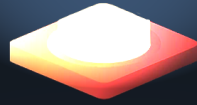




FIND IT!

Future Innovation and Discovery IT!

DATA ANALYTIC COMPETITION



PETUNJUK PESERTA

Deskripsi Data

Data berisi informasi tentang buku yang diambil dari beberapa sumber. Tujuan yang ingin dicapai adalah memprediksi harga jual dari suatu buku berdasarkan fitur-fitur yang sesuai.

Pembagian File

Dataset terbagi dalam 3 file:

1. File public-train.csv. Jumlah data sebanyak 3550 baris, dengan delimiter berupa pipe “|”. Digunakan peserta untuk melatih model Machine Learning.
2. File public-test.csv. Jumlah data sebanyak 500 baris, dengan delimiter berupa pipe “|”. Digunakan peserta untuk menguji model yang dibuat.
3. File private-test.csv. Jumlah data sebanyak 500 baris, dengan delimiter berupa pipe “|”. Digunakan panitia untuk menguji model yang dibuat oleh peserta. File ini **tidak** dibagikan.

Keterangan Kolom

No	Nama Kolom	Keterangan
1	author_id	ID dari penulis buku
2	description	Deskripsi singkat dari buku
3	bookformat	Jenis cetakan buku. Contohnya Hardcover, Paperback, Kindle, dan lain-lain
4	bookedition	Edisi cetakan buku
5	pages	Jumlah halaman
6	published_date	Tanggal diterbitkan pertama kali
7	publisher_id	ID penerbit buku
8	reading_age	Rekomendasi usia minimal untuk membaca buku
9	lexile_measure	Tingkat kesulitan teks dalam skala <i>lexile</i>
10	grade_level	Level pembaca berdasarkan US grade level
11	weight	Berat buku dalam gram
12	rating_value_0	Rata-rata skor dari pembaca pada suatu website portal buku (kode 0)
13	rating_value_1	Rata-rata skor dari pembaca pada suatu website portal buku (kode 1)
14	rating_count_0	Jumlah pembaca yang memberi skor pada website portal buku (kode 0)
15	rating_count_1	Jumlah pembaca yang memberi skor pada website portal buku (kode 1)
16	dimension_0	Ukuran dimensi buku dalam cm (bisa panjang, lebar, atau tebal)
17	dimension_1	Ukuran dimensi buku dalam cm (bisa panjang, lebar, atau tebal)
18	dimension_2	Ukuran dimensi buku dalam cm (bisa panjang, lebar, atau tebal)
19	genre_0	Genre dari buku. Urutan menunjukkan seberapa cocok suatu genre terhadap buku tersebut menurut pembacanya
20	genre_1	
21	genre_2	
22	genre_3	
23	genre_4	
24	genre_5	
25	genre_6	
26	genre_7	
27	genre_8	
28	genre_9	
29	genre_weight_0	

30	genre_weight_1	Bobot dari masing-masing genre. Urutan sesuai dengan fitur genre pada kolom sebelumnya.
31	genre_weight_2	
32	genre_weight_3	
33	genre_weight_4	
34	genre_weight_5	
35	genre_weight_6	
36	genre_weight_7	
37	genre_weight_8	
38	genre_weight_9	
39	price	Harga jual buku pada suatu website yang diambil pada 3-4 Mei 2021, dalam rupiah. Ini adalah target yang butuh diprediksi.

Tugas

Peserta membuat model menggunakan data public-train dan mengujinya menggunakan public-test. Model yang dibuat oleh peserta diserahkan ke panitia dalam bentuk *source code* yang bisa dieksekusi langsung oleh panitia. Panitia akan menguji model dari peserta menggunakan private-test.

Hasil prediksi menggunakan public-test dan private-test akan dihitung dalam RMSE dan menjadi salah satu komponen penilaian.