

PREDIKSI TOTAL TRANSAKSI DI APLIKASI GOJEK

Sukma Sekar Devita 1^a, Najma Ulya Zahra 2^b, Gede Nayaka Baswara 3^c, Khansa Mutiara Kheeda 4^d, Freya Gracia 5^e

^a(162112133015) Teknologi Sains Data, Fakultas Teknologi Maju dan Multidisiplin, Universitas Airlangga, Surabaya

^b(162112133059) Teknologi Sains Data, Fakultas Teknologi Maju dan Multidisiplin, Universitas Airlangga, Surabaya

^c(162112133065) Teknologi Sains Data, Fakultas Teknologi Maju dan Multidisiplin, Universitas Airlangga, Surabaya

^d(162112133110) Teknologi Sains Data, Fakultas Teknologi Maju dan Multidisiplin, Universitas Airlangga, Surabaya

^e(162112133111) Teknologi Sains Data, Fakultas Teknologi Maju dan Multidisiplin, Universitas Airlangga, Surabaya

Abstrak

Pada penelitian ini, peneliti menganalisis faktor-faktor apa saja yang mempengaruhi total pengeluaran menggunakan aplikasi Gojek selama 6 bulan. Data dihimpun dari mahasiswa di Indonesia dalam rentang waktu 7 hari dimulai tanggal 25 November 2022 sampai 1 Desember 2022. Analisis dilakukan setelah melalui proses eksplorasi dan visualisasi data. Tujuan dari eksplorasi data adalah untuk memahami data yang dimiliki, seperti pendeteksian outlier pada data, distribusi data yang dimiliki, korelasi antar variabel data, dan lain-lain. Jika terdapat outlier pada data, apakah anomali tersebut terjadi secara MNAR (Missing Not at Random) atau lainnya. Tujuan dari visualisasi data adalah untuk memudahkan dalam memahami data secara visual atau gambar. Namun, pengamatan visual tidak akurat. Maka dari itu diperlukan perhitungan statistik. Setelah dilakukan analisis, diketahui terdapat korelasi yang kuat antara total pengeluaran transaksi Gojek dengan besaran UKT per semester. Pada model regresi diketahui faktor jenis tempat tinggal kos/kontrakan menambah pengeluaran total transaksi Gojek paling besar.

Kata Kunci : visualisasi data, outlier, MNAR, analisis regresi

1. Pendahuluan

1.1 Latar belakang Masalah

Gojek merupakan salah satu jasa yang menyediakan layanan transportasi ojek berbasis aplikasi seperti transportasi motor atau mobil yang bisa dipesan secara *online* oleh masyarakat luas. Gojek berdiri pertama kali pada 13 Oktober 2010 khususnya di DKI Jakarta. Pada saat itu, sistemnya masih sederhana dan jumlah *driver* juga terbatas sehingga cara pemesanan pun masih dilakukan melalui *call center*. Seiring waktu, Gojek mulai berkembang yakni pada tahun 2015 Gojek meluncurkan aplikasi yang bisa diunduh melalui *smartphone* dan mulai merambah jangkauannya ke kota-kota besar lainnya selain Jakarta. Program ini ternyata menuai reaksi positif dari masyarakat luas. Dengan memanfaatkan kecanggihan teknologi *mobile* inilah Gojek berhasil merevolusi industri transportasi ojek. Saat ini, aplikasi Gojek sudah diunduh lebih dari 20 juta orang di berbagai daerah. Selain lebih mudah dalam

memesan Gojek, *customer* juga dimanjakan dengan tarif yang murah meriah. Gojek menjadi solusi alternatif masyarakat dalam bepergian, pesan antar makanan, hingga pengiriman barang.

Adanya aplikasi yang memuat berbagai fitur didalamnya membuat Gojek semakin dikenal masyarakat luas. Selain fitur ojek *online* untuk penumpang (Go Ride dan Go Car), Gojek juga mengembangkan fitur resto *online* (Go Food), fitur pengantaran barang (Go Box dan Go Send), fitur belanja kebutuhan sehari-hari (Go Mart), dan berbagai fitur pembayaran lainnya. Semakin hari, layanan Gojek pun semakin berkembang mengikuti *demand* konsumen yang terus meningkat. Jika konsumen puas dan nyaman, maka mereka akan menggunakan jasa ini berulang kali sampai bisa menjadi pelanggan setia (*loyal customer*) sehingga Gojek menjadi *market leader* dan menjadi perusahaan yang maju serta dipercaya oleh pelanggan di Indonesia. Berdasarkan uraian yang telah disebutkan sebelumnya, maka peneliti ingin melakukan analisis untuk mengetahui faktor-faktor yang mempengaruhi total transaksi mahasiswa di aplikasi Gojek.

