IMDb

IMDb adalah sebuah basis data daring informasi yang berkaitan dengan film, acara televisi, video rumahan, permainan video, acara internet, termasuk juga daftar pemeran, biografi kru produksi dan personil, ringkasan alur cerita, trivia, dan ulasan serta penilaian oleh penggemar. Situs web ini sekarang dimiliki oleh Amazon.com. Koleksi informasi film yang ditampilkan cukup lengkap dan dapat juga dilihat informasi film lama maupun film baru yang akan rilis di bioskop. Dalam IMDb terdapat komunitas yang dapat berkontribusi langsung untuk menuangkan ulasan film dan memberikan rating pada film yang diinginkan. Tidak hanya kaum awam, para pakar juga memiliki wadah sendiri untuk memberi rating dan menuangkan ulasan secara profesional pada film tersebut.

Tujuan Penelitian

Pada kali ini kami akan mencoba mengclusterkan film dengan K-Means Clustering pada situs IMDb dengan 3 metode untuk penentuan cluster optimum dan memilih metode yang tepat untuk menentukan jumlah cluster terbaik yang akan digunakan untuk penelitian lebih lanjut.

Metode

- **Analisis Pengelompokan/Clustering** merupakan proses membagi data dalam suatu himpunan ke dalam beberapa kelompok yang kesamaan datanya dalam suatu kelompok lebih besar daripada kesamaan data tersebut dengan data dalam kelompok lain.
- **Metode Elbow** merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah cluster terbaik dengan cara melihat persentase hasil perbandingan antara jumlah cluster yang akan membentuk siku pada suatu titik.
- Pendekatan **Metode Silhouette** digunakan untuk melihat kualitas dan kekuatan cluster, seberapa baik suatu objek ditempatkan dalam suatu cluster.Lebar siluet rata-rata yang tinggi menunjukkan pengelompokkan yang baik.
- **Metode Gap Statistic** membandingkan total variasi intracluster untuk nilai k yang berbeda dengan nilai yang diharapkan di bawah distribusi referensi nol dari data (yaitu distribusi tanpa pengelompokan yang jelas).
- **K-means** merupakan suatu algoritma dalam pengelompokan secara partisi yang memisahkan data ke dalam kelompok yang berbeda-beda.

Dashboard Bisnis Analitik Sakaoni Rofi P dan Haiva Qurrota A Pengantar Data Visualisasi data Clustering K-Means Source Code

Raw Data | Data Final dengan Metode Elbow Klustering K-means

100 records

Search:

film	genre	runtime	rating	votes
Gisaengchung	Comedy	132	8.6	533.468
Black Panther	Action	134	7.3	620.825
Snowpiercer	Action	126	7.1	316.522
Kol	Crime	112	7.1	12.303
Dark Waters	Biography	126	7.6	57.151
Avengers: Age of Ultron	Action	141	7.3	736.654
Oldeuboi	Action	120	8.4	512.172
Ah-ga-ssi	Drama	145	8.1	111.543
Cloud Atlas	Action	172	7.4	343.91
Salinui chueok	Action	131	8.1	137.097
Train to Busan 2	Action	116	5.4	18.429
Lucy	Action	89	6.4	444.18
Busanhaeng	Action	118	7.6	163.626
Training Day	Crime	122	7.7	388.073

Dashboard Bisnis Analitik | Sakaoni Rofi P dan Haiva Qurrota A | Pengantar | Data | **Visualisasi data** | Clustering K-Means | Source Code

Variabel:

- (•) Runtime
- ( ) Rating
- ( ) Votes

Number of bins:

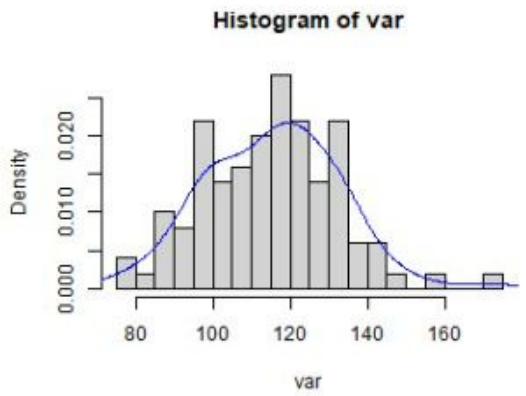
20

Bandwidth adjustment:

