

SKRIPSI

**PREDIKSI DAN ANALISIS TINDAKAN *FRAUD* KLAIM
ASURANSI KENDARAAN DENGAN METODE ANALISIS
DISKRIMINAN LINEAR DAN KUADRATIK**



JASON ORLANDO

NPM: 6161901026

**PROGRAM STUDI MATEMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2023**

FINAL PROJECT

**PREDICTION AND ANALYSIS OF FRAUDS IN AUTO
INSURANCE CLAIMS USING LINEAR AND QUADRATIC
DISCRIMINANT ANALYSIS**



JASON ORLANDO

NPM: 6161901026

**DEPARTMENT OF MATHEMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2023**

LEMBAR PENGESAHAN

PREDIKSI DAN ANALISIS TINDAKAN *FRAUD* KLAIM ASURANSI KENDARAAN DENGAN METODE ANALISIS DISKRIMINAN LINEAR DAN KUADRATIK

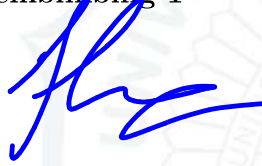
Jason Orlando

NPM: 6161901026

Bandung, 19 Agustus 2023

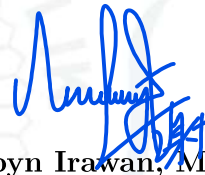
Menyetujui,

Pembimbing 1



Felivia Kusnadi, M.Act.Sc.

Pembimbing 2



Robyn Irawan, M.Sc.

Ketua Penguji



Jonathan Hoseana, Ph.D.

Anggota Penguji



Rizky Reza Fauzi, D.Phil.Math.

Mengetahui,

Ketua Program Studi



Dr. Livia Owen

PERNYATAAN

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

PREDIKSI DAN ANALISIS TINDAKAN *FRAUD* KLAIM ASURANSI KENDARAAN DENGAN METODE ANALISIS DISKRIMINAN LINEAR DAN KUADRATIK

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,
19 Agustus 2023

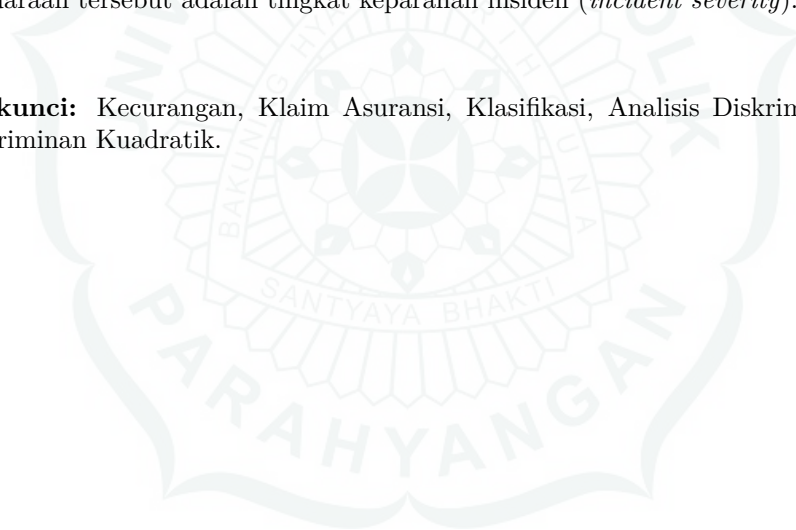


Jason Orlando
NPM: 6161901026

ABSTRAK

Tindak kecurangan klaim asuransi dapat menyebabkan perusahaan asuransi mengalami kerugian. Oleh karena itu, perusahaan asuransi perlu mencegah terjadinya tindak kecurangan klaim dengan cara memprediksi klaim-klaim yang mengindikasikan terjadinya tindak kecurangan. Pada penelitian ini, data klaim asuransi kendaraan roda empat di Amerika Serikat yang memuat informasi tindak kecurangan klaim diklasifikasikan menggunakan metode analisis diskriminan linear dan kuadrat. Kedua model yang diperoleh dibandingkan menggunakan *confusion matrix*, *receiver operating characteristic (ROC)*, dan *area under the curve (AUC)*. Dengan bantuan perangkat lunak RStudio, model analisis diskriminan linear ditemukan memiliki tingkat akurasi, presisi, spesifisitas, dan *F1-score* yang lebih baik dibandingkan model analisis diskriminan kuadrat. Variabel yang memberikan pengaruh paling besar pada tindak kecurangan klaim asuransi kendaraan tersebut adalah tingkat keparahan insiden (*incident severity*).

