

SKRIPSI

**IMPLEMENTASI METODE ANALISIS KOMPONEN UTAMA
DAN REGRESI LINEAR BERGANDA DALAM
MEMPREDIKSI HARGA JUAL MOBIL BEKAS**



STEPHANIE STELLA CHRISDIANNATA

NPM: 6161901014

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

FINAL PROJECT

**IMPLEMENTATION OF PRINCIPAL COMPONENT
ANALYSIS AND MULTIPLE LINEAR REGRESSION
METHODS FOR PREDICTING THE SALE PRICE OF USED
CARS**



STEPHANIE STELLA CHRISDIANNATA

NPM: 6161901014

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

LEMBAR PENGESAHAN

IMPLEMENTASI METODE ANALISIS KOMPONEN UTAMA DAN REGRESI LINEAR BERGANDA DALAM MEMPREDIKSI HARGA JUAL MOBIL BEKAS

Stephanie Stella Chrisdiannata

NPM: 6161901014

Bandung, 04 Agustus 2023

Menyetujui,

Pembimbing 1

Pembimbing 2



Dr. Livia Owen



Rizky Reza Fauzi, D.Phil.Math.

Ketua Penguji

Anggota Penguji



Maria Anastasia, M.Si., M.Act.Sc.



Liem Chin, M.Si.

Mengetahui,

Ketua Program Studi



Dr. Livia Owen

PERNYATAAN

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

IMPLEMENTASI METODE ANALISIS KOMPONEN UTAMA DAN REGRESI LINEAR BERGANDA DALAM MEMPREDIKSI HARGA JUAL MOBIL BEKAS

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,
04 Agustus 2023



Stephanie Stella Chrisdiannata
NPM: 6161901014

ABSTRAK

Industri otomotif merupakan salah satu sektor ekonomi yang memiliki dampak yang luas pada ekonomi global, baik dari segi pertumbuhan ekonomi, penciptaan lapangan kerja, maupun perkembangan teknologi. Dalam pasar otomotif, pasar mobil bekas telah menjadi pasar yang semakin kompetitif dan dinamis. Penentuan harga jual beli mobil bekas memerlukan beberapa pertimbangan untuk memperoleh keuntungan yang tepat dengan harga jual yang sesuai di harga pasar. Dalam hal ini, akan digunakan pendekatan statistika untuk menentukan harga jual yang sesuai dengan kondisi mobil dan dalam rentang harga yang sesuai dengan harga pasar. Penelitian ini akan berfokus pada penentuan harga jual mobil bekas yang dianalisis menggunakan metode Analisis Komponen Utama (PCA) dan Regresi Linear Berganda. Data yang digunakan terdiri dari tiga juta data mobil bekas yang terdaftar di salah satu pasar otomotif *online* di Amerika Serikat. Proses analisis dimulai dengan melakukan pembersihan data. Kemudian dilanjutkan dengan mencari variabel-variabel penting, atau komponen utama, yang merupakan hasil reduksi dimensi menggunakan metode analisis komponen utama. Dengan analisis komponen utama, dimensi data berhasil direduksi dan dapat dijamin telah memenuhi asumsi non-multikolinearitas, sehingga proses analisis dapat dilanjutkan dengan mengonstruksi model prediksi harga jual yang melibatkan komponen penting menggunakan metode regresi linear berganda. Diperoleh suatu model prediksi yang cukup baik, tetapi belum menjadi model terbaik karena terdapat salah satu asumsi regresi yang belum terpenuhi, yaitu homoskedastisitas.

Kata-kata kunci: penentuan harga jual; mobil bekas; analisis komponen utama; regresi linear berganda.

