

1. Jelaskan tentang skill - skill yang diperlukan oleh data scientist untuk menyelesaikan pekerjaannya!

Secara garis besar, meliputi :

Computer Science/IT, Domains/Business Knowledge, Math and Statistics.

Secara singkat, berikut merupakan skills atau Data Science Lyfe Cycle :

Business Understanding :

Tahap awal yang perlu dimiliki oleh Data Scientist untuk memahami masalah bisnis yang perlu diselesaikan dengan problem solving yang tepat, didukung dengan pemahaman tujuan bisnis, sasaran, dan konteks informasi implisit dari hasil analisis data.

Data Collection :

Pengumpulan data yang diperlukan untuk memecahkan masalah. Tahap ini diperlukannya untuk mengerti, mengidentifikasi sumber data, memeriksa kualitas, validitas, dan memilih jenis data.

Data Mining/KDD :

Penggalian informasi atau penemuan pengetahuan dari data dalam basis data yang besar. Tahap ini melibatkan serangkaian tahap mulai dari pemahaman bisnis, pemilihan dan pengumpulan data, pembersihan data, transformasi data, pemodelan, evaluasi dan visualisasi data.

2. Berikan contoh-contoh penerapan data science

Finance :

- **Prediksi Risiko Kredit:** Penggunaan data mining dan teknik Machine Learning untuk menganalisis data kreditur dapat membantu memprediksi risiko kredit dan memberikan penilaian risiko yang lebih akurat.
- **Analisis Saham :** Data science juga dapat digunakan untuk melakukan analisis portofolio, memprediksi pergerakan pasar saham, dan memperkirakan return investasi.
- **Fraud Detection:** Data science juga dapat membantu dalam mendeteksi aktivitas penipuan keuangan seperti fraud dalam kartu kredit, skimming ATM, dan lain-lain.

Marketing :

- **Analisis Sentimen** adalah metode Data Science yang dapat digunakan untuk memahami opini dan pandangan konsumen terhadap merek atau produk tertentu. Dapat dilakukan untuk menganalisis media sosial, review produk, atau survei konsumen untuk memperoleh wawasan yang lebih baik mengenai persepsi konsumen terhadap merek atau produk.
- **Personalisasi pemasaran** adalah metode Data Science yang dapat digunakan untuk menyediakan pengalaman pemasaran yang disesuaikan dengan preferensi dan perilaku pelanggan. Dapat dilakukan untuk mengumpulkan data perilaku konsumen dan preferensi, dan menggunakan algoritma Machine Learning untuk menyesuaikan pesan pemasaran dan menawarkan produk yang lebih relevan kepada pelanggan.
- **Pemilihan lokasi toko** adalah metode Data Science yang dapat digunakan untuk menentukan lokasi toko yang optimal untuk merek atau produk tertentu. Dapat dilakukan untuk menganalisis data demografis dan geografis, data pesaing, dan data penjualan untuk menentukan lokasi toko yang tepat untuk menjangkau pelanggan potensial.

2. Berikan contoh-contoh penerapan data science

Operation :

- Manajemen inventaris : teknik analisis data untuk membantu perusahaan mengelola inventaris dengan lebih efisien dan menghindari kekurangan atau kelebihan stok yang berlebihan.
- Internet of Things (IoT) untuk memantau mesin dan peralatan produksi secara real-time, atau penggunaan algoritma Machine Learning untuk mengoptimalkan jadwal produksi dan pengiriman barang.
- Prediksi supply chain : data scientist menggunakan algoritma Machine Learning untuk memprediksi permintaan produk atau bahan baku di masa depan dan mengelola rantai pasokan dengan lebih efektif.

