan daerah kendaraan dikemudikan akan digunakan dalam pemodelan.
6. Variabel prediktor numerik yang digunakan, yaitu usia pengemudi dan usia kendaraan.
7. Variabel respon numerik, yaitu total klaim annual untuk peristiwa pencurian, kebararan, bertabrakan total, bertabrakan parsial, dan lain-lain.
8. Variabel respon diasumsikan saling bebas.
9. Seluruh nilai acak yang dihasilkan berasal dari 'random.seed(110)'
10. Setiap *hidden layer* yang dimodelkan adalah *dense layer* dengan paling sedikit satu neuron.
11. Setiap *hidden layer* yang dimodelkan dan *output layer* diaktivasi dengan suatu fungsi aktivasi tertentu.
12. Regularisasi yang digunakan adalah lasso (L1) dan ridge (L2).
13. Pemodelan ANN dan GLMNN menggunakan asumsi pemodelan jangka pendek.

## 1.5 Manfaat Penelitian

Manfaat teoretis dan praktis dari penelitian ini adalah sebagai berikut:

### 1.5.1 Manfaat Teoretis

1. Dapat memahami algoritma EvoNN untuk mengoptimalkan model ANN dalam memodelkan ekspektasi total klaim annual asuransi kendaraan bermotor di Brazil.
2. Dapat memahami jumlah *hidden layer*, jumlah neuron, dan besar regularisasi yang tepat pada model ANN dalam memodelkan ekspektasi total klaim annual asuransi kendaraan bermotor di Brazil.

3. Dapat memahami transformasi GLM sebagai GLMNN dalam memodelkan ekspektasi total klaim annual asuransi kendaraan bermotor di Brazil.
4. Dapat memahami perbandingan APOE dari model ANN terhadap model regresi linier berganda pada data validasi.

### 1.5.2 Manfaat Praktis

1. Dapat mengoptimalkan model ANN dalam memodelkan ekspektasi total klaim annual pemegang polis asuransi kendaraan bermotor berdasarkan *rating factor* pemegang polisnya.
2. Dapat menghitung ekspektasi total klaim annual pemegang polis asuransi kendaraan bermotor berdasarkan *rating factor* pemegang polisnya.
3. Dapat menganjurkan metode baru dalam menentukan ekspektasi total klaim annual pemegang polis asuransi kendaraan bermotor berdasarkan *rating factor* pemegang polisnya.

### 1.6 Sistematika Penulisan

Sistematika penulisan dari penelitian ini adalah sebagai berikut:

#### 1. **BAB I - PENDAHULUAN**

Pada Bab I, dijelaskan latar belakang penelitian, rumusan masalah, tujuan penelitian, batasan masalah, serta manfaat penelitian secara teoretis dan praktis. Sebagai penutup, akan dijelaskan sistematika penulisan penelitian.

#### 2. **BAB II - LANDASAN TEORI**

Pada Bab II, dijelaskan seluruh teori meliputi istilah, definisi, notasi, dan teorema matematika yang digunakan pada penelitian ini. Teori yang akan dijabarkan meliputi model ANN, model GLM, fungsi biaya yang digunakan, dan penyelesaian parameter model. Selain itu, penelitian-penelitian lain yang relevan dengan penelitian ini juga akan dibahas.

#### 3. **BAB III - METODOLOGI**

Pada Bab III, dijelaskan langkah-langkah penelitian untuk mencapai tujuan penelitian. Langkah-langkah tersebut disajikan dalam bentuk diagram alir dan penjelasan terperinci.

#### 4. **BAB IV - ANALISIS DAN PEMBAHASAN**

Pada Bab IV, dijelaskan hasil penelitian dan analisis dari hasil penelitian. Setiap hasil akan dibahas secara terperinci. Melalui analisis ini, setiap rumusan masalah akan dijawab.

#### 5. **BAB V - KESIMPULAN DAN SARAN**

Pada Bab V, dijelaskan kesimpulan dari hasil penelitian dan analisis dari hasil penelitian. Kelebihan penelitian, kekurangan penelitian, dan rekomendasi yang dapat digunakan untuk penelitian selanjutnya.