1.2 Rumusan Masalah

- 1) Apa saja faktor-faktor yang mempengaruhi total transaksi di aplikasi Gojek pada mahasiswa?
- 2) Bagaimana korelasi variabel prediktor (faktor-faktor yang mempengaruhi) dengan variabel respons (total transaksi)?
- 3) Seberapa besar variabel prediktor dapat mempengaruhi variabel respons?

1.3 Tujuan

- 1) Mengetahui faktor-faktor apa saja yang mempengaruhi total transaksi di aplikasi Gojek pada mahasiswa.
- 2) Mengetahui korelasi dari variabel prediktor (faktor-faktor yang mempengaruhi) dengan variabel respons (total transaksi).
- 3) Mengetahui besaran proporsi variasi dari variabel respons yang dijelaskan oleh variabel prediktor.

2. Landasan Teori

2.1 Analisis Data

Analisis data merupakan proses pemeriksaan, pembersihan, transformasi, dan pemodelan data untuk mencari informasi yang bermanfaat, membuat kesimpulan atas suatu fenomena, dan mendukung dalam pengambilan keputusan. Terdapat berbagai aspek dan pendekatan dalam analisis data yang dapat digunakan sesuai domain data yang berbeda. Dalam aplikasi statistik, analisis data dapat dibagi menjadi tiga yaitu statistika deskriptif, *exploratory data analysis* (EDA), dan *confirmatory data analysis* (CDA). Proses analisis data mencakup *data processing*, *data cleaning*, *data exploratory*, dan pemodelan data.

2.2 Visualisasi Data

Menurut Anton Santoso, manusia merupakan makhluk visual. 90% informasi yang dikirim ke otak berbentuk visual, 50% kapasitas otak bekerja memproses informasi berupa visual. Kemampuan otak manusia dapat memproses gambar 60.000 kali lebih cepat dari kemampuan memproses teks. Oleh karena itu, tujuan utama dari visualisasi data adalah untuk memberikan informasi secara jelas dan efisien melalui grafik informasi, plot, dan grafik statistik. Data-data numerik dapat divisualisasikan dengan titik, garis, atau batang untuk memberi pesan kuantitatif.

2.3 Analisis Regresi

Analisis regresi adalah suatu proses statistik untuk memprediksi hubungan antar variabel, yaitu berupa teknik-teknik pemodelan dan melakukan analisis beberapa variabel untuk melihat hubungan antara satu variabel dependen dan satu atau lebih variabel independen (Amstrong, 2012:689). Analisis

regresi dengan satu variabel independen disebut dengan analisis regresi sederhana. sedangkan analisis regresi dengan lebih dari satu variabel independen disebut analisis regresi berganda.

3. Sumber Data dan Metodologi

3.1 Sumber Data

Data di dapatkan dari kusioner yang telah kami buat yang berisikan total pengeluaran gojek, Uang Kuliah Tunggal (UKT), uang saku perbulannya, membawa kendaraan atau tidak, jarak tempuh tempat tinggal dengan kampus, menggunakan promo atau tidak saat menggunakan aplikasi gojek, cara mendapatkan makan, mempunyai m-banking atau tidak dan banyaknya *m-banking* yang di miliki oleh para mahasiswa. Kami menggunakan seluruh mahasiswa yang ada di Indonesia dalam penelitian yang kami lakukan. Proses pengumpulan data dalam penelitian ini memanfaatkan media *Google Form* yang dilakukan selama 7 hari, yaitu dimulai pada tanggal 25 November 2022 hingga 1 Desember 2022.