ABSTRACT

The automotive industry is one of the most important economic sectors in the world and has a broad impact on the global economy, in terms of economic growth, job creation and technological development. In the automotive market, used car sales have become an increasingly competitive and dynamic market. Determination of the buying and selling price of a used car requires several considerations are needed to get the right price that is in accordance with market prices. In this case, a statistical approach will be used to determine the selling price according to the condition of the car and matches the market price. This study will focus on determining the selling price of used cars which are analyzed using Principal Component Analysis (PCA) and Multiple Linear Regression methods. The data that used in analysis consists of three million used car that registered in one of the online automotive markets in the United States. The analysis process begins by cleaning the data. Then proceed with looking for the important variables, which called principal components, which are the result of dimension reduction using the principal component analysis method. With principal component analysis, the dimensions of the data have been reduced and can be guaranteed to meet the assumption of non-multicollinearity, so the analysis process can be continued by constructing a selling price prediction model with the significant component using multiple linear regression methods. A fairly good predictive model is obtained, but it is not the best model yet because there is one of the unfulfilled regression assumptions, called homoscedasticity.

Keywords: price prediction; used cars; principal component analysis; multiple linear regression.

KATA PENGANTAR

Puji dan syukur dipanjatkan kepada Tuhan Yang Maha Esa, atas berkat dan rahmat-Nya penulis diberi kesempatan untuk dapat berproses dalam menyelesaikan skripsi ini. Skripsi berjudul “Implementasi Metode Analisis Komponen Utama dan Regresi Linear Berganda dalam Memprediksi Harga Jual Mobil Bekas” merupakan salah satu syarat dalam menyelesaikan Strata-1 Program Studi Matematika di Fakultas Teknologi Informasi dan Sains, Universitas Katolik Parahyangan, Bandung.

Dalam penyusunan skripsi ini, penulis mendapatkan berbagai bantuan dan dukungan dari berbagai pihak, sehingga penulis mampu menyelesaikan penulisannya. Oleh karena itu, penulis ingin mengucapkan terima kasih yang tulus kepada:

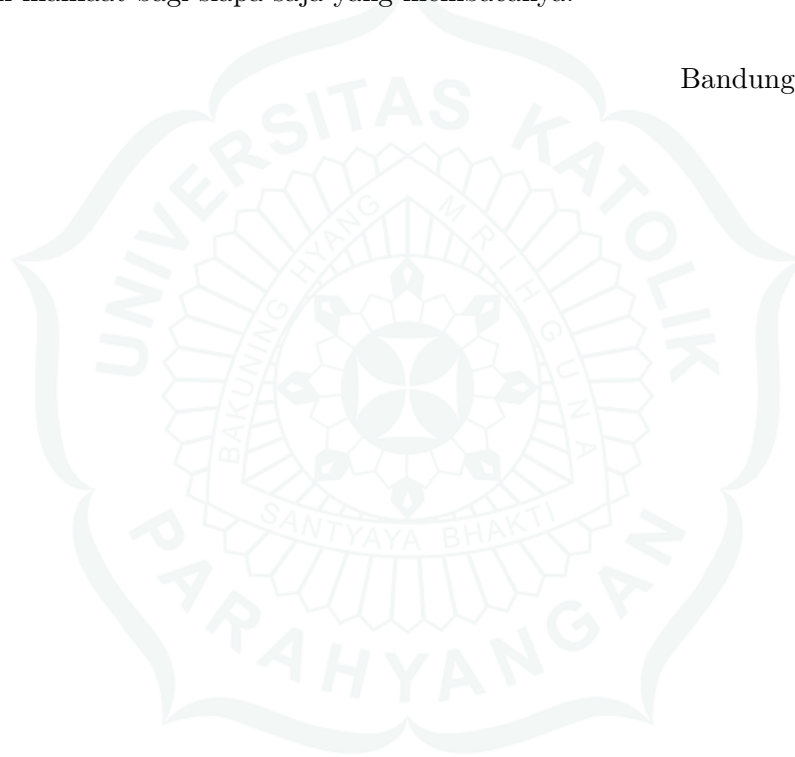
- Keluarga yang senantiasa memberikan doa dan dukungan dalam proses penyusunan skripsi.
- Bapak Rizky Reza Fauzi, D.Phil.Math. selaku dosen pembimbing pilihan yang senantiasa memberikan saran topik, ilmu, dorongan, arahan, dan kesabaran dalam membimbing penulis menyelesaikan skripsi.
- Ibu Dr. Livia Owen selaku dosen pembimbing yang senantiasa memberikan ilmu, dorongan, arahan, dan kesabaran dalam membimbing penulis menyelesaikan skripsi.
- Bapak Dr. Andreas Parama Wijaya selaku Kepala Laboratorium Pemodelan dan Simulasi yang senantiasa memberikan akses, sarana prasarana, dan saran yang diperlukan dalam pengolahan data skripsi.
- Bapak Liem Chin, M. Si. selaku dosen penguji yang telah memberikan kritik dan saran dalam pengembangan skripsi.
- Ibu Maria Anastasia, M. Si., MActSc. selaku dosen penguji yang telah memberikan kritik dan saran dalam pengembangan skripsi.
- Bapak Agus Sukmana, M. Sc. selaku dosen mata kuliah Statistika Multivariat yang senantiasa memberikan ilmu dan wawasan tambahan terkait topik skripsi.
- Bapak Dr. Daniel Salim selaku koordinator skripsi yang telah memberikan arahan dalam proses penyusunan skripsi.
- Bapak Janto Vincent Sulungbudi, S. Si. selaku dosen mata kuliah *Machine Learning* dalam Fisika yang telah memberikan ilmu dan wawasan tambahan terkait topik skripsi.
- Seluruh dosen FTIS, khususnya dosen Program Studi Matematika yang telah memberikan ilmu dan wawasan selama masa perkuliahan.
- Seluruh staf Tata Usaha FTIS yang telah memberikan bantuan administrasi selama masa perkuliahan.
- Alexander S, Verra Andriani, dan Aspira Rahmadini sebagai teman dekat yang senantiasa mendengarkan, memberikan hiburan, dukungan, semangat, dan saran dalam proses penyusunan skripsi ini.

