

# Prediksi harga konsultasi berdasarkan pedoman billing Inkindo

Fikri Yudhistira Ajiwijaya  
February 2nd 2023

## Latar belakang

Inkindo, singkatan dari Ikatan Nasional Konsultan Indonesia, adalah sebuah badan asosiasi konsultasi. Setiap tahunnya Inkindo melakukan publikasi pedoman billing yang bermaksud memberikan standar minimal harga jasa konsultasi profesional.

Pada pedoman ini, harga ditetapkan oleh beberapa macam metrik dari konsultan, seperti gelar, sertifikasi, tahun pengalaman, dan indeks daerah, dan metrik dari klien, seperti dalam negeri atau luar negeri. Pedoman ini biasa digunakan oleh calon klien untuk menetapkan budget mereka.

## Permasalahan

Pedoman billing yang dipublikasikan oleh Inkindo pada beberapa terahun terakhir adalah pedoman untuk tahun-1 (tahun min satu). Misalnya, pada tahun 2021, pedoman yang dipublikasikan adalah pedoman billing untuk tahun 2020.

Dibutuhkan metode untuk mendapatkan billing pada tahun-tahun dimana Inkindo tidak menyediakan pedomannya. Metode ini diperlukan untuk memenuhi 2 macam kebutuhan, yaitu menetapkan harga tawaran pada tahun saat itu, dan penyusunan anggaran pada tempo yang akan datang.

## Pekerjaan sebelumnya

Usaha membangun model untuk memprediksi harga jasa konsultasi yang tidak jauh dari pedoman billing Inkindo sudah pernah dilakukan. Laporan akhir dari usaha ini dapat dilihat pada file terlampir bernama “Forecasting Tahun 2022”.

Berdasarkan historical data Inkindo, harga jasa konsultasi tahun ke tahun konsisten naik dalam jangkauan 1 juta rupiah dari 2017 sampai 2020. Pada tahun 2021, kenaikan yang dialami cukup drastis yaitu dalam jangkauan 2 juta rupiah.

Dilakukan analisis untuk menentukan apakah harga pada 2021 tersebut adalah sebuah anomali, apakah pada tahun 2022 billing akan tetap naik, dan seberapa besar kenaikan tersebut.

Analisis tersebut dilakukan pada tahun 2022. Data yang digunakan berasal dari 5 tahun terakhir, yaitu dari pedoman Inkindo tahun 2017 sampai tahun 2021. Harapan hasil yang didapat adalah harga jasa konsultasi untuk tahun 2022 tidak turun.

Metode yang digunakan untuk membangun model yang digunakan pada analisis tersebut adalah *simple linear regression*, yaitu analisis regresi linear dengan 1 regressor/variabel bebas dan 1 response/variabel tak bebas. Regressor yang digunakan adalah tahun, dengan billing sebagai response, sesuai dari tertera pada pedoman Inkindo.

Analisis yang dilakukan terbatas untuk billing pada klien dalam negeri. Cara perhitungan antara konsultan bersertifikasi dan tidak bersertifikasi adalah menggeserkan tahun pengalamannya, sehingga hanya perlu dilakukan analisis pada perhitungan tak bersertifikasi, dan hasilnya dapat disesuaikan. Metrik indeks daerah hanyalah faktor pengali dari billing, yaitu untuk DKI Jakarta dengan indeks 1.00, sehingga dapat diabaikan untuk tujuan analisis ini.

Untuk semua permutasi tahun pengalaman dan gelar yang ada dari pedoman Inkindo, dibuat model *simple linear regression*-nya. Kemudian, dari model yang didapat, dibuat estimasi harga untuk tahun 2022, dengan mengambil titik pada garis model di tahun tersebut.

Hasil dari analisis yang pernah dilakukan ini adalah semua harga jasa konsultasi untuk semua tahun pengalaman dan gelar di tahun 2022 turun dari tahun 2021. Hasil ini dapat diverifikasi langsung pada angka yang tertera pada file laporan akhir yang dilampirkan.

Analisis yang pernah dilakukan ini akan dibahas lebih lanjut dan dikaji ulang pada bagian *model analysis* dan *model interpretation* di bab selanjutnya.