Table 1. Variabel Penelitian

Variabel	Keterangan
Variabel Respon	
pengeluaran	Variabel respon dalam penelitian ini adalah jumlah pengeluaran dalam menggunakan aplikasi gojek dalam 6 bulan yang diperoleh oleh responden. Pengeluaran dihitung dari tanggal 25 Mei sampai dengan 25 November.
Variabel Prediktor	
ukt	Selama kita kuliah kita membayar Uang Kuliah Tunggal (UKT) per-semester. Oleh karena itulah, variabel ini berbentuk variabel kontinu yang ingin kita ketahui apakah berpengaruh terhadap pengeluaran dalam menggunakan aplikasi gojek.
uang_saku	Pada variabel kontinu ini kita ingin mengetahui apakah banyaknya uang yang diperoleh tiap bulannya berpengaruh terhadap pengeluaran dalam menggunakan aplikasi gojek.
kendaraan	Variabel ini merupakan variabel kategorik yang memiliki 2 pilihan jawaban, yaitu ya (Mempunyai kendaraan pribadi) atau tidak (tidak memiliki kendaraan pribadi). Dalam penelitian ini, kami ingin mengetahui apakah mempunyai kendaraan pribadi atau tidak berpengaruh terhadap pengeluaran dalam menggunakan aplikasi gojek.
tempat_tinggal	Variabel ini merupakan variabel kategorik yang memiliki 2 pilihan jawaban, yaitu kos/kontrak atau rumah pribadi. Dalam penelitian ini, kami ingin mengetahui apakah tinggal di kontrak/kos atau rumah pribadi berpengaruh terhadap pengeluaran dalam menggunakan aplikasi gojek.
jarak	Pada variabel kontinu ini kita ingin mengetahui apakah jarak mahasiswa pergi ke kampus dari berpengaruh terhadap pengeluaran dalam menggunakan aplikasi gojek.
promo	Variabel ini merupakan variabel kategorik yang memiliki 4 pilihan jawaban, yaitu sangat jarang, jarang, sering, atau sangat sering dalam menggunakan promo yang tersedia di aplikasi gojek. Dalam penelitian ini, kami ingin mengetahui apakah sering atau jarang menggunakan promo akan berpengaruh terhadap pengeluaran dalam menggunakan aplikasi gojek.
makan	Variabel ini merupakan variabel kategorik yang memiliki 3 pilihan jawaban, yaitu dari orang tua, masak sendiri atau beli jadi dalam mendapatkan makanan sehari-hari. Dalam penelitian ini, kami ingin mengetahui apakah hal-hal tersebut dapat berpengaruh terhadap pengeluaran dalam menggunakan aplikasi gojek.
mbank	Variabel ini merupakan variabel kategorik yang memiliki 2 pilihan jawaban, yaitu ya (Mempunyai <i>M-banking</i>) atau tidak (Tidak mempunyai <i>M-banking</i>). Dalam penelitian ini, kami ingin mengetahui apakah mempunyai <i>M-banking</i> atau tidak berpengaruh terhadap pengeluaran dalam menggunakan aplikasi gojek.
banyak_mbank	Dalam variabel ini kita ingin mengetahui banyak <i>M-banking</i> yang dimiliki oleh mahasiswa maka dari itu, variabel ini menjadi variabel kontinu yang ingin kita ketahui apakah juga berpengaruh terhadap pengeluaran dalam menggunakan aplikasi gojek.

3.2 Metodologi

Dalam penelitian ini, metode yang digunakan adalah metode kuantitatif deskriptif. Menurut Robert Donmoyer (dalam Norjanah: 2014), metode penelitian kuantitatif adalah pendekatan terhadap kajian empirik untuk mengumpulkan, menganalisis, dan menampilkan data dalam bentuk numerik. Sedangkan metode deskriptif adalah metode penelitian yang dipakai untuk memberi deskripsi masalah yang sedang

berlangsung. Tujuan dari metode deskriptif adalah memberi gambaran mengenai sesuatu yang terjadi sebagaimana pada saat penelitian dilakukan.

