Rakamin Academy

Homework

Estimasi Waktu Pengerjaan

 3 - 5 jam

Jumlah Soal

 4 Soal

Total Point

 100 poin

Teknis Pengerjaan

1. Pekerjaan dilakukan secara **berkelompok, sesuai kelompok Final Project**
2. Masing-masing anggota kelompok tetap perlu submit ke LMS (jadi bukan perwakilan)
3. Menggunakan dataset berikut ini: Klik
4. File yang perlu dikumpulkan:
   - File **jupyter notebook** (.ipynb) yang berisi source code.
   - File **laporan homework** (.pdf) yang berisi rangkuman dari apa saja yang telah dilakukan.
5. Upload hasil pengerjaan melalui LMS.
   - Masukkan semua file ke dalam **1 file** dengan format **ZIP**.
   - Nama File:

     `Supervised - <Nama Kelompok>.zip`

Youtube Views Prediction

Salah satu implementasi regresi yang dapat dilakukan adalah dengan melakukan prediksi views pada video youtube dengan menggunakan angka statistik atau atribut lain pada videonya.
Data statistik ada pada file `youtube_statistics.xlsx`

	trending_date	title	channel_title	category_id	publish_time	tags	views	likes	dislikes	comment_count	comments_disabled	ratings_disabled	video_error_or_removed	description	No_tags	desc_len	len_title	publish_date
0	2017-11-14	Sharry Mann: Cute Munda (Song Teaser) | Parmi...	Lokdhun Punjabi	1	12:20:39	sharry mann|"sharry mann new song"|"sharry man...	1096327	33966	798	882	False	False	False	Presenting Sharry Mann latest Punjabi Song Cu...	15	920	81	2017-11-12
1	2017-11-14	पीरियड्स के समय, पेट पर पति करता ऐसा, देखकर द'...	HJ NEWS	25	05:43:56	पीरियड्स के समय|"पेट पर पति करता ऐसा"|"देखकर द...	590101	735	904	0	True	False	False	पीरियड्स के समय, पेट पर पति करता ऐसा, देखकर द...	19	2232	58	2017-11-13
2	2017-11-14	Stylish Star Allu Arjun @ ChaySam Wedding Rece...	TFPC	24	15:48:08	Stylish Star Allu Arjun @ ChaySam Wedding Rece...	473988	2011	243	149	False	False	False	Watch Stylish Star Allu Arjun @ ChaySam Weddin...	14	482	58	2017-11-12
3	2017-11-14	Eruma Saani | Tamil vs English	Eruma Saani	23	07:08:48	Eruma Saani|"Tamil Comedy Videos"|"Films"|"Mov...	1242680	70353	1624	2684	False	False	False	This video showcases the difference between pe...	20	263	30	2017-11-12
4	2017-11-14	why Samantha became EMOTIONAL @ Samantha naga ...	Filmylooks	24	01:14:16	Filmylooks|"latest news"|"telugu movies"|"telu...	464015	492	293	66	False	False	False	why Samantha became EMOTIONAL @ Samantha naga ...	11	753	88	2017-11-13

Definisi masing-masing kolom

- trending_date: tanggal ketika video trending
- title: judul video
- channel_title: nama channel
- category_id: kategori video dalam label encoding
- publish_time: waktu publish video
- tags: tag yang digunakan pada video
- views: jumlah views video
- likes: jumlah likes video
- dislikes: jumlah dislikes video
- comment_count: jumlah komentar pada video
- comments_disabled: apakah status komentar dinonaktifkan pada video
- ratings_disabled: apakah rating dinonaktifkan pada video
- video_error_or_removed: apakah video error atau sudah dihapus saat ini
- description: deskripsi video
- No_tags: jumlah tags yang digunakan
- desc_len: panjang kata deskripsi video
- len_title: panjang kata judul video
- publish_date: tanggal publish video

Tugas teman-teman sebagai Data Scientist

- Kerjakan secara berkelompok (team final project)
- Dataset: youtube_statistics.xlsx
  - `df = pd.read_excel('youtube_statistics.xlsx')`
- (1) Lakukan EDA dan preprocessing sederhana (30 point)
  - Jelaskan Fitur mana yang sebaiknya digunakan dari hasil EDA?
- (2) Lakukan feature engineering (10 point)
  - Apakah ada feature tambahan lain yang mendukung? Jelaskan mengapa menggunakan feature tersebut.

Referensi tambahan:
- Feature selection: Google
- Feature engineering terkait date: https://towardsdatascience.com/feature-engineering-on-date-time-data-90f6e954e6b8

Tugas teman-teman sebagai Data Scientist

- (3) Lakukan training model & **prediksi views** sebagai variabel target (40 point)
  - Dapat menggunakan model linear regression ataupun algoritma lainnya (30 point)
  - Lakukan tuning hyperparameter, cari mana model yang paling baik (10 point)
- (4) Evaluasi model dengan metrics RMSE dan R^2 (20 point)
  - Jelaskan & berikan analisis mengapa memilih model tersebut sebagai model akhir yang digunakan.

Referensi tambahan:

- Feature selection: Google
- Feature engineering terkait date: https://towardsdatascience.com/feature-engineering-on-date-time-data-90f6e954e6b8

Submission

- Submit berupa file notebook .ipynb dan document report .pdf
- Template report dapat dilihat disini