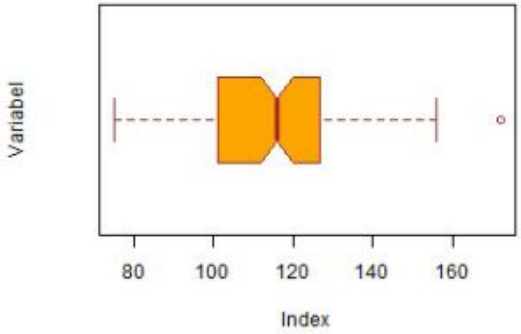
0.2 1 2

0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2

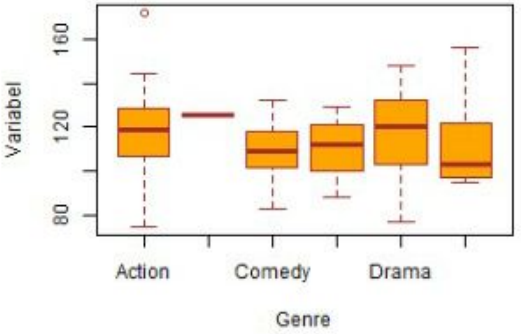
Histogram



Boxplot



Boxplots for each Genre



Summary Runtime

```
   Min. 1st Qu.  Median    Mean 3rd Qu.
  Max.
   75.0   101.0   116.0   115.5   126.2
172.0
```