- Liona Cecilia dan Vanessa Vindy Limarcia yang senantiasa memberikan semangat dan dukungan dalam proses penyusunan skripsi.
- Angel Wijaya yang senantiasa mendengarkan dan bertukar pikiran sebagai teman seperjuangan jurusan matematika dari universitas berbeda.
- Yustinus Onny, Bambang Saputra, Rusdy, Rosadi Rusdi, Noviandani Prastyo, Hadi Purboyo, dan Muhammad Aru Agam yang senantiasa memberikan semangat, penghiburan, dan doa selama proses penyusunan skripsi.
- Semua pihak lain yang tidak dapat disebutkan satu per satu yang telah memberikan dukungan, doa, penghiburan, dan membantu dalam proses penyusunan skripsi.

Penulis menyadari bahwa penulisan skripsi ini masih jauh dari kata sempurna karena keterbatasan ilmu dan pengalaman yang dimiliki penulis. Oleh karena itu, penulis menerima berbagai kritik dan saran yang membangun untuk dapat menyempurnakan skripsi ini. Akhir kata, semoga skripsi ini dapat memberikan manfaat bagi siapa saja yang membacanya.

Bandung, 04 Agustus 2023

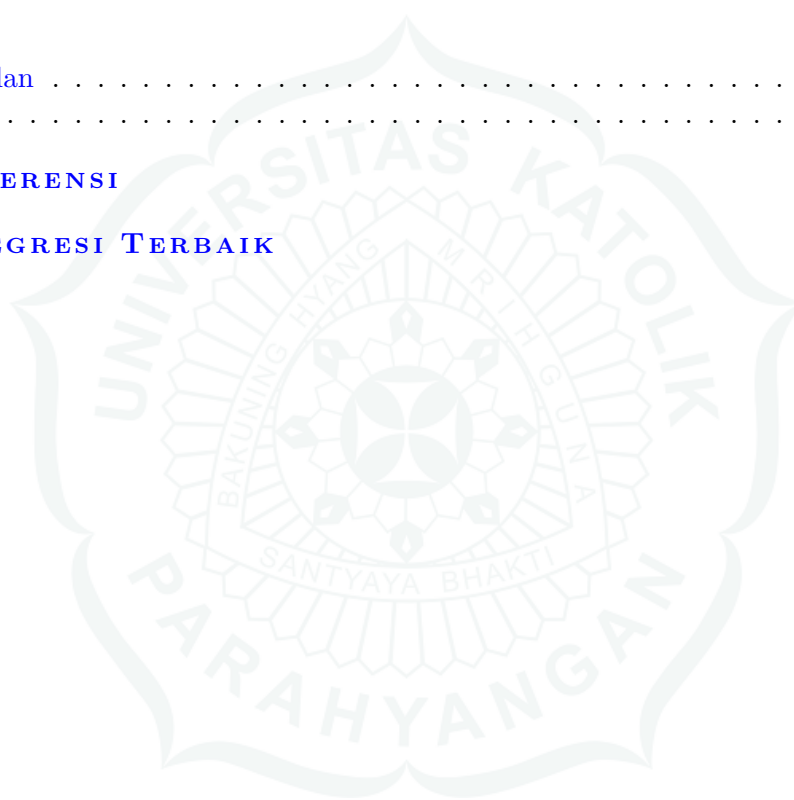