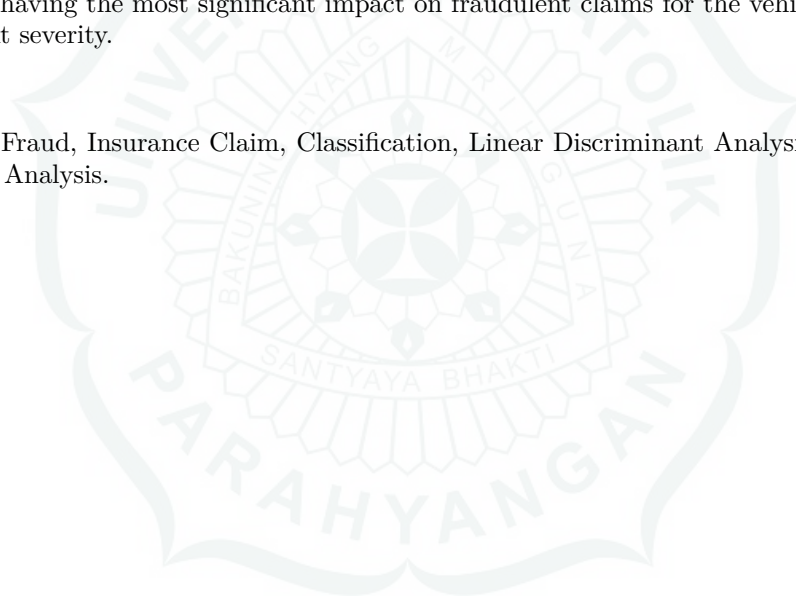
Kata-kata kunci: Kecurangan, Klaim Asuransi, Klasifikasi, Analisis Diskriminan Linear, Analisis Diskriminan Kuadrat.



ABSTRACT

Insurance claim frauds could lead to financial losses for insurance companies. Accordingly, insurance companies need to prevent occurrences of fraudulent claims by predicting claims that indicate fraudulent activities. In this research, we use the linear and quadratic discriminant analysis classification methods to classify data on fraudulent claims for a four-wheeled vehicle insurance in the United States. We compare the two resulting models using a confusion matrix, receiver operating characteristic (ROC), and area under the curve (AUC). By using RStudio software, we found that the linear discriminant analysis model demonstrates higher accuracy, precision, specificity, and F1-score rate compared to the quadratic discriminant analysis model. The variable having the most significant impact on fraudulent claims for the vehicle insurance is the incident severity.

Keywords: Fraud, Insurance Claim, Classification, Linear Discriminant Analysis, Quadratic Discriminant Analysis.



Untuk Papa & Mama Tercinta



KATA PENGANTAR

Puji syukur kepada Tuhan yang Maha Esa, karena atas berkatNya, penulis dapat menyelesaikan skripsi dengan judul “Prediksi dan Analisis Tindakan *Fraud* Klaim Asuransi Kendaraan dengan Metode Analisis Diskriminan Linear dan Kuadratik”. 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.

Dalam proses pembuatan skripsi ini, penulis ingin menyampaikan terima kasih kepada:

1. Papa, mama, dan cici yang selalu memberikan doa, semangat, dukungan, dan nasihat kepada penulis.
2. Ibu Felivia Kusnadi, M.Act.Sc. dan Bapak Robyn Irawan, M.Sc. selaku dosen pembimbing yang telah memberikan saran dan nasihat dari awal proses skripsi sehingga penulis dapat menyelesaikan skripsi dengan baik.
3. Bapak Jonathan Hoseana, Ph.D. selaku ketua penguji dan Bapak Rizky Reza Fauzi, D.Phil.Math. selaku anggota penguji yang telah memberikan saran dan kritik sehingga penyusunan skripsi ini menjadi lebih baik.
4. Bapak Dr. Daniel Salim selaku koordinator skripsi yang senantiasa membantu proses penyusunan skripsi sehingga dapat berjalan dengan lancar.
5. Bapak Dr. Julius Dharma Lesmonoselaku dosen wali yang telah memberikan bantuan dan arahan sejak awal perkuliahan.
6. Seluruh dosen FTIS khususnya dosen Program Studi Matematika yang telah memberikan ilmu dan pengetahuan selama proses perkuliahan.
7. Seluruh staf Tata Usaha FTIS yang telah memberikan bantuan dan arahan dalam bidang administrasi selama masa perkuliahan.
8. Reinard sebagai sahabat yang selalu memberikan dukungan dan bantuan serta telah berjuang bersama sejak SMA hingga akhir masa perkuliahan.
9. Ryu sebagai sahabat sejak SMA yang selalu mendukung dan menghibur selama berjuang bersama di Universitas Katolik Parahyangan.
10. Irena, Igna, Novaldy, dan Ivan selaku sahabat yang selalu memberikan dukungan dan motivasi selama masa perkuliahan.
11. Nathan dan Vincent selaku teman pares yang selalu menemani dan menghibur sejak awal perkuliahan.
12. Teman-teman jurusan matematika angkatan 2019 atas kebersamaan, kerja sama, dan bantuannya selama perkuliahan.
13. Semua pihak yang telah berjasa selama masa perkuliahan dan penyusunan skripsi.

