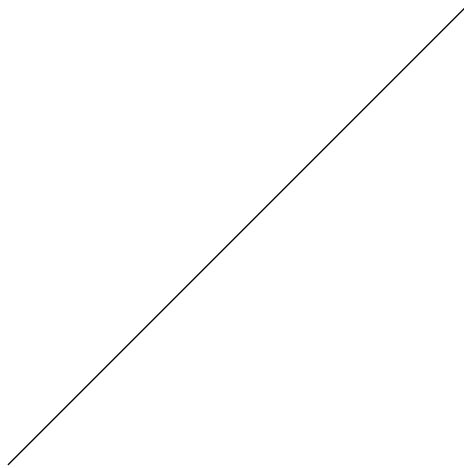


Laporan Hasil Tugas Akhir Big data



Nama : Muhammad yulianto

Nim : 5170411218

Kelas : Big data B

1. Motivasi

Alasan saya menggunakan iseng aja karena suka game Dota 2 juga dan menurut saya menarik juga game kesukaan dibuat seperti ini. Manfaatnya sepertinya tidak ada dan dataset yang saya gunakan diambil dari website kaggle.com dibuat oleh Devin Azelmo.

2. Tujuan

Tujuan saya untuk mengelompokkan durasi game Dota 2 setiap match dan saya kategorikan menjadi 3 bagian early, middle dan late game. Metode yang saya gunakan yaitu menggunakan pengelompokan data atau klastering dengan mengelompokkan data durasi game setiap match.

3. Deskripsi

Data yang saya gunakan yaitu milik Devin Azelmo dan diambil dari website kaggle.com dan jenis datasetnya tidak terstruktur dengan ukuran file 2.56MB dengan keterangan 13 kolom dan jumlah datanya 50.000 baris.

Kolom – kolom berisi match_id, start_time, duration, status_tower, status_barrack, game_mode, dll.

Kemudian jenis datanya

match_id	50000	non-null	int64
start_time	50000	non-null	int64
duration	50000	non-null	int64
tower_status_radiant	50000	non-null	int64
tower_status_dire	50000	non-null	int64
barracks_status_dire	50000	non-null	int64
barracks_status_radiant	50000	non-null	int64
first_blood_time	50000	non-null	int64
game_mode	50000	non-null	int64
radiant_win	50000	non-null	bool
negative_votes	50000	non-null	int64
positive_votes	50000	non-null	int64
cluster	50000	non-null	int64

4. Analisis

1. Preprocessing data

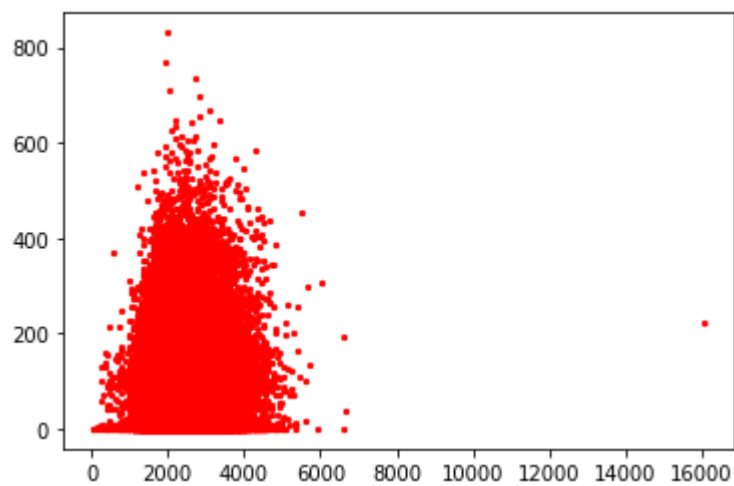
Mengolah data set yang digunakan dari mengambil jenis tipe data apa dan mengambil data yang akan digunakan.

Disini saya menggunakan data duration dan first blood time untuk klasteringnya

	duration	first_blood_time
0	2375	1
1	2582	221
2	2716	190
3	3085	40
4	1887	58

2. Menvisualisasi data

Dengan aplha = 1 diperoleh visual seperti berikut



3. Membuat cluster

- Disini proses dari mengubah data kedalam bentuk array

```
[[2375    1]
 [2582  221]
 [2716  190]
 ...
 [2672  130]
 [2211  645]
 [2788  326]]
```

- kemudian bentuk array tersebut diproses dengan menggunakan MinMaxScaler dan diubah kembali menjadi array yang baru.

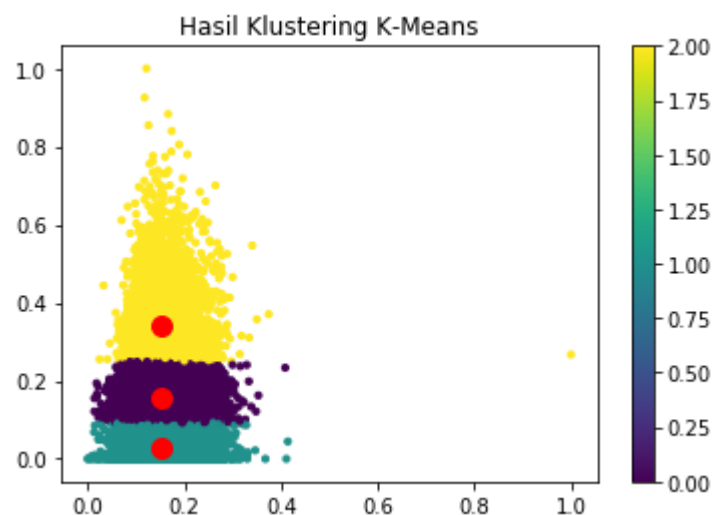
```
array([[0.14494931, 0.00120337],
       [0.15790462, 0.26594465],
       [0.16629115, 0.22864019],
       ...,
       [0.16353736, 0.15643803],
       [0.13468519, 0.77617329],
       [0.17079735, 0.39229844]])
```

- Lalu membuat banyaknya cluster yang mau dibuat, disini saya buat 3 cluster kemudian menghitung nilai K (cluster) untuk menentukan titik tengah cluster dan didapat sebagai berikut

```
[[0.15103439 0.15761331]
 [0.15092318 0.02620612]
 [0.15384376 0.34141974]]
```

- Terakhir memvisualisasikan cluster

Hasil titik tengah tersebut digunakan untuk membuat cluster untuk data lain



5. Diskusi

Dilihat dari sebaran data yang dicluster diperoleh tingkatan durasi game dari early game (game cepat), sedang, dan late game (game panjang) yang didapat dari gabungan durasi dan first blood time. Kemungkinan adanya miss clustering saat pengelompokan data mungkin karena nilai yang didapat hampir sama dengan yang

lain. Mungkin gabungan dari kolom tertentu bisa jadi mendapatkan hasil yang lebih baik saat clustering dan juga menurut saya dataset ini bisa digunakan untuk klasifikasi karena ada kolom target untuk mengklasifikasikan antara menang dan kalah.

6. Kesimpulan

Saya menyimpulkan dari analisis diatas bahwa proses clustering bisa dikatakan tercapai dengan tujuan awal yaitu untuk mengelompokkan data tersebut. Mungkin dengan kombinasi dari kolom lain bisa mendapatkan hasil yang lebih bagus.