Penulis



DAFTAR ISI

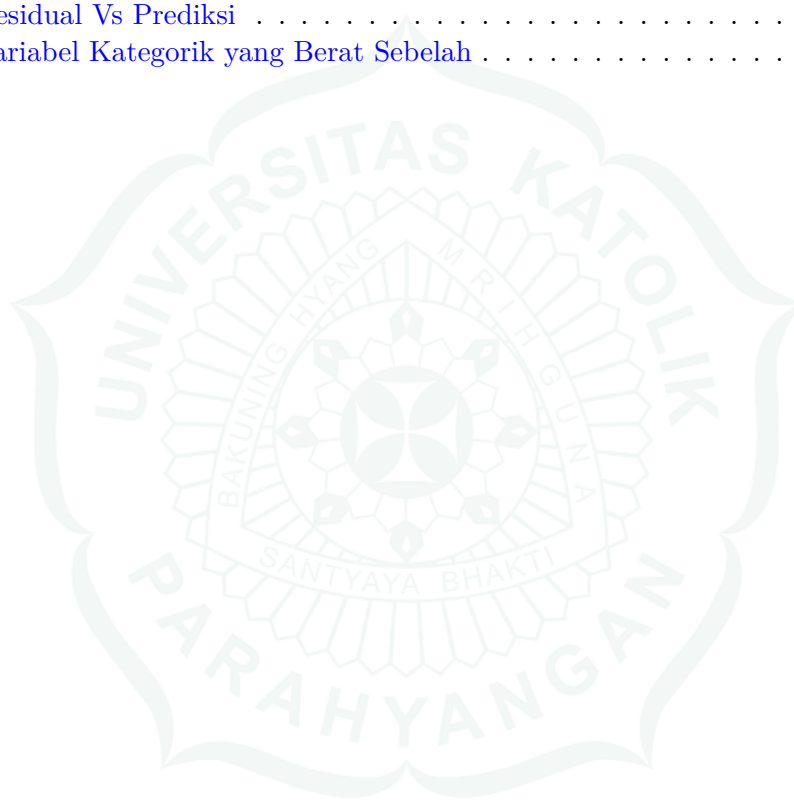
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
1.4 <i>State of the Art</i>	2
1.5 Sistematika Pembahasan	3
2 LANDASAN TEORI	4
2.1 Kegiatan Pemasaran	4
2.1.1 Elemen Harga dalam <i>Marketing Mix</i>	5
2.1.2 <i>Digital Marketing</i>	5
2.2 Transformasi Data	6
2.2.1 Standardisasi	6
2.2.2 Matriks Variansi Kovariansi	7
2.2.3 Nilai Eigen dan Vektor Eigen	9
2.3 Analisis Komponen Utama	10
2.3.1 Pendekatan Geometris Komponen Utama	11
2.3.2 Kriteria Pemilihan Komponen Utama	13
2.3.3 Korelasi antara Variabel Awal dengan Komponen Utama	15
2.4 Regresi Linear Berganda	16
2.5 Uji Asumsi Regresi Linear	17
2.5.1 <i>Akaike Information Criterion</i> (AIC)	18
2.5.2 Asumsi Non-Multikolinearitas	18
2.5.3 Asumsi Homoskedastisitas	18
2.5.4 Asumsi Normalitas	19
2.5.5 Asumsi Non-Autokorelasi	19
2.5.6 <i>Mean Squared Error</i> (MSE)	19
3 METODE PENELITIAN	21
3.1 Data	21
3.2 Prosedur Pengerjaan	24
4 HASIL DAN PEMBAHASAN	27
4.1 Pengolahan Data PCA	27
4.1.1 Persiapan Data	27

