

Kelompok 3

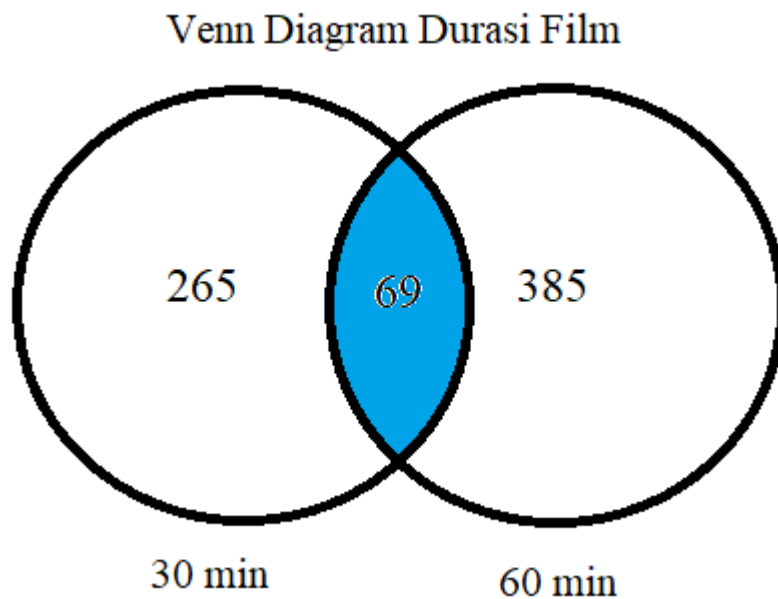
Nama	NIM	Pembagian Tugas
Anisa Meilia	H1D021003	Gabungan dan Komplemen
Haryo Bimantoro	H1D021071	Irisan
Faris Akbar Abimanyu	H1D021072	Beda Setangkep
Muhammad Salman Farrisi	H1D021088	Selisih dan Perkalian Kartesian

1. DATASET YANG DIGUNAKAN

Dataset yang digunakan adalah dataset IMDB Movie yang bersumber dari Kaggle. Dataset IMDb mencakup informasi dasar tentang film, seperti judul, genre, tahun rilis, rating, aktor dan durasi. Setiap film dapat memiliki lebih dari satu genre (misalnya, Action, Adventure, Drama), dan juga memiliki rating tertentu (misalnya, 7.5, 8.0). Informasi ini bermanfaat untuk analisis terkait preferensi penonton, tren genre, dan kualitas film berdasarkan rating.

2. ANALISIS DAN VISUALISASI HIMPUNAN

A. Irisan



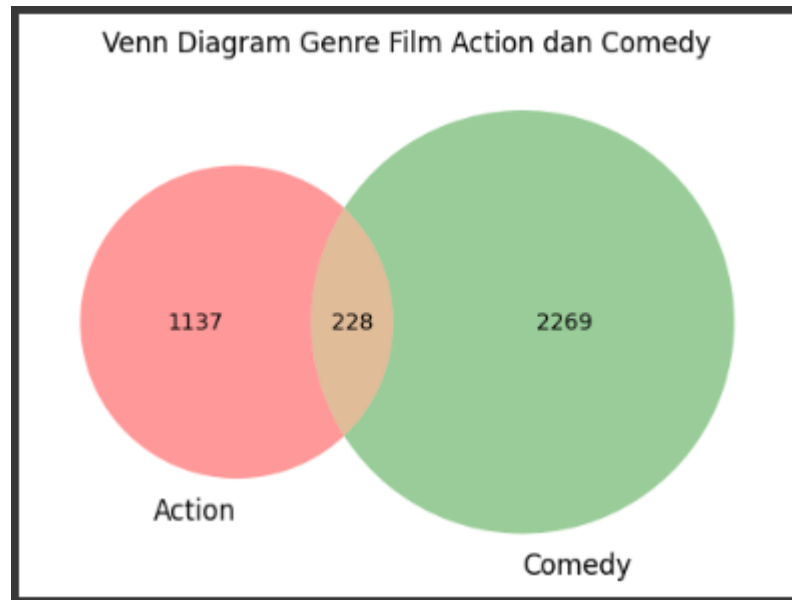
Notasi : $A \cap B = \{x \mid x \in A, x \in B\}$

Diagram venn menunjukkan Irisan antara dua himpunan dari Durasi pada beberapa film, yaitu durasi 30 min (A) dan durasi fil 60 min (B) berdasarkan dataset IMDb. Pada Himpunan A yaitu himpunan yang mempunyai film-film yang durasinya adalah 30 menit, sedangkan Himpunan B yaitu himpunan yang film-film tersebut mempunyai durasi film 60 menit. Irisan antar himpunan A dan B, ditulis sebagai $A \cap B$ adalah himpunan yang mencakup irisan film yang lebih besar atau lebih kecil dari kedua lingkaran tersebut. Area gabungan pada Diagram Venn adalah seluruh wilayah yang mencakup kedua lingkaran, yaitu:

- Film yang memiliki durasi 30 menit pada bagian lingkaran A
- Film yang memiliki durasi 60 menit pada bagian lingkaran B
- Film yang memiliki durasi lebih besar atau lebih kecil pada bagian lingkaran yang di tindih di tengah

Terlihat pada diagram venn diatas terdapat 49 film yang durasinya ditengah-tengah 30 menit dan 60 menit.