Penulis menyadari bahwa skripsi ini masih memiliki kekurangan. Oleh karena itu, penulis terbuka terhadap kritik dan saran agar skripsi ini dapat menjadi lebih baik. Akhir kata, penulis berharap semoga skripsi ini dapat bermanfaat bagi pembaca.

Bandung, 19 Agustus 2023

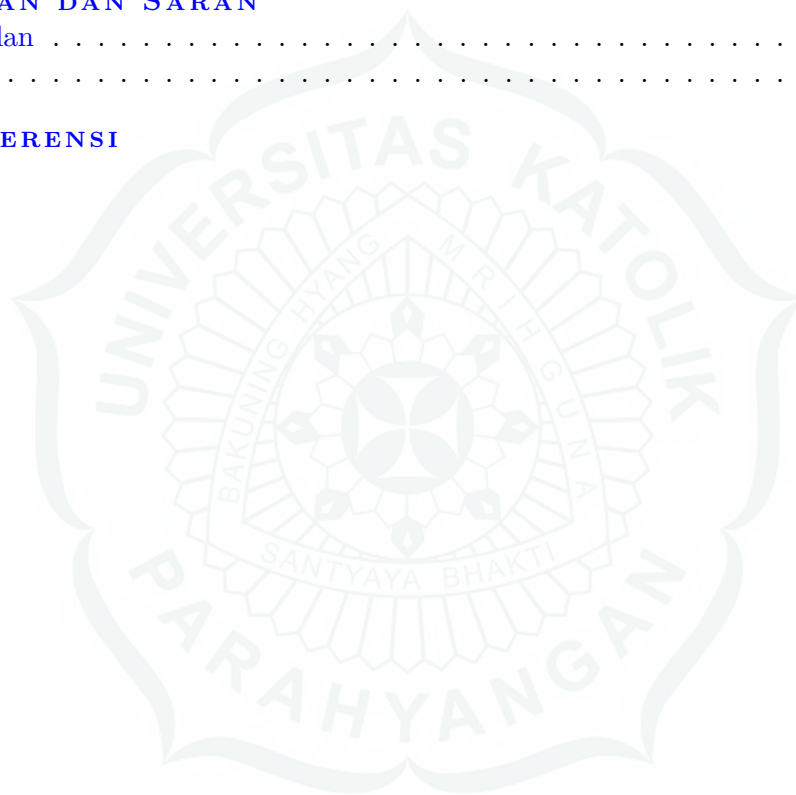
Penulis



DAFTAR ISI

KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
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
2 LANDASAN TEORI	4
2.1 <i>Fraud</i> dalam Asuransi	4
2.2 <i>Machine Learning</i>	5
2.3 Standardisasi Data	5
2.4 Teorema Bayes	6
2.5 Matriks Kovariansi	6
2.6 Analisis Diskriminan Linear dan Kuadratik untuk Kelas Biner	7
2.6.1 <i>Decision Boundary</i>	7
2.6.2 Fungsi Diskriminan Kuadratik	8
2.6.3 Fungsi Diskriminan Linear	8
2.6.4 Estimasi Kelas Suatu Observasi	9
2.6.5 Koefisien Analisis Diskriminan Linear	9
2.7 <i>Confusion Matrix</i>	10
2.7.1 Akurasi	10
2.7.2 Presisi	10
2.7.3 Sensitivitas	10
2.7.4 Spesifisitas	11
2.7.5 <i>F1-Score</i>	11
2.8 <i>Receiver Operating Characteristic (ROC)</i> dan <i>Area Under the Curve (AUC)</i>	11
3 ANALISIS DISKRIMINAN LINEAR DAN KUADRATIK	13
3.1 Algoritma Analisis Diskriminan Linear	13
3.2 Algoritma Analisis Diskriminan Kuadratik	14
3.3 Contoh Numerik Analisis Diskriminan Linear dan Kuadratik	14
3.3.1 <i>Data Dummy</i>	14
3.3.2 Penghitungan Numerik <i>Data Dummy</i> Menggunakan Analisis Diskriminan Linear dan Kuadratik	15
3.4 Diagram Alir Metode Penelitian	21

