



MySkill | *#RintisKarirImpian*

Portfolio - Short Class

Data Science Introduction

Owner: Ika Nurfitriani

Build your skill and portfolio via myskill.id/bootcamp

Course Summary

Poin Belajar	Rangkuman
Pengenalan Data Science	<ul style="list-style-type: none">• Data science mengkombinasikan matematika, statistik, pemrograman khusus, analisis lanjutan, kecerdasan buatan, dan machine learning untuk mendapatkan insight dari data yang dimiliki.• Proses data science: pengumpulan data, pembersihan data, eksplorasi menggunakan metode statistik, pemodelan untuk prediksi, dan interpretasi hasil.• Alat yang digunakan: bahasa pemrograman seperti Python dan R, platform visualisasi seperti Tableau, Power BI, dan Looker Studio. Serta basis data seperti MySQL, PostgreSQL, dan MongoDB.• Data scientist menggunakan statistika dan machine learning untuk memprediksi dan menjawab <i>key business questions</i>, data engineer yang membangun dan mengoptimasi sistem data analyst untuk bekerja, dan data analyst yang membagikan "value" dengan cara mengambil data dan mengkomunikasikan hasil kepada stakeholder.• Contoh pengaplikasian pada data science yakni sistem rekomendasi, penilaian risiko kredit, segmentasi pasar, deteksi penipuan, serta pengenalan objek, teks dan suara.
Mengumpulkan dan Menyimpan Data	<p>Asal Data:</p> <ul style="list-style-type: none">• Company Data: Dikumpulkan oleh perusahaan untuk membantu pembuatan keputusan, seperti data logistik, survei, dan data pelanggan.• Open Data: Data gratis, open source, serta dapat digunakan, dibagikan dan diubah oleh siapapun, seperti API dan <i>public records</i>.

Course Summary



Poin Belajar	Rangkuman
Mengumpulkan dan Menyimpan Data	<p>Tipe Data:</p> <ul style="list-style-type: none">• Data Kualitatif: Tidak dapat diukur tapi dapat diobservasi. Contoh: Berwarna biru.• Data Kuantitatif: Angka yang dapat diukur. Contoh: Tinggi badan 155 cm.• Lainnya: Image data, Spatial data, Text Data, dan Network data <p>Hal yang Perlu Diperhatikan saat Menyimpan Data:</p> <ol style="list-style-type: none">1. Lokasi: Parallel storage atau cloud.2. Tipe Data: Tidak terstruktur, bentuk tabel, bentuk dokumen, atau relational database.3. Proses Retrieval: Bahasa query SQL atau nonSQL.
Menyiapkan Data	<p>Data mentah biasanya berantakan. Dengan merapikan data akan membantu mengurangi error, hasil yang salah, dan bias dalam algoritma. Contoh-contoh merapikan data:</p> <ul style="list-style-type: none">• Menghilangkan Duplikasi• Menggunakan ID Unik• Membuat Data Homogen• Penyesuaian Tipe Data• Mengatasi Missing Value. <p>Missing value bisa muncul karena error, data entry dan dilihat sebagai entitas yang valid. Mengatasi missing value bisa dengan disubstitusikan, dikeluarkan, atau dibiarkan.</p>



Mini Task

Simaklah tabel di slide di samping lalu siapkanlah data tersebut menggunakan materi yang sudah dijelaskan, yakni mencakup menghilangkan duplikasi data, menggunakan ID unik, membuat data menjadi homogen, menyesuaikan tipe data, dan mengatasi *missing value*.

Nama	Umur	Tinggi Badan	Negara
Teresa	"23"	165	Belgia
Alex	"20"		USA
Patricia	"18"	181	Prancis
Michelle	"22"	172	SG
Anggun	"21"	156	Indonesia
Michelle	"22"	172	SG
John	"25"	169	Kanada
Timothy	"29"	180	USA
Timothy	"29"	180	USA
Airlangga	"20"		ID

Hasil Persiapan Data

Menghilangkan Duplikasi Data

Nama	Umur	Tinggi Badan	Negara
Teresa	"23"	165	Belgia
Alex	"20"		USA
Patricia	"18"	181	Prancis
Michelle	"22"	172	SG
Anggun	"21"	156	Indonesia
John	"25"	169	Kanada
Timothy	"29"	180	USA
Airlangga	"20"		ID

Hasil Persiapan Data

Menggunakan ID Unik

No ID	Nama	Umur	Tinggi Badan	Negara
0	Teresa	"23"	165	Belgia
1	Alex	"20"		USA
2	Patricia	"18"	181	Prancis
3	Michelle	"22"	172	SG
4	Anggun	"21"	156	Indonesia
5	John	"25"	169	Kanada
6	Timothy	"29"	180	USA
7	Airlangga	"20"		ID

Hasil Persiapan Data

Membuat Data Homogen

No ID	Nama	Umur	Tinggi Badan	Negara
0	Teresa	"23"	165	Belgia
1	Alex	"20"		Amerika Serikat
2	Patricia	"18"	181	Prancis
3	Michelle	"22"	172	Singapura
4	Anggun	"21"	156	Indonesia
5	John	"25"	169	Kanada
6	Timothy	"29"	180	Amerika Serikat
7	Airlangga	"20"		Indonesia

Hasil Persiapan Data

Menyesuaikan Tipe Data

No ID	Nama	Umur	Tinggi Badan	Negara
0	Teresa	23	165	Belgia
1	Alex	20		Amerika Serikat
2	Patricia	18	181	Prancis
3	Michelle	22	172	Singapura
4	Anggun	21	156	Indonesia
5	John	25	169	Kanada
6	Timothy	29	180	Amerika Serikat
7	Airlangga	20		Indonesia

Hasil Persiapan Data

Mengatasi Missing Value (Menggunakan Mean/Rata-rata)

No ID	Nama	Umur	Tinggi Badan	Negara
0	Teresa	23	165	Belgia
1	Alex	20	170	Amerika Serikat
2	Patricia	18	181	Prancis
3	Michelle	22	172	Singapura
4	Anggun	21	156	Indonesia
5	John	25	169	Kanada
6	Timothy	29	180	Amerika Serikat
7	Airlangga	20	170	Indonesia

Follow me!

Instagram : @ikanurfitriani_

Twitter : @ikanurfitriani_

LinkedIn : <https://www.linkedin.com/in/ikanurfitriani>

Short Class Data Science and Analysis
by @myskill.id