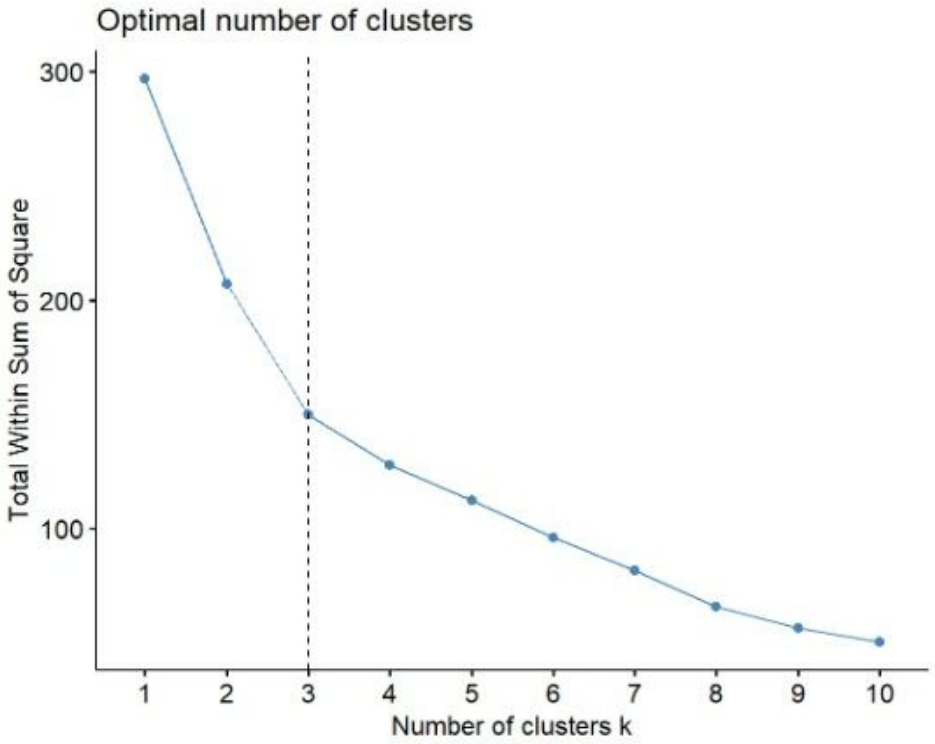
Summary Rating

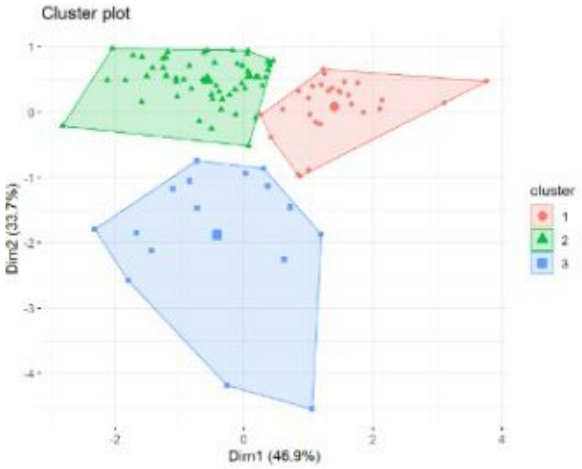
```
   Min. 1st Qu.  Median    Mean 3rd Qu.
  Max.
   3.100   6.275   7.000   6.868   7.600
8.600
```

Summary Votes

```
   Min. 1st Qu.  Median    Mean 3rd Qu.
  Max.
   1.022  13.273  56.760 135.036 159.176 9
70.000
```

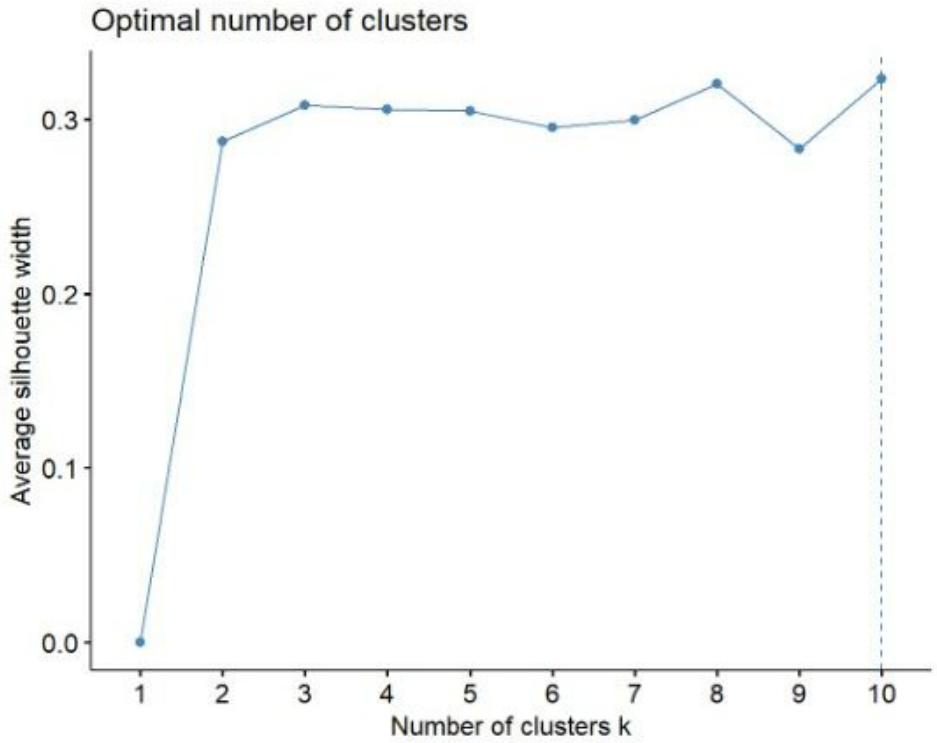
Metode Elbow | Metode *Silhouette* | Metode *Gap Statistic*