4. Analisis dan Pembahasan

4.1 Mengubah tipe data

```
$ Total Transaksi di GOJEK (6 bulan terakhir)      : num [1:65]
142000 402500 112000 2262683 2941956 ...
$ Jarak dari kos/rumah Anda ke kampus (dalam KM)   : num [1:65]
3 10 1.5 7.5 0.5 2 16 1.3 0.5 2 ...
```

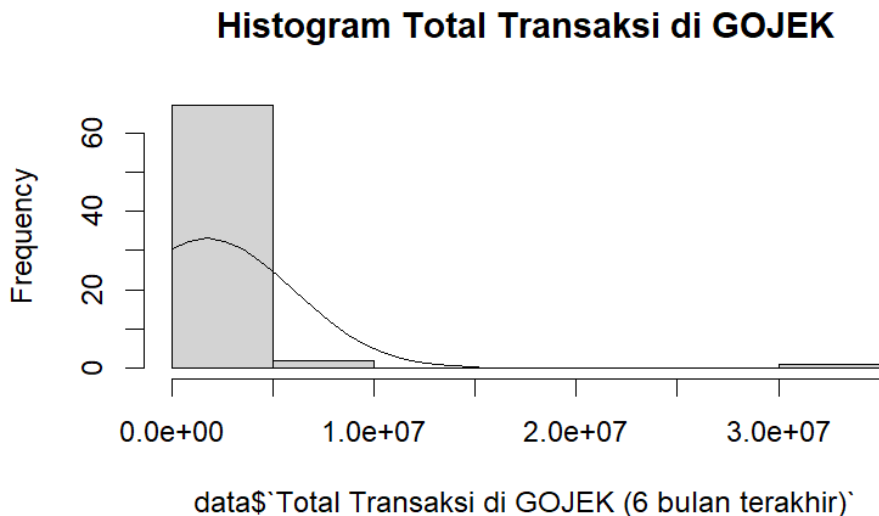
Untuk variabel “Total Transaksi di GOJEK (6 bulan terakhir)” dan “Jarak dari kos/rumah Anda ke kampus (dalam KM)” pada awalnya berupa character maka kami mengubah data tersebut menjadi numeric menggunakan fungsi `as.numeric()`.

4.2 Variabel Dummy

Selanjutnya, peneliti melakukan perubahan pada variabel kualitatif menjadi variabel dummy agar bisa dilihat korelasinya dan bisa dilakukan analisis regresi linear berganda pada variabel-variabel kualitatif ini.

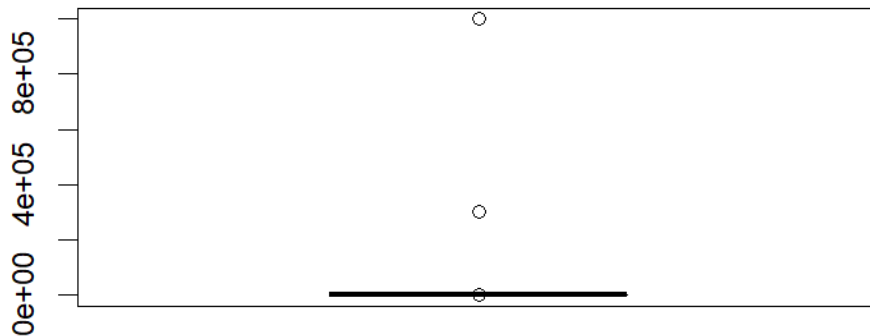
- Untuk variabel dummy bagaimana cara mendapatkan makanan, akan diberi kode 1 untuk “Dari Orang Tua”, 2 untuk “Masak Sendiri” dan 3 untuk “Beli Jadi”.
- Untuk variabel dummy ada atau tidaknya kendaraan, akan diberi kode 1 untuk “Ya” dan 0 untuk “Tidak”.
- Untuk variabel dummy tempat tinggal, akan diberi kode 1 untuk “Rumah” dan 0 untuk “Kos/Kontrak”.
- Untuk variabel dummy ada atau tidaknya m-banking, akan diberi kode 1 untuk “Ya” dan 0 untuk “Tidak”.

4.3 Sebaran Data



Dari plot diatas dapat kita ketahui bahwa data total transaksi GOJEK memiliki skewness positif. Data pada total transaksi GOJEK cenderung pada angka $< 1.000.000$.