## **Pekerjaan sekarang**

Analisi yang dilakukan terbatas untuk billing pada klien dalam negeri. Metrik indeks daerah diabaikan pula.

## **Data collection**

Salah satu kelemahan dari model sebelumnya adalah kurangnya banyak data. Karena ia hanya menggunakan 1 regressor, yaitu tahun, dan banyak historical data yang dimiliki hanya 5 tahun kebelakang, maka ada resiko terjadinya *overfitting* pada model. Kemudian, karena data yang sedikit, model yang dihasilkan cenderung tidak *mampu*, dijelaskan lebih lanjut pada bagian *model analysis*.

Oleh karena itu, pada analisis ini, dilakukan pencarian data lebih banyak terlebih dahulu. Yaitu dengan mencari historical data Inkindo ke tahun kebelakang yang lebih jauh. Pencarian dilakukan melalui Google, dari berbagai macam sumber dimana terdapat pengarsipan pedoman billing Inkindo.

Hasil dari pengumpulan data ini adalah dokumen billing Inkindo untuk 13 tahun yang berbeda. Seluruh dokumen yang didapat berbentuk file PDF. Beberapa dari PDF ini adalah dokumen digital, dan beberapa adalah koleksi foto dokumen fisik. Tahun-tahun yang didapat diantaranya adalah 2023, 2021, 2019, 2018, 2017, 2016.

2015, 2014, 2013, 2011, 2010, 2008 dan 2007. Dokumen-dokumen ini dapat dilihat pada lampiran folder “Inkindo Data” dengan format PDF.

### **Data entry**

Karena seluruh historical data pedoman billing Inkindo ada dalam bentuk PDF yang tidak bisa dibaca langsung oleh program, maka perlu dilakukan data entry manual untuk menjadikan data-data dalam PDF ke bentuk tekstual

Pada proses data entry ini, struktur data (format tabel) tetap dipertahankan tanpa dirubah sama sekali. Data ini dimasukkan ke dalam format TXT, yang kemudian dikonversikan ke format CSV. Hasil data entry ini dapat dilihat pada lampiran folder “Inkindo Data”, dalam bentuk kedua format tersebut.

### **Data cleaning**

Salah satu kesulitan yang dihadapi ketika mengkoleksi historical data pedoman billing Inkindo adalah adanya perubahan struktur table dan juga interpretasi dari metrik-metrik yang disediakan.

Oleh karena itu, agar dapat dilakukan analisis dengan konteks yang konsisten, seluruh historical data tersebut perlu diselaraskan agar memiliki interpretasi yang sama. Metode yang digunakan dalam usaha penyelarasan data ini adalah, seluruh data yang ada dipaparkan langsung secara bersampingan, dan dilihat apakah perubahan harga dari tahun ke tahunnya wajar, dan disesuaikan dengan melakukan penggeseran.

Dari proses tersebut, diputuskan bahwa data yang sudah digeserkan, sebagai *selaras*. Hasil dari penyelarasan data ini dapat dilihat di dalam lampiran folder “Inkindo Align” pada file yang bernama “Inkindo Align.xlsx”.

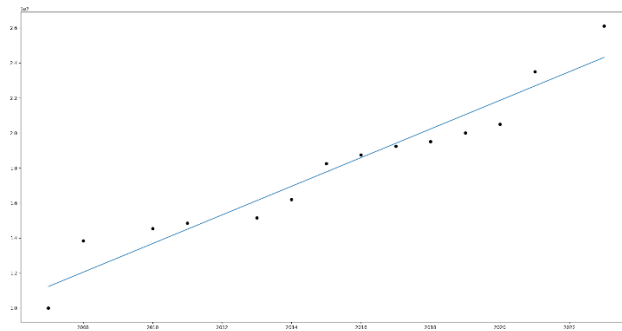
Setelah penyelarasan data ini, diputuskan untuk melakukan interpretasi metrik pada seluruh data sesuai dengan format yang digunakan pada tahun 2019. Pada format ini, dapat diperhatikan bahwa kalkulasi sertifikasi hanya menggeserkan tahun pengalamannya. Maka, untuk kebutuhan analisis ini, diambil tahun pengalaman tanpa sertifikasi, dan hasil yang diperoleh nanti dapat disesuaikan.