4.1.2	Analisis Komponen Utama	28
4.1.3	Evaluasi PCA	29
4.1.4	<i>Loadings</i>	31
4.2	Pengolahan Data Regresi Linear Berganda	32
4.2.1	Persiapan Data	32
4.2.2	Pengujian Model Prediksi	32
4.2.3	Uji Asumsi Normalitas	33
4.2.4	Uji Asumsi Non-Autokorelasi	33
4.2.5	Uji Asumsi Multikolinearitas	33
4.2.6	Uji <i>Mean Squared Error</i> (MSE)	34
4.3	Interpretasi Model Regresi	34
4.4	Contoh Prediksi	39
4.5	Evaluasi Lanjutan	40
4.5.1	Uji Asumsi Homoskedastisitas	40
5	PENUTUP	44
5.1	Kesimpulan	44
5.2	Saran	45
	DAFTAR REFERENSI	46
	A MODEL REGRESI TERBAIK	48



DAFTAR GAMBAR

2.1	Contoh grafik <i>scree</i>	14
4.1	Peta Korelasi Antar Variabel Bebas	28
4.2	<i>Scree Plot</i>	30
4.3	Peta Non-Multikolinearitas	34
4.4	Grafik Residual Vs Prediksi	41
4.5	Grafik Variabel Kategorik yang Berat Sebelah	42



DAFTAR TABEL

3.1	Data Numerik	21
3.2	Data Numerik (Lanjutan Tabel 3.1)	22
3.3	Data Non-Numerik	23
4.1	Variansi Masing-masing Komponen Utama	29
4.2	<i>Loadings</i> PCA	31
4.3	Label Penamaan Komponen Utama	31
4.4	Variabel Bebas Kategori Tipe Badan Mobil	35
4.5	Variabel Bebas Kategori Kepemilikan Bisnis atau Pemerintah	35
4.6	Variabel Bebas Kategori Kerusakan Fisik	35
4.7	Variabel Bebas Kategori Tipe Bahan Bakar	36
4.8	Variabel Bebas Kategori Riwayat Kecelakaan	36
4.9	Variabel Bebas Kategori Riwayat Angkutan Umum	36
4.10	Variabel Bebas Kategori Keluaran Baru	37
4.11	Variabel Bebas Kategori Merk Mobil	37
4.12	Variabel Bebas Kategori <i>Maximum Seating</i>	38
4.13	Variabel Bebas Kategori Riwayat Kerusakan Parah	38
4.14	Variabel Bebas Kategori Jenis Transmisi	38
4.15	Variabel Bebas Kategori Sistem Penggerak Roda	39
4.16	Spesifikasi Mobil Bekas Simulasi 1 dan 3	40
4.17	Spesifikasi Mobil Bekas Simulasi 4	41
4.18	Spesifikasi Mobil Bekas Simulasi 4 (Lanjutan Tabel 4.17)	42
A.1	Model Regresi Terbaik	48
A.2	Model Regresi Terbaik (Lanjutan Tabel A.1)	49