4	HASIL DAN PEMBAHASAN	23
4.1	Pengolahan Data	23
4.1.1	Pembersihan Data	23
4.1.2	Analisis Data Eksploratif	24
4.1.3	Deskripsi Data	25
4.1.4	Transformasi Data	26
4.1.5	Pembagian Data Latih dan Data Uji	27
4.2	Penerapan Model Pada Data	27
4.2.1	Estimasi Parameter	27
4.2.2	Penerapan Model Analisis Diskriminan Linear Pada Data	28
4.2.3	Penerapan Model Analisis Diskriminan Kuadratik Pada Data	29
4.3	Perbandingan Model	31
4.4	<i>Variable Importance</i>	31
5	KESIMPULAN DAN SARAN	32
5.1	Kesimpulan	32
5.2	Saran	32
	DAFTAR REFERENSI	33



DAFTAR GAMBAR

2.1	Peluang kedua kategori kelas ekuivalen di x^*	7
2.2	Kurva ROC	12
3.1	<i>Plot</i> prediksi analisis diskriminan linear	18
3.2	<i>Plot</i> prediksi analisis diskriminan kuadratik	20
3.3	Diagram alir metode penelitian	22
4.1	<i>Bar plot</i> frekuensi kedua kategori kelas dari variabel respon	24
4.2	<i>Box plot</i> variabel bebas <i>bodily injuries</i> terhadap variabel respon	24
4.3	<i>Box plot</i> variabel bebas <i>incident hour of the day</i> terhadap variabel respon	24
4.4	<i>Box plot</i> variabel bebas <i>months as customer</i> terhadap variabel respon	24
4.5	<i>Box plot</i> variabel bebas <i>policy deductible</i> terhadap variabel respon	25
4.6	<i>Box plot</i> variabel bebas <i>umbrella limit</i> terhadap variabel respon	25
4.7	<i>Box plot</i> variabel bebas <i>number of vehicles involved</i> terhadap variabel respon	25
4.8	<i>Bar plot</i> variabel bebas <i>incident type</i> terhadap variabel respon	25
4.9	<i>Plot</i> kurva ROC model analisis diskriminan linear	29
4.10	<i>Plot</i> kurva ROC model analisis diskriminan kuadratik	30

DAFTAR TABEL

2.1	<i>Confusion matrix</i>	10
3.1	Contoh data latih	14
3.2	Contoh data uji	15
3.3	Hasil klasifikasi data latih menggunakan analisis diskriminan linear	18
3.4	Hasil klasifikasi data latih menggunakan analisis diskriminan kuadratik	20
4.1	Keterangan variabel	26
4.2	Lima observasi pertama dari data (1)	26
4.3	Lima observasi pertama dari data (2)	26
4.4	Lima observasi pertama dari data setelah dilakukan standardisasi (1)	27
4.5	Lima observasi pertama dari data setelah dilakukan standardisasi (2)	27
4.6	<i>Prior probability</i> masing-masing kategori kelas	27
4.7	<i>Mean</i> masing-masing kategori kelas	28
4.8	<i>Confusion matrix</i> dari model analisis diskriminan linear terhadap data latih	28
4.9	<i>Confusion matrix</i> dari model analisis diskriminan linear terhadap data uji	29
4.10	Nilai akurasi, presisi, sensitivitas, spesifisitas, dan <i>F1-Score</i> model analisis diskriminan linear	29
4.11	<i>Confusion matrix</i> dari model analisis diskriminan kuadratik terhadap data latih	29
4.12	<i>Confusion matrix</i> dari model analisis diskriminan kuadratik terhadap data uji	30
4.13	Nilai akurasi, presisi, sensitivitas, spesifisitas, dan <i>F1-Score</i> model analisis diskriminan kuadratik	30
4.14	Perbandingan model analisis diskriminan linear dan kuadratik	31
4.15	Koefisien dari model analisis diskriminan linear	31