Hasil dari penyelarasan metrik ini dapat dilihat di dalam lampiran folder “Inkindo Align” pada file yang bernama “Inkindo Model.xlsx”. Data yang tertera pada file “Inkindo Model.xlsx” ini disebut sebagai *wide format*, dengan *hierarchical column index*. Agar memudahkan pembacaan data oleh program, data tersebut dikonversikan menjadi *long format* tanpa *hierarchical index*.

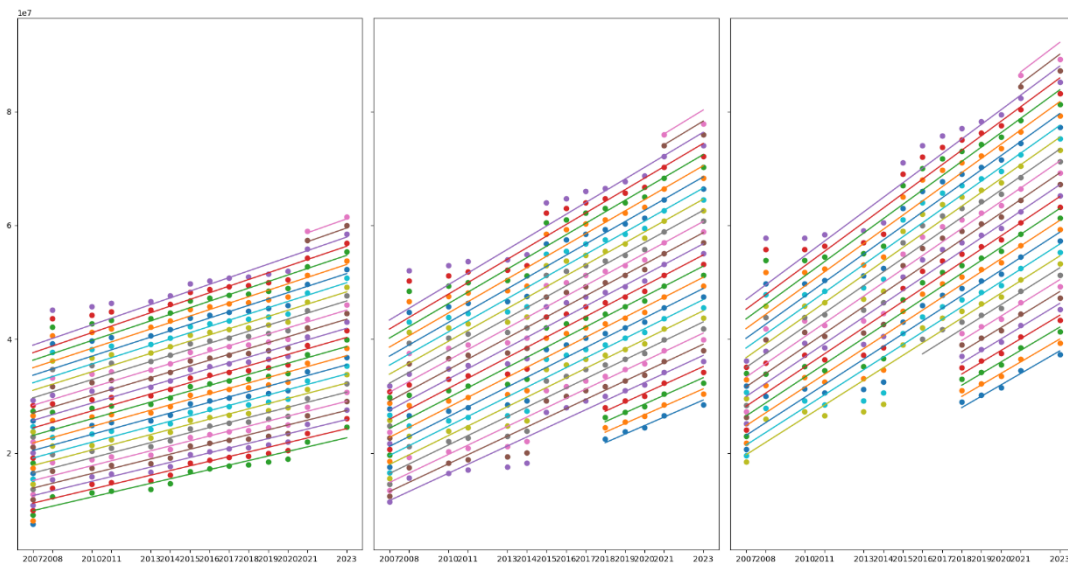
Pada tahap ini, pengkonversian dilakukan menggunakan Python script. File script ini dapat diakses dalam lampiran folder “Inkindo Align” dengan file bernama “long.py”. Bila script ini dijalankan, akan menghasilkan versi data dalam *long format* dengan ekstensi CSV. Hasil file ini dapat dilihat dalam lampiran folder “Inkindo Align” pada file bernama “Inkindo.csv”. File “Inkindo.csv” inilah yang akan dijadikan sebagai sumber data yang dibaca oleh program untuk perlakuan analisis.

## Data exploration

Kenaikan harga jasa konsultasi oleh pedoman Inkindo pada jangkauan yang melebihi 1 juta rupiah pernah terjadi sebelumnya, yaitu pada tahun 2007 ke tahun 2008, dan tahun 2014 ke tahun 2015. Tetapi lonjakan ini selalu terjadi satu kali. Sedangkan pada saat ini, sudah terjadi dua kali berturut-turut, yaitu dari tahun 2020 ke tahun 2021, dan dari tahun 2022 ke tahun 2023. Figur berikut berdasarkan gelar S1 dengan pengalaman 3 tahun.



Dicoba multiple linear regression, figur kiri adalah S1, figur tengah adalah S2, figur kanan adalah S2. Warna menunjukkan tahun pengalaman pada gelar, garis menunjukkan garis model, titik menunjukkan data Inkindo. Perhatikan trend perubahan tahun ke tahun harga jasa konsultasi yang diberikan oleh Inkindo.