Sales :

- Sales Opportunity Analysis : Data Science dapat digunakan untuk mengidentifikasi peluang penjualan baru dan memprediksi potensi penjualan dengan analisis data seperti data pelanggan, data historis penjualan, dan data pasar. Hal ini memungkinkan perusahaan untuk mengembangkan strategi penjualan yang lebih efektif dan memaksimalkan potensi penjualan.
- Sales Performance Prediction : Data Science dapat digunakan untuk memprediksi kinerja penjualan di masa depan dengan menganalisis data historis penjualan, data pasar, dan faktor-faktor lainnya. Hal ini memungkinkan perusahaan untuk mengambil tindakan yang tepat untuk meningkatkan kinerja penjualan di masa depan.
- Sales Optimization : Data Science dapat digunakan untuk mengoptimalkan proses penjualan dengan mengidentifikasi dan mengurangi waktu tunggu dan waktu berbicara, serta memperbaiki pengalaman pelanggan selama proses penjualan. Hal ini dapat meningkatkan efisiensi penjualan dan meningkatkan kepuasan pelanggan.

3. Berikan analisis CRISP-DM sederhana yang bisa kamu dapatkan dari data harga rumah di Boston!

Business Understanding :

Tujuan analisis :

1. Mengembangkan model prediksi harga rumah yang akurat didukung dengan akurasi RSquared yang tinggi.
2. Menentukan faktor-faktor utama yang mempengaruhi harga rumah di Boston.
3. Visualisasi Data untuk memperoleh informasi implisit mengenai faktor-faktor harga rumah di Boston.

Data Understanding :

Memahami karakteristik, tipe data, dan variasi nilai pada setiap atribut dataset.

Secara garis besar, terdapat 18 atribut dari dataset. Adapun variabel dependen sebagai price sedangkan variabel independen selain atribut price.

Data Preparation :

Menyiapkan data untuk digunakan dalam proses penggalian data.

3. Berikan analisis CRISP-DM sederhana yang bisa kamu dapatkan dari data harga rumah di Boston!

Data Preprocessing :

- Data Cleaning :

Pada tahap ini, saya menghapus atribut country, date, dan street. Alasan menghapus atribut tersebut, pada atribut country hanya berisi 1 nilai yaitu USA. Lalu, untuk date hanya sebagai kapan rumah dijual, dalam hal ini menurut saya tidak berguna sebagai atribut faktor-faktor utama harga rumah. Terakhir, tidak menggunakan atribut street karena nilainya terlalu unique untuk ditrain dalam model.

- Data Transformation :

Pada tahap ini, saya melakukan Label Encoding pada atribut bertipe Object dan melakukan perubahan nilai atribut dari yr_renovated menjadi binary.

- Data Rescaling :

Saya melakukan scaling data dengan teknik logaritma terhadap atribut price. Hal ini untuk menyeimbangkan nilainya. Selain itu, saya juga menerapkan ke semua atribut dengan teknik PowerTransformer dengan metode Yeo Johnson.

EDA :

- Korelasi Matrix
- Scatter Plot
- Bar Plot

Data Postprocessing :

- Modeling

Saya membangun model dengan teknik multiple linear regression model. Saya menggunakan atribut price sebagai variabel Y. Dalam hal ini juga saya menggunakan data dummy yang dari output hasil encoding data transformasi.

- Evaluasi Model

Model yang didapatkan memiliki nilai RSquared sebesar 0.97. Selain itu, terdapat nilai measurement Error yaitu MSE sebesar 0.023 dan MAE sebesar 0.037.

3. Jelaskan feature - feature data yang ada, dan berikan hipotesis feature apa yang berpengaruh terhadap harga rumah!

- date: tanggal rumah dijual
- price: harga rumah yang dijual
- bedrooms: jumlah kamar tidur
- bathrooms: jumlah kamar mandi
- sqft_living: ukuran persegi ruang tamu
- sqft_lot: ukuran persegi dari lot
- floors: total lantai di rumah
- waterfront: apakah rumah berada di waterfront (bool)
- view: memiliki spesial view
- condition: kondisi rumah
- sqft_above: ukuran persegi rumah selain ruang bawah tanah
- sqft_basement: luas ruang bawah tanah
- yr_built: tahun dibangun
- yr_renovated: tahun renovasi rumah
- street : alamat jalan rumah
- city : letak kota rumah
- country : negara letak rumah