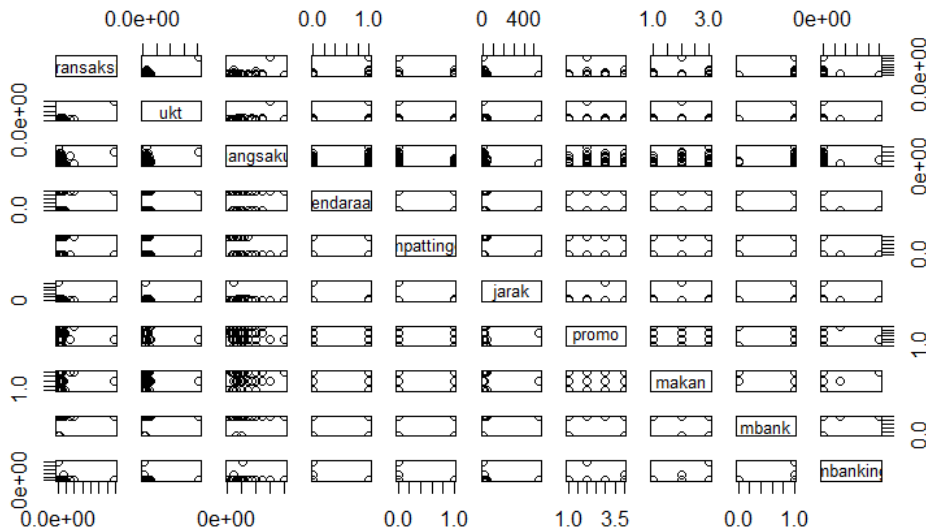
4.4 Outlier



```
> quartiles <- quantile(data$`Jika ada, berapa mobile banking yang anda miliki?`, probs=c(.25, .75), na.rm
= FALSE)
> IQR <- IQR(data$`Jika ada, berapa mobile banking yang anda miliki?`)
> Lower <- quartiles[1] - 1.5*IQR
> Upper <- quartiles[2] + 1.5*IQR
> data <- subset(data, data$`Jika ada, berapa mobile banking yang anda miliki?` > Lower & data$`Jika ada,
berapa mobile banking yang anda miliki?` < Upper)
> dim(data)
[1] 65 18
```

Langkah selanjutnya yaitu mendeteksi outlier, pada data yang kami miliki untuk variabel lain outlier tersebut dibutuhkan maka dari itu kami tidak menghapus outlier pada variabel-variabel lain. Namun pada variabel jumlah mobile banking didapat outlier yang diduga terjadi akibat dari human error, maka dari itu kami memutuskan untuk menghapus variabel tersebut.

4.5 Korelasi



```
> cor(data$`Total Transaksi di GOJEK (6 bulan terakhir)`, data$`Besaran UKT Anda`)
[1] 0.926724
```

Untuk korelasi dari variabel total transaksi dan besaran ukt memiliki korelasi positif yang sangat tinggi yaitu sebesar 0.92 yang berarti bahwa semakin besar besaran ukt maka total transaksi di GOJEK nya akan semakin besar.

```
> cor(data$`Total Transaksi di GOJEK (6 bulan terakhir)` , data$`Besaran uang saku (per bulan)` )
[1] 0.3401427
```

Untuk korelasi dari variabel total transaksi dan besaran uang saku memiliki korelasi positif yang cukup rendah yaitu sebesar 0.34 yang berarti besaran uang saku tidak mempengaruhi total transaksi di GOJEK.

```
> cor(data$`Total Transaksi di GOJEK (6 bulan terakhir)` , data$`Jarak dari kos/rumah Anda ke kampus (dalam KM)` )
[1] -0.07497948
```

Untuk korelasi dari variabel total transaksi dan jarak antara rumah dan kampus memiliki korelasi negatif yang hampir tidak ada yaitu sebesar -0.07 yang berarti bahwa jarak antara rumah dan kampus tidak mempengaruhi total transaksi di GOJEK.

```
> cor(data$`Total Transaksi di GOJEK (6 bulan terakhir)` , data$`Ketika menggunakan aplikasi GOJEK, seberapa sering Anda menggunakan "Promo" yang tersedia?` )
[1] -0.05624124
```

Untuk korelasi dari variabel total transaksi dan seringnya penggunaan promo pada GOJEK memiliki korelasi negatif yang hampir tidak ada yaitu sebesar -0.05 yang berarti bahwa seringnya penggunaan promo pada GOJEK tidak mempengaruhi total transaksi di GOJEK.

```
> cor(data$`Total Transaksi di GOJEK (6 bulan terakhir)` , data$`Jika ada, berapa mobile banking yang anda miliki?` )
[1] 0.0929563
```