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Dewasa ini tindakan *fraud* (kecurangan) di seluruh dunia semakin sering terjadi. Menurut *Association of Certified Fraud Examiners (ACFE)*, *fraud* merupakan suatu tindak kriminal yang dilakukan secara sengaja oleh suatu pihak dengan tujuan memperoleh keuntungan sehingga menyebabkan kerugian bagi pihak lain¹. Tindakan *fraud* sering terjadi di seluruh dunia, misalnya korupsi, pencucian uang, dan penggelapan pajak, dan didorong oleh beberapa faktor seperti kondisi ekonomi, sifat keserakahan, dan lemahnya hukum. Tentu setiap individu maupun kelompok tidak ingin menjadi korban dari tindakan *fraud* karena tidak ingin mengalami kerugian. Tindakan *fraud* ini bisa terjadi di mana saja, termasuk di industri asuransi.

Tindakan *fraud* juga menjadi masalah umum yang dihadapi perusahaan asuransi, yang biasanya dilakukan oleh pihak tertanggung atau pemegang polis. Namun, tidak tertutup kemungkinan bahwa tindakan *fraud* dilakukan oleh pihak agen asuransi [1, hal. 2]. Bentuk *fraud* dalam asuransi dapat berupa penyembunyian fakta untuk memperkecil biaya premi hingga rekayasa klaim dengan tujuan memperoleh keuntungan dari uang pertanggungan yang tidak seharusnya. Tindakan *fraud* dapat terjadi di setiap sektor asuransi, termasuk sektor asuransi kendaraan bermotor. Beberapa contoh tindakan *fraud* yang sering ditemukan dalam asuransi kendaraan adalah klaim kecelakaan atau kerusakan yang direncanakan, tindakan melebih-lebihkan biaya perbaikan kendaraan, dan pemalsuan data [2, hal. 159–163]. Pada skripsi ini akan dibahas tindakan *fraud* yang terjadi pada klaim asuransi kendaraan bermotor di Amerika Serikat.

Polis asuransi yang diterbitkan oleh suatu perusahaan asuransi tentu sangat banyak. Semakin banyak polis asuransi yang diterbitkan, semakin banyak juga tindakan *fraud* yang mungkin terjadi, sehingga dapat terjadi kerugian yang cukup besar bagi perusahaan asuransi. Untuk mengatasi tindakan *fraud* yang sudah terjadi, perlu dibentuk divisi yang bertanggung jawab untuk menyelidiki kebenaran klaim-klaim yang diajukan. Namun, akan lebih baik jika tindakan *fraud* dapat dicegah dengan cara memprediksi klaim-klaim asuransi kendaraan yang mengindikasikan terjadinya tindakan *fraud*. Oleh karena itu, sangat penting bagi perusahaan asuransi untuk dapat memprediksi jenis klaim yang benar dan jenis klaim yang merupakan tindakan *fraud*.

Suatu perusahaan asuransi tentunya memiliki nasabah yang sangat banyak, sehingga tidak memungkinkan bagi perusahaan untuk melakukan pemeriksaan secara personal karena akan menghabiskan banyak waktu dan biaya. Oleh karena itu, penelitian ini bertujuan untuk mengetahui kontribusi faktor-faktor yang memengaruhi terjadinya tindakan *fraud*, dengan menggunakan metode

¹ <https://www.acfe.com/fraud-resources/fraud-101-what-is-fraud>

klasifikasi yang merupakan bagian dari studi pembelajaran mesin (*machine learning*). Penggunaan metode ini bertujuan untuk memprediksi jenis klaim yang mengindikasikan terjadinya tindakan *fraud* dengan memanfaatkan variabel-variabel bebas yang mungkin menjadi faktor-faktor terjadinya tindakan *fraud*. Dengan demikian, metode klasifikasi ini diharapkan dapat membantu suatu perusahaan asuransi untuk memperkecil risiko kerugian akibat tindakan *fraud*.

Pada penelitian-penelitian sebelumnya, beberapa metode klasifikasi sudah pernah diterapkan untuk mengklasifikasi tindakan *fraud* klaim asuransi, salah satunya adalah metode regresi logistik [3, hal. 1]. Namun, metode regresi logistik dianggap kurang stabil apabila setiap kategori kelas respon terpisah dengan baik [4, hal. 138]. Selain regresi logistik, metode klasifikasi *K-Nearest Neighbors* (KNN) juga pernah digunakan untuk mengklasifikasi data asuransi kendaraan untuk mendeteksi tindakan-tindakan *fraud* [5, hal. 4]. Namun, pada penelitian tersebut, metode KNN menghasilkan performa yang kurang baik dalam mendeteksi tindakan *fraud* klaim asuransi kendaraan.