- $X = \text{bedrooms, bathrooms, sqft_living, sqft_lot, floors, waterfront, view, condition, sqft_above, sqft_basement, yr_built, yr_renovated, city, statezip}$
- $Y = \text{price}$

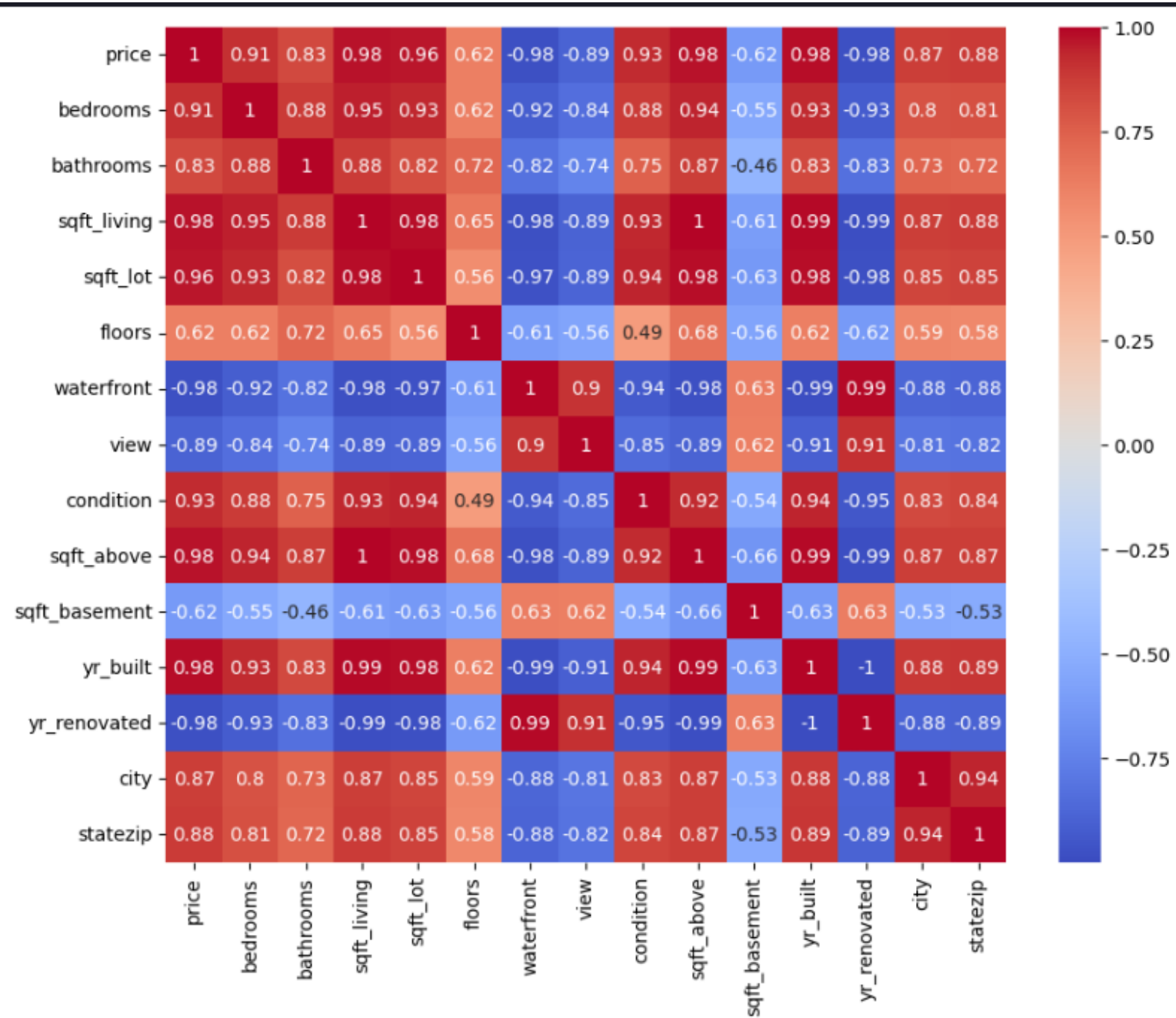
Hipotesis :