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Secara global, industri otomotif menjadi salah satu aspek penting dalam sektor manufaktur. Berdasarkan data dari situs IbisWorld¹, pendapatan per kapita global dalam industri otomotif naik 1,2% secara tahunan selama lima tahun hingga 2021. Selain itu, dari banyaknya unit otomotif, khususnya unit mobil yang terjual di seluruh dunia, terjadi peningkatan dalam beberapa tahun terakhir berdasarkan data dari *Statista Research Department*². Banyaknya unit yang terjual tersebut menggambarkan kebutuhan masyarakat yang cukup tinggi akan penggunaan mobil, khususnya mobil pribadi. Pada praktiknya, mobil dapat memenuhi kebutuhan penggunanya seperti memberikan kemudahan dalam berpergian, misal pergi bekerja, sekolah, liburan, atau bahkan untuk hal-hal darurat seperti pergi ke rumah sakit. Selain itu, masyarakat juga telah memiliki kemudahan dalam memperoleh satu unit mobil di era digital ini, yaitu melalui pasar otomotif *online*, di mana pengguna dapat mencari dan memilih tipe, warna, spesifikasi, hingga harga mobil hanya dengan mengakses internet.

Dalam pasar otomotif, terdapat berbagai bentuk layanan transaksi jual-beli mobil, mulai dari mobil baru, mobil bekas, layanan sewa mobil, dan sebagainya. Tidak sedikit masyarakat yang telah memiliki mobil berkeinginan untuk membeli mobil varian terbaru, sehingga memberikan pilihan bagi penggunanya untuk menyimpan mobil lamanya atau menjualnya terlebih dahulu sebelum menggantinya dengan mobil yang baru. Mobil bekas dapat dijual melalui lelang, jasa bursa mobil bekas, ataupun secara pribadi. Dalam hal ini, penjual memiliki tantangan untuk menentukan harga jual yang sesuai, karena banyaknya faktor pertimbangan yang perlu disesuaikan.

Pada penelitian ini, dilakukan analisis untuk menentukan harga jual mobil bekas agar dapat memperoleh laba yang sesuai, tetapi tidak jauh dari harga pasar. Dengan mempertimbangkan banyaknya faktor yang dapat menjadi penentu harga jual, digunakan metode analisis komponen utama atau *Principal Component Analysis* (PCA) untuk memilih komponen mana saja yang memiliki pengaruh besar terhadap harga jualnya, sehingga dimensi data dapat tereduksi. Berdasarkan penelitian [1], analisis komponen utama dapat mengekstrak informasi dari sejumlah besar variabel. Hasil analisisnya dapat membantu proses pengambilan keputusan dari kegiatan pemasaran, khususnya dalam analisis kepuasan pelanggan. Kemudian, analisis akan dilanjutkan dengan menggunakan metode regresi linear, khususnya regresi linear berganda, untuk menentukan harga jual mobil bekas berdasarkan komponen penting yang diperoleh sebelumnya. Pada penelitian [2], dikatakan bahwa

¹<https://www.ibisworld.com/>

²<https://www.statista.com/statistics/200002/international-car-sales-since-1990/>

metode regresi linear berganda mampu mengestimasi harga jual mobil bekas dengan beberapa variabel bebas.

Dalam penelitian ini, digunakan data harga jual mobil bekas dari salah satu pasar otomotif *online* di Amerika Serikat, yaitu CarGurus. Dengan menggunakan data tersebut, dilakukan implementasi metode analisis komponen utama dan metode regresi linear berganda. Proses analisis dimulai dengan melakukan pembersihan data, lalu mencari variabel-variabel penting yang merupakan komponen utama hasil reduksi dimensi menggunakan metode analisis komponen utama, dan komponen utama tersebut digunakan dalam mengonstruksi suatu model prediksi menggunakan metode regresi linear berganda. Dengan dimensi yang lebih rendah, diharapkan proses konstruksi model dapat dilakukan dengan lebih sederhana tanpa menghilangkan informasi data awal secara signifikan. Hasil konstruksi model tersebut diharapkan mampu memberikan prediksi harga jual mobil bekas yang baik, yakni cukup memberikan keuntungan bagi penjual tanpa menyimpang terlalu jauh dari harga pasar.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan sebelumnya, rumusan masalah yang bersesuaian adalah sebagai berikut

1. Bagaimana cara kerja metode analisis komponen utama dalam mereduksi dimensi data?
2. Bagaimana peran metode analisis komponen utama dalam memenuhi syarat regresi?
3. Bagaimana prediksi harga jual mobil bekas dari model yang dikonstruksi menggunakan metode regresi linear berganda dan dilengkapi dengan komponen utama?