Pada tahun 2018, telah dilakukan penelitian yang membandingkan metode KNN dengan metode analisis diskriminan linear. Hasil dari penelitian ini menunjukkan bahwa metode analisis diskriminan linear memiliki tingkat akurasi yang lebih baik daripada KNN [6, hal. 3]. Metode analisis diskriminan linear memiliki beberapa kelebihan, yaitu memiliki performa yang lebih baik apabila sebaran dari variabel-variabel bebas pada data mendekati distribusi normal untuk setiap kategori kelas, serta dapat menentukan variabel-variabel bebas yang berperan penting dalam proses klasifikasi [4, hal. 138]. Pada skripsi ini akan digunakan metode analisis diskriminan linear dan kuadratik untuk mengklasifikasi tindakan *fraud* dalam asuransi kendaraan.

Penelitian ini menggunakan data klaim asuransi kendaraan roda empat di Amerika Serikat yang memuat informasi tindak kecurangan klaim. Skripsi ini menggunakan data yang bukan dari Indonesia karena kurangnya sumber data mengenai kecurangan klaim asuransi kendaraan roda empat di Indonesia. Data yang digunakan pada skripsi ini juga mempunyai cukup banyak variabel yang dapat dimanfaatkan, sehingga dapat memaksimalkan penelitian.

1.2 Rumusan Masalah

Berikut rumusan masalah-masalah yang dibahas pada skripsi ini.

1. Bagaimana cara mendeteksi tindakan *fraud* klaim asuransi kendaraan di Amerika Serikat?
2. Bagaimana perbandingan performa model analisis diskriminan linear dan analisis diskriminan kuadratik dalam mengklasifikasikan tindak *fraud* klaim asuransi kendaraan di Amerika Serikat?
3. Variabel apakah yang memberikan pengaruh paling besar terhadap tindakan *fraud* klaim asuransi kendaraan di Amerika Serikat?

1.3 Tujuan

Tujuan dari penelitian ini adalah:

1. mendeteksi tindakan *fraud* klaim asuransi kendaraan di Amerika Serikat menggunakan analisis diskriminan linear dan kuadratik;

2. membandingkan performa model analisis diskriminan linear dan analisis diskriminan kuadratik dalam mengklasifikasikan tindak *fraud* klaim asuransi kendaraan di Amerika Serikat menggunakan *confusion matrix*, *receiver operating characteristic* (ROC), dan *area under the curve* (AUC) dari masing-masing model;
3. menentukan variabel yang memberikan pengaruh paling besar terhadap tindakan *fraud* klaim asuransi kendaraan di Amerika Serikat berdasarkan koefisien setiap variabel bebas yang dihasilkan oleh model analisis diskriminan linear.

1.4 *State of the Art*

Pada penelitian-penelitian sebelumnya, beberapa metode seperti regresi logistik dan *K-Nearest Neighbors* (KNN) banyak digunakan untuk mengklasifikasi tindak *fraud* klaim asuransi [3, 5]. Pada skripsi ini akan digunakan metode yang berbeda, yaitu analisis diskriminan linear, serta versi kuadratnya yaitu analisis diskriminan kuadratik. Salah satu kelebihan metode analisis diskriminan linear adalah dapat menghasilkan koefisien masing-masing variabel bebas yang berguna untuk menentukan variabel yang berperan penting dalam pembentukan model [7, hal. 105]. Oleh karena itu, akan dianalisis juga variabel yang memberikan pengaruh paling besar terhadap tindakan *fraud* klaim asuransi kendaraan roda empat di Amerika Serikat.

Pada skripsi ini juga akan diperlihatkan grafik *partition plot* yang merupakan visualisasi dari klasifikasi menggunakan analisis diskriminan linear dan kuadratik. Melalui grafik ini juga dapat diperlihatkan perbedaan yang signifikan antara model analisis diskriminan linear dan kuadratik. Terdapat beberapa metode evaluasi model yang banyak digunakan seperti tingkat akurasi, presisi, sensitivitas, spesifisitas, dan *F1-score* [6, 8]. Pada skripsi ini akan ditambahkan juga metode evaluasi model menggunakan kurva ROC dan nilai AUC.