Untuk korelasi dari variabel total transaksi dan jumlah m-banking yang dimiliki memiliki korelasi positif yang sangat rendah yaitu sebesar 0.09 yang berarti bahwa jumlah m-banking yang dimiliki tidak mempengaruhi total transaksi di GOJEK.

```
> cor(data$`Total Transaksi di GOJEK (6 bulan terakhir)` , data$dummy_makanan)
[1] -0.03964753
```

Untuk korelasi dari variabel total transaksi dan cara mendapatkan makanan memiliki korelasi negatif yang sangat rendah yaitu sebesar -0.39 yang berarti bahwa cara mendapatkan makanan tidak mempengaruhi total transaksi di GOJEK.

```
> biserial.cor(data$`Total Transaksi di GOJEK (6 bulan terakhir)` , data$dummy_kendaraan)
[1] -0.06058518
```

Untuk korelasi dari total transaksi dan ada atau tidaknya kendaraan memiliki korelasi negatif yang sangat rendah yaitu sebesar -0.06 yang berarti kedua variabel tersebut tidak saling mempengaruhi.

```
> biserial.cor(data$`Total Transaksi di GOJEK (6 bulan terakhir)` , data$dummy_rumah)
[1] 0.03172756
```

Untuk korelasi dari total transaksi dan tempat tinggal memiliki korelasi positif yang sangat rendah yaitu sebesar 0.03 yang berarti kedua variabel tersebut tidak saling mempengaruhi.

```
> biserial.cor(data$`Total Transaksi di GOJEK (6 bulan terakhir)` , data$dummy_mbanking)
[1] -0.05972302
```

Untuk korelasi dari total transaksi dan ada atau tidaknya mbanking memiliki korelasi negatif yang sangat rendah yaitu sebesar -0.05 yang berarti kedua variabel tersebut tidak saling mempengaruhi.

4.6 Regresi

Coefficients:		Estima
te	(Intercept)	-2.291e+
06		
data\$`Besaran UKT Anda`		1.654e-
01		
data\$`Besaran uang saku (per bulan)`		4.226e-
02		
data\$`Jarak dari kos/rumah Anda ke kampus (dalam KM)`		-4.348e+
04		
data\$`Ketika menggunakan aplikasi GOJEK, seberapa sering Anda menggunakan "Promo" yang tersedia?`		1.308e+
05		
data\$`Jika ada, berapa mobile banking yang anda miliki?`		2.205e+
04		
data\$dummy_kendaraan		-2.261e+
05		
data\$dummy_rumah		9.876e+
05		
data\$dummy_makanan		3.800e+
05		
data\$dummy_mbanking		5.675e+
05		

Model regresi merupakan persamaan matematik yang dapat meramalkan nilai suatu variabel dependen dari variabel independen (Walpole, 1982: 340). Pada “Analisis Faktor-faktor Pengeluaran Gojek” didapatkan model regresi seperti di bawah ini:

$$Y = -2.291 + 0.165X_1 + 0.042X_2 - 4.348X_3 + 1.308X_4 + 2.205X_5 - 2.261X_6 + 9.876X_7 + 3.8X_8 + 5.675X_9$$

Jika semua variabelnya konstan, nilai dari Y (total transaksi GOJEK) adalah -2.291. Bertambahnya variabel UKT tiap 1 satuan menambah nilai Y sebesar 0.165. Bertambahnya variabel uang saku per bulan tiap 1 satuan menambah nilai Y sebesar 0.042. Bertambahnya variabel jarak tiap 1 satuan mengurangi nilai Y sebesar 4.348. Kecenderungan dalam menggunakan promo menambah nilai Y sebesar 1.308. Bertambahnya jumlah *mobile banking* tiap 1 satuan menambah nilai Y sebesar 2.205. Keberadaan kendaraan mengurangi nilai Y sebesar -2.261. Tempat tinggal kos/kontrakan menambah nilai Y sebesar 9.876. Cara mendapatkan makanan (dari orang tua, masak sendiri, atau beli jadi) menambah nilai Y sebesar 3.8 tiap 1 satuan. Dan kepemilikan *mobile banking* menambah nilai Y sebesar 5.675.