1.3 Tujuan

Berikut dipaparkan tujuan penulisan penelitian ini, yaitu

1. Mengonstruksi komponen utama sebagai suatu cara untuk mereduksi dimensi data yang sangat besar.
2. Mengonstruksi komponen utama, sehingga syarat regresi seperti asumsi non-multikolinearitas dijamin dapat terpenuhi.
3. Mengevaluasi model regresi linear yang digunakan untuk prediksi harga jual mobil bekas jika komponen utama berperan sebagai variabel bebas.

1.4 *State of the Art*

Metode analisis komponen utama (PCA) memiliki kelebihan dalam mereduksi dimensi data, sehingga proses analisis selanjutnya menjadi lebih sederhana karena variabel yang dilibatkan menjadi lebih sedikit. Selain itu, metode ini juga mampu menjamin data yang digunakan tidak bersifat multikolinearitas, artinya masing-masing komponen utama terjamin tidak memiliki korelasi satu sama lain. Di samping kelebihan yang dimiliki, metode PCA memiliki kekurangan, yaitu terdapat informasi yang hilang akibat reduksi itu sendiri. Meskipun demikian, informasi yang hilang masih

dapat diminimalisasi. Selain itu, komponen utama yang dikonstruksi merupakan kombinasi linear dari variabel asli, sehingga dapat memberikan kesulitan dalam kegiatan interpretasi [3]. Untuk menanggulangi hal tersebut, dilakukan penamaan pada masing-masing komponen utama sesuai nilai *loadings* yang dihasilkan dari proses ini.

Metode regresi linear berganda memiliki kelebihan dalam memberikan kemudahan implementasi dan interpretasi, karena konsep dasarnya adalah persamaan linear yang mudah dipahami [4]. Namun, metode ini memiliki kekurangan, yakni penggunaannya yang hanya terbatas untuk mengakomodasi hubungan linear. Selain itu, metode ini perlu memenuhi berbagai asumsi yang terkadang sulit untuk dipenuhi, antara lain asumsi non-multikolinearitas, homoskedastisitas, non-autokorelasi, dan normalitas. Terdapat berbagai pengujian untuk memeriksa masing-masing asumsi tersebut, di mana jika ada yang tidak terpenuhi, artinya diperlukan perlakuan khusus, misalnya dengan transformasi ataupun menggunakan metode lain. Metode ini juga dapat dikatakan cukup sensitif terhadap pencilan, karena dapat mengakibatkan hasil prediksi yang keliru.

Dalam skripsi ini, untuk menjamin terpenuhinya asumsi non-multikolinearitas pada model regresi, dilakukan analisis komponen utama pada variabel bebas numerik sebelum dilanjutkan ke analisis regresi. Hal ini dapat dilakukan karena sifat dari komponen-komponen utama yang terbentuk dapat dipastikan saling bebas linear [5]. Dengan demikian, diharapkan model regresi yang terkonstruksi memberikan hasil yang optimal tanpa harus menghilangkan variabel numerik manapun.

1.5 Sistematika Pembahasan

penelitian ini akan membahas mengenai:

- **Bab 1: Pendahuluan** Bab ini berisi latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, dan sistematika pembahasan.
- **Bab 2: Landasan Teori** Bab ini berisi landasan teori terkait *digital marketing* dan transformasi data.
- **Bab 3: Metode Penelitian** Bab ini berisi data dan prosedur pengerjaan yang digunakan dalam skripsi ini.
- **Bab 4: Hasil dan Pembahasan** Bab ini berisi persiapan data, hasil, dan interpretasi analisis.
- **Bab 5: Kesimpulan dan Saran** Bab ini berisi kesimpulan dari pembuatan skripsi ini dan saran untuk penelitian selanjutnya.