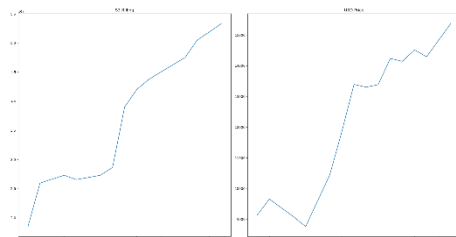


Dapat diamati bahwa perubahan data S1 adalah yang paling reguler, sedangkan perubahan data S3 yang paling sangat bervariasi. Dapat dilihat pada gelar yang sama, tahun pengalaman berbeda, garis model hanya bergeser ke atas. Sedangkan pada gelar yang berbeda dengan tahun yang sama, garis model tetap mulai dari awalan yang sama tetapi dengan kemiringan yang berbeda.

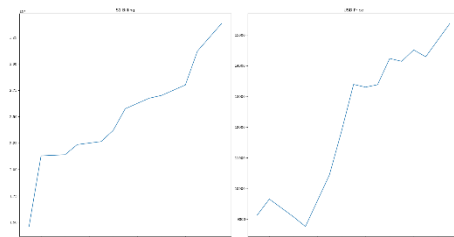
Pada data Inkindo di S3 dapat diperhatikan pula bahwa asumsi harga jasa konsultasi yang akan selalu naik tidak sepenuhnya benar. Pada pergantian tahun 2010 ke tahun 2011, billing untuk gelar S3 mengalami penurunan.

Dicoba dicari data lain yang dapat menjelaskan variasi yang diamati pada data Inkindo untuk gelas S1 dan S3.

Dicoba dibandingkan harga jasa konsultasi dengan harga dolar. Figur kiri untuk jasa S3, figur kanan untuk jasa S1. Subfigur kiri untuk harga jasa, subfigur kanan untuk harga dolar. Dapat dilihat trend harga dolar mirip dengan billing S3, tetapi tidak dengan billing S1.

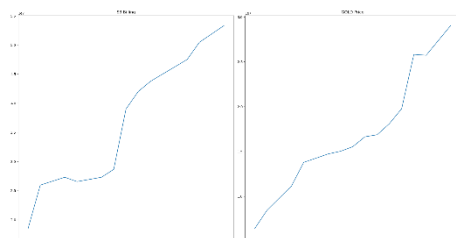


S3 Billing vs USD

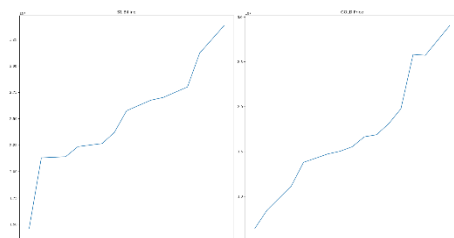


S1 Billing vs USD

Dicoba dibandingkan harga jasa konsultasi dengan harga emas. Figur kiri untuk jasa S3, figur kanan untuk jasa S1. Subfigur kiri untuk harga jasa, subfigur kanan untuk harga emas. Dapat dilihat trend harga emas mirip dengan billing S1, tetapi tidak dengan billing S3.



S3 Billing vs GOLD



S1 Billing vs GOLD

## Model analysis

Signifikan pada statistik berbeda dengan signifikan pada bisnis. Signifikan pada statistik artinya pengamatan yang didapat bukanlah *coincidence* (tidak terjadi secara kebetulan). Signifikan pada bisnis adalah perbedaan nilai tersebut signifikan atau tidak, misalnya perbedaan antara 100 ribu rupiah dengan 1 juta rupiah.

Model pada simple linear regression berbentuk,

$$y = b_0 + b_1x$$

dimana,  $b_0$  adalah titik perpotongan garis model dengan sumbu-y, dan  $b_1$  adalah gradien dari garis model.

Dapat dilakukan tes hipotesa untuk menolak hipotesa null  $b_1 = 0$  secara signifikan statistik. Bila ditolak, artinya  $b_1$  tidak nol, dan garis model bukanlah konstan (gradien model tidak nol).