Residual standard error: 1699000 on 55 degrees of freedom
 Multiple R-squared: 0.8695, Adjusted R-squared: 0.8482
 F-statistic: 40.72 on 9 and 55 DF, p-value: < 2.2e-16

Tabel diatas menunjukkan nilai koefisien determinasi (R-Square) dan koefisien determinasi yang disesuaikan (Adjusted R-Square). Dari hasil perhitungan yang telah peneliti lakukan didapat R-Square sebesar 86.95% yang berarti variabel independen menjelaskan sebanyak 86.95% terhadap variabel y yaitu total pengeluaran pada aplikasi GOJEK dan sisanya dijelaskan oleh variabel lain. Namun, nilai R-Square ini seringkali bermasalah, yaitu setiap bertambahnya variabel maka nilai nya akan meningkat. Maka dari itu untuk mengatasi hal tersebut, dapat menggunakan nilai Adjusted R-Square dengan interpretasi yang sama yaitu variabel y atau total pengeluaran pada GOJEK dijelaskan sebanyak 84.82% oleh variabel independen sisanya dijelaskan oleh variabel lain.

5. Kesimpulan

5.1 Kesimpulan

Berdasarkan hasil dan analisis pembahasan, maka dapat diambil kesimpulan sebagai berikut:

1. Pada tahap *preprocessing data* ditemukan adanya *outlier*
2. Terdapat 3 variabel yang berkorelasi positif (ada pertalian langsung) dengan variabel respons (pengeluaran) yaitu variabel ukt, uang saku, dan banyak m-banking. Sedangkan, yang berkorelasi negatif (hubungan invers) terhadap variabel respons adalah jarak ke kampus, promo, makan.
3. Didapatkan model regresinya

$$Y = -2.291 + 0.165X_1 + 0.042X_2 - 4.348X_3 + 1.308X_4 + 2.205X_5 - 2.261X_6 + 9.876X_7 + 3.8X_8 + 5.675X_9$$

Artinya, jika semua variabelnya konstan, nilai dari Y (total transaksi GOJEK) adalah -2.291. Faktor jenis tempat tinggal kos/kontrakan menjadi faktor penambah pengeluaran total transaksi Gojek yang paling besar.

4. Didapatkan R-square sebesar 86.95% dan Adjusted R-Squarenya sebesar 84.82% yang berarti variabel independen menjelaskan sebanyak 86.95% (R-Square) dan/atau 84.82% (Adjusted R-Square) terhadap variabel y yaitu total pengeluaran pada aplikasi GOJEK dan sisanya dijelaskan oleh variabel lain.

5.1 Saran

Dari hasil penelitian yang diperoleh, terdapat beberapa saran dan masukan yang bisa diberikan oleh peneliti yaitu,

1. Terdapat *outlier* pada variabel jumlah m-banking akibat *human error*, sehingga menyebabkan data tidak akurat karena data mengandung kesalahan nilai atribut atau nilai yang menyimpang dari yang diharapkan. Pada tahap instrumen pengumpulan data peneliti dapat memberikan intruksi atau memberikan format pertanyaan yang tidak menimbulkan kesalahpahaman pada saat *input* data oleh *responder*.

Daftar Pustaka

- [1] Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J., 1993. *Classification And Regression Tree*. New York: Chapman AND Hall.
- [2] Nello, C. & Taylor, J. S., 2000. *An Introduction to Support Vector and Other Kernel-based Learning Methods*. s.l.:Cambridge University Press.
- [3] Few, S., *Eenie, Meenie, Minie, Moe: Selecting the Right Graph for Your Message*, https://www.perceptualedge.com/articles/ie/the_right_graph.pdf, Diakses Pada 9 Desember 2022
- [4] Widyan, I.R., 2017, *Platform Visualisasi Data untuk Pemerintah Amsterdam sebagai Solusi Pembersihan Kota Secara Efektif*, https://repository.its.ac.id/46815/1/2913100014-Undergraduate_Thesis.pdf, Diakses Pada 9 Desember 2022
- [5] Armstrong, Scott J, 2012, *Illusion in Regression Analysis*, International Journal Forecasting, Vol. 28, 689-693