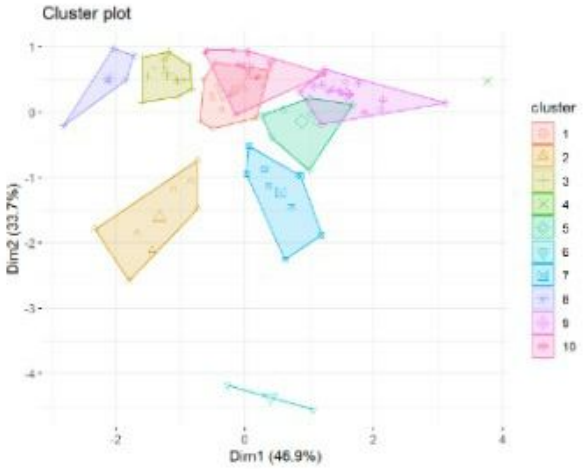




K Means dengan K=3

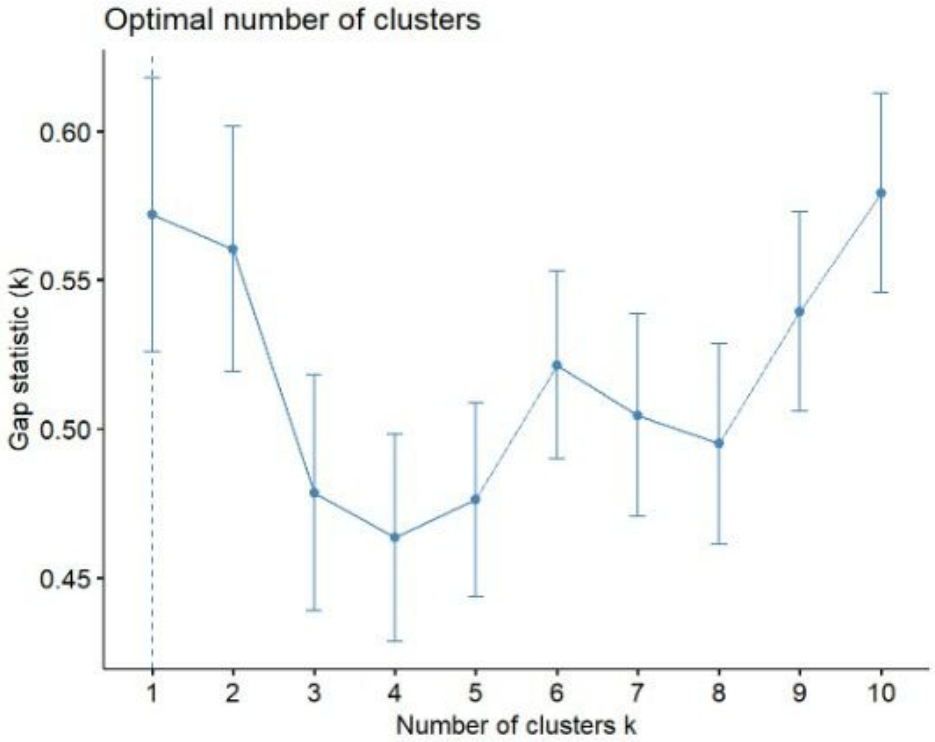
Metode Elbow | Metode *Silhouette* | Metode *Gap Statistic*

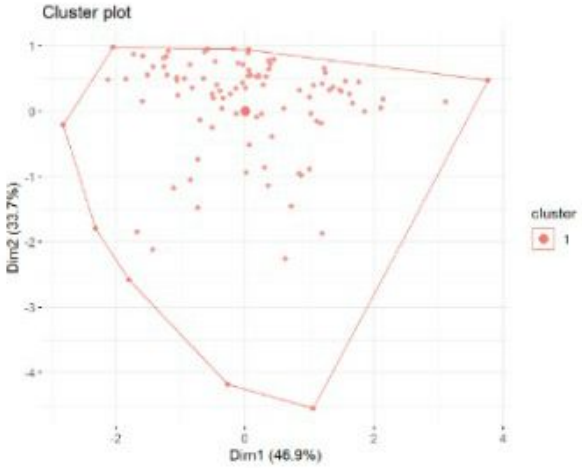




K Means dengan K=10

Metode Elbow | Metode *Silhouette* | Metode *Gap Statistic*





K Means dengan K=1