1. Apakah lokasi menjadi faktor yang paling penting dalam menentukan harga rumah di Boston?
2. Bagaimana ukuran rumah mempengaruhi harga rumah di Boston?
3. Bagaimana pengaruh jumlah kamar tidur, kamar mandi, ruang tamu, memiliki fitur tambahan seperti basement, waterfront, kondisi/usia dan pemandangan?

3. Jelaskan feature - feature data yang ada, dan berikan hipotesis feature apa yang berpengaruh terhadap harga rumah!

Hasil analisis :

- Ukuran rumah juga menjadi faktor penting dalam menentukan harga rumah di Boston. Rumah yang lebih besar dengan lebih banyak kamar tidur, kamar mandi, atau ruang tamu akan memiliki harga yang lebih tinggi daripada rumah dengan ukuran yang lebih kecil.
- Usia rumah juga berpengaruh pada harga rumah di Boston. Rumah yang lebih tua mungkin memiliki harga yang lebih rendah karena adanya keausan atau kebutuhan untuk renovasi.
- Lokasi merupakan faktor yang paling penting dalam menentukan harga rumah di Boston. Rumah yang berada di lokasi yang strategis, seperti dekat dengan pusat kota atau akses transportasi yang mudah, akan memiliki harga yang lebih tinggi daripada rumah yang berada di lokasi yang terpencil.



	Slope
bedrooms	-0.038722
bathrooms	0.028511
sqft_living	-0.007262
sqft_lot	-0.026677
floors	0.007253
waterfront	-0.181513
view	0.024360
condition	0.004136
sqft_above	0.191413
sqft_basement	0.011293
yr_built	-1.652233
yr_renovated	-2.311981
city	0.006335
statezip	0.015223

Intercept: -0.0019969782186210045

Mean Squared Error: 0.023557457995490312
Mean Absolute Error: 0.037807679456466584
R-squared: 0.9763976217655786

3. Apa outcome yang dapat diharapkan dari data tersebut?

Model Multiple Linear Regression :

$$Y' = -0.001996 - 0.038(\text{bedrooms}) + 0.028(\text{bathrooms}) - 0.007262(\text{sqft_living}) - 0.0266(\text{sqft_lot}) + 0.0072(\text{floors}) - 0.181(\text{waterfront}) + 0.024(\text{view}) + 0.0041(\text{condition}) + 0.191(\text{sqft_above}) + 0.0112(\text{sqft_basement}) - 1.652(\text{yr_built}) - 2.311(\text{yr_renovated}) + 0.0063(\text{city}) + 0.0152(\text{statezip})$$

Melalui Model tersebut, maka outcome yang diharapkan :

- Menentukan faktor-faktor apa saja yang mempengaruhi harga rumah di Boston. Hal ini dapat membantu para pengembang properti atau investor untuk membuat keputusan yang lebih baik dalam memilih lokasi dan jenis properti yang akan dibangun atau dibeli.
- Memprediksi harga rumah di masa depan. Dengan menggunakan data historis dan teknik analisis yang tepat, kita dapat membuat model prediksi harga rumah di Boston yang akurat. Hal ini dapat membantu pembeli dan penjual untuk menentukan harga yang tepat pada waktu yang tepat.
- Memahami kondisi pasar properti di Boston. Analisis harga rumah juga dapat memberikan wawasan tentang bagaimana pasar properti sedang berjalan di Boston. Informasi ini dapat membantu para pembeli dan penjual untuk membuat keputusan yang tepat dalam membeli atau menjual properti mereka.

Link :

https://github.com/rfdharma/Mini_Camp_DataScience_StartupCampus/blob/master/model.ipynb