Dilakukan tes hipotesa ini pada model simple linear regression yang sudah pernah dikerjakan, yaitu dengan data 5 tahun. Pada model gelar S1 dan 3 tahun pengalaman, nilai  $b_0 = -1878820000$  dan  $b_1 = 940000$ . Menghasilkan tes hipotesa untuk  $b_1$ ,  $p = 0.05125$ , dimana ia gagal menolak hipotesa null pada level signifikan 0.05. Sehingga, model yang pernah dikerjakan sebenarnya tidak *mampu*.

Dicoba simple linear regression yang sama, untuk gelar S1 dan 3 tahun pengalaman, tetapi dengan banyak data 13 tahun yang berbeda. Didapat nilai  $b_0 = -1630838934.427$  dan  $b_1 = 818169.3989076116$ . Menghasilkan tes hipotesa untuk  $b_1$ ,  $p = 0.00000002$ , dimana ia berhasil menolak hipotesa null pada level signifikan 0.05. Sehingga, model ini *mampu*.

Dapat disimpulkan, bahwa metodologi yang pernah dilakukan sebenarnya cukup, tetapi terbatas oleh banyaknya data yang dimiliki.

Akan tetapi, kelemahan dari model yang pernah dilakukan adalah pengabaianya pada metrik-metrik yang lain. Setiap permutasi dari gelar dan tahun pengalaman dibuat menjadi model-model yang independen dari satu dan lainnya.

Sehingga, untuk kombinasi dimana data yang ada tidak cukup, karena dari Inkindo tidak tersedia poin data tersebut, menyebabkan “kebolongan” pada data. Maka, model tidak dapat dilakukan, atau memperburuk *mampunya* model tersebut. Contohnya pada kasus S2, dimana beberapa model yang dibuat hanya memiliki data sebanyak 4 tahun, karena Inkindo tidak memberikannya untuk tahun 2017.

Sebenarnya, adanya data hanya sampai 5 tahun terakhir tidak seharusnya menjadi suatu hambatan, karena banyaknya data secara total sudah sangat banyak. Oleh karena itu, diperlukan metode lain dari *simple linear regression* kepada *multiple linear regression*.

*Multiple linear regression* adalah analisis regresi linear dengan lebih dari satu regressor/variabel bebas, dan satu response/variabel tak bebas.

Pada model *multiple linear regression*, seluruh dimensi data dapat digunakan, sehingga meskipun hanya terdapat data 5 tahun terakhir, total banyak poin data yang tersedia untuk membuat model menjadi dalam jangkauan 200 atau lebih. Pada analisis ini, karena ketersediannya historical data yang lebih banyak lagi, maka digunakan semua data yang ada, yaitu dalam jangkauan 900 atau lebih.

Model pada multiple linear regression berbentuk,

$$y = b_0 + b_1x_1 + b_2x_2 + \dots$$

dimana,  $b_0$  adalah titik perpotongan garis model dengan sumbu y, dan  $b_1, b_2, \dots$  adalah kontribusi gradien regressor/variabel bebas  $x_1, x_2, \dots$  kepada garis model.

Terdapat pula regressor yang memiliki sifat *categorical* (kategori), dan perlu diberi perlakuan khusus kedalam model, sebagai *dummy variable*.

Misalnya, untuk metrik gelar, ia adalah variabel bebas bersifat kategori. Terdapat 3 kategori untuk regressor gelar, yaitu S1, S2 dan S3. Maka akan menghasilkan 3 variabel dummy,  $d_0$ ,  $d_1$ , dan  $d_2$ . Nilai dari variabel dummy  $d_0$  adalah 1 bila gelar adalah S1 atau 0 bila gelar bukan S1. Hal yang sama berlaku untuk variabel dummy  $d_1$  dan  $d_2$ .

Model dengan 2 regressor, dimana salah satunya adalah categorical yang memiliki 2 kategori, berbentuk

$$y = d_0b_0 + d_0b_1x + d_1b_2 + d_1b_3x$$

dimana  $b_0$  adalah kontribusi *intercept* oleh kategori 1,  $b_1$  adalah kontribusi gradien oleh kategori 1,  $b_2$  adalah kontribusi *intercept* oleh kategori 2, dan  $b_3$  adalah kontribusi gradien oleh kategori 2.

