

### Quiz 3: Modeling

#### Format Pelaporan

Jawaban harus dikirimkan dalam format pdf. Anda memiliki batas panjang laporan yaitu tujuh halaman. Ukuran huruf tidak boleh lebih kecil dari 10, margin disesuaikan, grafik dan tabel harus disematkan dalam laporan dan disesuaikan skalanya. Lampirkan kode program atau file pendukung lainnya.

#### Latar Belakang

Projek berbagi sepeda adalah varian dari persewaan sepeda tradisional, di mana proses persewaan dan pengembalian sangat otomatis dan fleksibel, biasanya, sepeda dapat disewa di satu lokasi dan dikembalikan di lokasi lain tanpa harus berurusan dengan manusia. Saat ini ada beberapa ratus program berbagi sepeda di banyak kota yang berbeda, dan potensinya sangat menarik.

Anda sebagai seorang analis data diminta oleh salah satu perusahaan persewaan sepeda di negara X untuk memprediksi tingkat persewaan sepeda harian dari variabel lingkungan dan musim. Karena anda belum mengetahui ada penelitian serupa, anda akan mencoba memberi mereka prediksi semacam itu menggunakan model regresi linier berganda. Anda bisa memanfaatkan kumpulan data pada file 'TestCases1.csv', yang diperoleh dari log penggunaan dua tahun sistem berbagi sepeda di negara tersebut.

Variabelnya adalah sebagai berikut:

Variabel	Deskripsi
tanggal	Tanggal lengkap dalam format tahun-bulan-hari
musim	Musim dalam setahun (1 tahun 4 musim)
tahun	Tahun 2020 dan 2021
bulan	Bulan Januari s.d Desember
holiday	Status hari, hari biasa atau hari libur nasional
weekday	Nama hari dalam sepekan
weather	Jenis cuaca
temp	Suhu ternormalisasi (degC)
hum	Humiditas ternormalisasi
windspeed	Kecepatan angin ternormalisasi
registered	Jumlah sepeda yang disewa oleh pengguna terdaftar
count	Jumlah total sepeda yang disewa

Catatan: Anda mungkin akan menemukan hasil yang lebih mudah untuk ditafsirkan jika mengubah beberapa variabel. Perhatikan, menggunakan semua variabel prediktor dapat mengakibatkan masalah kolinearitas.

#### Poin Pertanyaan

Poin-poin berikut ini sangat menarik bagi klien sehingga anda dapat membahasnya dalam laporan.

1. Apakah lebih banyak pengguna terdaftar yang menyewa sepeda pada hari tertentu dapat menyebabkan prediksi total penyewaan sepeda yang lebih tinggi?
2. Apa hubungan antara suhu dan jumlah sepeda yang disewa?
3. Apakah hubungan antara suhu dan jumlah sepeda yang disewa sama dalam dua tahun?
4. Apa hubungan antara kelembaban dan jumlah sepeda yang disewa?

5. Apakah hubungan antara kelembaban dan jumlah penyewaan sepeda akan berbeda dalam kondisi cuaca yang berbeda?
6. Apakah jenis musim mempengaruhi jumlah sepeda yang disewa?
7. Faktor apa yang paling berpengaruh terhadap jumlah sepeda yang disewa?
8. Kapan waktu yang tepat untuk perusahaan melakukan peningkatan layanan?
9. Apakah penyewaan sepeda tahun 2020 sama dengan tahun 2021?
10. Bagaimana tren prediksi jumlah sepeda yang disewa pada 2 tahun kedepan?

## Garis Besar Pelaporan

Susunlah laporan hasil analisis (.pdf) yang mencakup poin-poin berikut:

1. **Pendahuluan:** Tulislah empat sampai lima kalimat yang memperkenalkan masalah penelitian, asumsi yang digunakan, dan menjelaskan hipotesis penelitian.
2. **Explanatory Data Analysis:**
  - Periksa variabel (prediktor dan respon) secara *univariate*, *bivariate*, dan *multivariate*.
  - Menyediakan tampilan grafis atau ringkasan numerik untuk semua variabel. Anda bisa gunakan *software* bantu tableau atau yang lain.
  - Anda membutuhkan EDA untuk semua pasangan variabel numerik, dan setidaknya untuk variabel kategori dan variabel respon. Kemudian beri penjelasan atas hasil yang diperoleh dan jawablah point pertanyaan yang diberikan.
3. **Pemodelan awal:** Mulailah dengan membangun *multivariate model* untuk memprediksi variabel prediktor, kemudian jawablah pertanyaan spesifik klien saat membangun model.
4. **Diagnostik/pemilihan model:**
  - Apakah asumsi dasar terpenuhi untuk model yang digunakan? Mengapa?
  - Transformasi apa yang digunakan? Mengapa?
  - Apakah ada *outlier* dalam sampel yang sangat mempengaruhi model? Identifikasi kandidat *outlier* dan putuskan bagaimana perlakuan yang diberikan. Berikan penjelasan detail.
  - Apakah ada variabel yang dikecualikan? Mengapa?
5. **Inferensi hasil:** Buat tabel yang merangkum model akhir anda (koefisien, kesalahan standar, interval kepercayaan, *p-value*). Berikan interpretasi dari semua koefisien dalam konteks masalah. Pastikan analisis yang anda lakukan sudah menjawab pertanyaan spesifik klien di atas.
6. **Lampiran:** Lampirkan file pendukung seperti kode program dan yang semisal.