Pada analisis ini, terdapat 3 metrik, dimana salah satunya adalah categorical yang memiliki 3 kategori. Maka, model yang dibuat menjadi,

$$y = d_0(b_0 + b_1x_1 + b_2x_2 + b_3x_1x_2) + d_1(b_4 + b_5x_1 + b_6x_2 + b_7x_1x_2) + d_2(b_8 + b_9x_1 + b_{10}x_2 + b_{11}x_1x_2)$$

dimana  $x_1$  adalah tahun,  $x_2$  adalah tahun pengalaman,  $d_0$  untuk S1,  $d_1$  untuk S2,  $d_3$  untuk S3, dan  $x_1x_2$  adalah interaksi antara tahun dengan tahun pengalaman.

Setelah menggunakan data yang ada dan membangun model menggunakan multiple linear regression sesuai dengan persamaan di atas, didapatkan nilai

$$\begin{aligned} b_0 &= -1494669807.76, & b_1 &= 747681.765, \\ b_2 &= -33887430.4, & b_3 &= 17542.98 \\ b_4 &= -2960475248.778, & b_5 &= 1476937.44, \\ b_6 &= -45292891.5, & b_7 &= 23357.496, \\ b_8 &= -3948063217.9, & b_9 &= 1969320.05, \\ b_{10} &= -45939723.28, & b_{11} &= 23738.87 \end{aligned}$$

dan tes signifikan statistik statistik

$$\begin{aligned} p_1 &= 4 \times 10^{-24}, p_2 = 2 \times 10^{-4}, p_3 = 1 \times 10^{-4}, \\ p_5 &= 1 \times 10^{-68}, p_6 = 3 \times 10^{-6}, p_7 = 1 \times 10^{-6}, \\ p_9 &= 1 \times 10^{-71}, p_{10} = 1 \times 10^{-4}, p_{11} = 5 \times 10^{-5} \end{aligned}$$

Pada umumnya, waktu bukanlah regressor yang baik. Tetapi pada kasus ini, model yang dihasilkan adalah *mampu*, sehingga ia dapat diterima secara analisis dan untuk keperluan prediksi. Dan mengejutkan bahwa seluruh *coefficient* melewati tes signifikan statistik, dan nilai dari *coefficient* tidak *trivial* pula secara bisnis.

## Model interpretation

Pada laporan yang pernah dilakukan sebelumnya, diberikan prediksi harga jasa konsultasi dengan mengambil titik pada garis model, dan didapati billing di tahun 2022 turun dari billing di tahun 2021. Sangat disayangkan bahwa sebenarnya laporan tersebut masih kurang cukup lengkap, dan interpretasi dari angka prediksi yang diberikan kurang tepat.

Setelah mendapatkan garis model, maka dapat dibuat prediksi harga jasa konsultasi. Akan tetapi, prediksi tersebut tidaklah berbentuk angka, melainkan berbentuk *interval* atau jangkauan.

*Level* dari jangkauan tersebut dapat disesuaikan agar bermakna dan berguna sesuai kebutuhan. Makna dari *level* adalah, seberapa cukup percaya proporsi harga jasa konsultasi akan jatuh pada jangkauan tersebut.

Semakin tinggi *level* ini, semakin lebar jangkauannya, untuk memadai kecukupan percaya bahwa harga akan berada dalam jangkauan. Semakin rendah *level* tersebut, semakin sempit jangkauan.

Perlu dipahami, misalnya pada level cukup percaya untuk proporsi 95% akan berada dalam jangkauan, tetapi jangkauan yang diberikan adalah 5 kali lebih lebar dari pada di level 90%, maka lebih berguna jangkauan pada level 90%.

Selain interval prediksi (*prediction interval*), terdapat juga interval rata-rata (*mean confidence interval*). Maksud dari interval rata-rata ini adalah, ketika melakukan kalkulasi rata-rata, ia akan bervariasi dari sampel ke sampel. Maka, dengan persentase level cukup untuk percaya diberikan interval dimana rata-rata populasi akan jatuh.

Sehingga, bila laporan yang pernah dilakukan dilengkapi, akan diberikan 5 macam angka, yaitu nilai titik di garis model, batas bawah interval rata-rata, batas atas interval rata-rata, batas bawah interval prediksi, dan batas atas interval prediksi.

Pada inteface laporan analisis ini, disediakan grafik beserta tabel yang menunjukkan lima informasi tersebut, dan titik-titik dimana harga jasa konsultasi Inkindo berada sebagai referensi.

## Analysis final report

Salah satu kelebihan dari multiple linear regression adalah prediksi/interpolasi yang dapat ia lakukan tidaklah terbatas pada 1 dimensi. Dapat dilakukan interpolasi kepada dimensi-dimensi lainnya.

Misalnya, keterbatasan dari model sebelumnya adalah pada kombinasi tahun pengalaman dan gelar yang tidak memiliki data, maka tidak dapat dibuat garis model.

Dengan multiple linear regression, hal tersebut tidak lagi menjadi suatu masalah. Bahkan, karena sifat ini pula mengapa multiple linear regression cenderung lebih *robust* terhadap “kebolongan” data.



Contohnya, untuk seorang konsultan yang baru saja lulus dan tidak memiliki tahun pengalaman, Inkindo tidak memberikan pedoman harga jasa konsultasi. Tetapi dengan model yang telah dibuat, dapat sajikan spekulatif garis model beserta prediksi-prediksinya untuk gelar S1 dengan 0 tahun pengalaman (nol, menunjukkan barunya kelulusan), sampai gelar S2 dengan 28 tahun pengalaman.

Oleh karena itu, laporan akhir pada analisis ini disajikan dalam bentuk interaktif, agar dapat memadai berbagai macam kebutuhan *input*. Ia disajikan dalam bentuk program web menggunakan Python.

Dibutuhkan minimal Python versi 3.7, dan menginstall packages yang digunakan oleh program dengan “pip install -r requirements.txt”. File program terlampir bernama “main.py”, dan dapat dijalankan dengan membuka terminal di folder yang sama dan menjalankan “python main.py”.

Pada laporan interaktif ini, disajikan grafik ilustrasi dari tabel yang menterakan angka. Bila terdapat lebih dari 1 tabel, disajikan juga grafik ringkasan untuk membandingkan garis model dari input metrik-metrik yang berbeda.

### **Suggestions**

Karena keperluan bisnis dimana harga konsultasi tidak seharusnya turun pada pergantian tahun, sebagai alternatif dari menggunakan harga pedoman Inkindo terakhir sebagai batas bawah, adalah menggunakan jangkauan dengan batas atas rata-rata dan batas atas prediksi.

Salah satu filosofi dari metode regresi adalah model yang didapat tidak digunakan jauh dari jangkauan yang wajar. Oleh karena itu, disarankan untuk melakukan prediksi harga, dibatasi maksimal untuk 2 tahun ke depan dari tahun terakhir yang diberikan oleh Inkindo.

Sebaiknya bila Inkindo telah mempublikasikan pedoman baru, data tersebut ditambahkan sebagai sumber untuk menyesuaikan kembali model yang ada.

### **Pekerjaan selanjutnya**

Beberapa catatan untuk penerus analisis ini.

Dapat dicoba dilakukan analisis tambahan terhadap indeks daerah. Apakah ia mengalami variasi dari tahun ke tahun? Dan apakah ia dapat digunakan sebagai regressor, atau diperlakukan sebagai response?

Dapat dicoba regressor selain tahun, melainkan hal-hal di luar seperti harga emas, dolar, minyak, gas, dll. Pada analisis ini, pendekatan tersebut hanya terbatas sebagai bagian dari explorasi, tetapi tidak menjadi bagian dari laporan akhir.

Beberapa saran metode model lain yang dapat dieksplorasi untuk melakukan analisis lanjutan termasuk moving average, multivariate linear regression, autoregressive model, non-linear regression, dan machine learning.