B. Gabungan



Notasi : $A \cup C = \{ x \mid x \in A \text{ atau } x \in C \}$

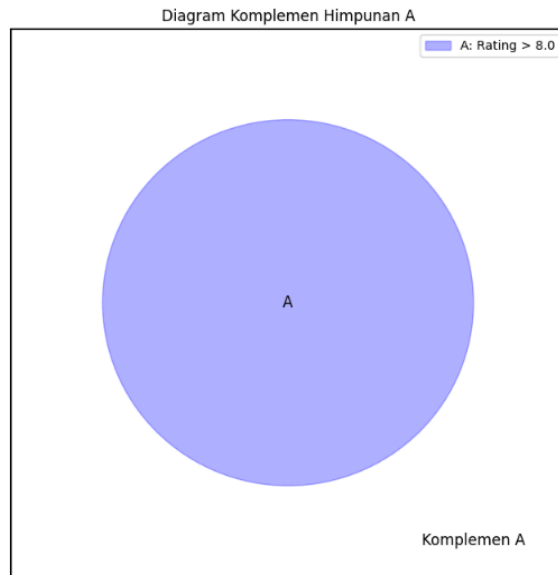
Diagram Venn diatas menggambarkan hubungan antara dua himpunan genre film, yaitu Action (A) dan Comedy (C), berdasarkan data dari dataset IMDb. Himpunan A (Action) berisi film-film yang memiliki genre Action. Himpunan C (Comedy) berisi film-film yang memiliki genre Comedy. Gabungan antara himpunan A dan C, yang ditulis sebagai $A \cup C$ (A union C), adalah himpunan yang mencakup semua film yang memiliki genre Action, Comedy, atau keduanya. Area gabungan di diagram Venn adalah seluruh wilayah yang mencakup kedua lingkaran, yaitu:

- Film yang hanya memiliki genre Action (bagian hanya dalam lingkaran A).
- Film yang hanya memiliki genre Comedy (bagian hanya dalam lingkaran C).
- Film yang memiliki kedua genre Action dan Comedy (bagian yang tumpang tindih antara kedua lingkaran).

Dapat dilihat pada diagram Venn gabungan diatas, maka gabungan genre Action dan Comedy akan mencakup kedua lingkaran (baik yang tumpang tindih maupun tidak).

Ini berarti film yang masuk dalam kategori Action atau Comedy, atau keduanya, akan berada di dalam gabungan ini.

C. Komplemen



$U = \{ \text{semua rating film} \}$

$A = \{ x | x > 8.0 \}$

$A^c = \{ \dots ; 7,8 ; 7,9 ; 8,0 \}$

Semesta (U) adalah semua rating film dalam dataset, yang digambarkan sebagai kotak. Himpunan A adalah film dengan rating lebih dari 8.0, yang digambarkan dalam lingkaran di dalam kotak semesta. Sedangkan Komplemen A adalah film dengan rating kurang dari 8.0 dan 8.0, yang digambarkan diluar lingkaran Himpunan A dan bagian Semesta. Diagram ini membantu memvisualisasikan pemisahan antara film dengan rating tinggi (>8.0) dan film dengan rating yang tidak memenuhi syarat (>8.0). Gabungan semesta dengan himpunan A dan komplemen A memberikan gambaran bahwa setiap film dalam dataset memiliki kategori yang jelas: apakah ia berada dalam himpunan A (rating tinggi) atau komplemen A (rating rendah).

D. Selisih

Selisih dalam himpunan, yang sering disebut juga sebagai perbedaan himpunan atau komplemen relatif, adalah operasi yang menghasilkan himpunan baru dengan elemen-elemen yang ada pada himpunan pertama tetapi tidak ada pada himpunan kedua. Secara formal, jika (A) dan (B) adalah dua himpunan, maka selisih antara (A) dan (B) (dilambangkan dengan (A - B)) adalah himpunan yang terdiri dari elemen-elemen (x) yang ada di (A) namun tidak ada di (B).

Secara matematis, dituliskan sebagai:

$$[A - B = \{ x \text{ mid } x \text{ in } A \text{ dan } x \text{ notin } B \}]$$

Sebagai contoh, jika (A = {1, 2, 3, 4}) dan (B = {3, 4, 5, 6}), maka selisih (A - B = {1, 2}), karena elemen 1 dan 2 ada di (A) tetapi tidak ada di (B).

Selisih ini sangat berguna dalam banyak konteks, terutama dalam teori himpunan, di mana kita sering kali tertarik untuk mengetahui elemen-elemen yang unik atau berbeda antara dua kelompok data atau entitas.

Dari dataset yang kita gunakan, kita akan mengambil kolom *year* dan *title*. Dari kolom *year* kita akan melakukan filter lagi. Pada kesempatan kali ini, kita akan mengambil untuk kolom *year* yang bernilai 2021 dan 2022. Artinya kita akan memiliki 2 himpunan, yaitu himpunan 2021, himpunan 2022, dan irisan antara himpunan 2021 dan 2022. Himpunan 2021 kita sebut Himpunan A dan Himpunan 2022 kita sebut Himpunan B.

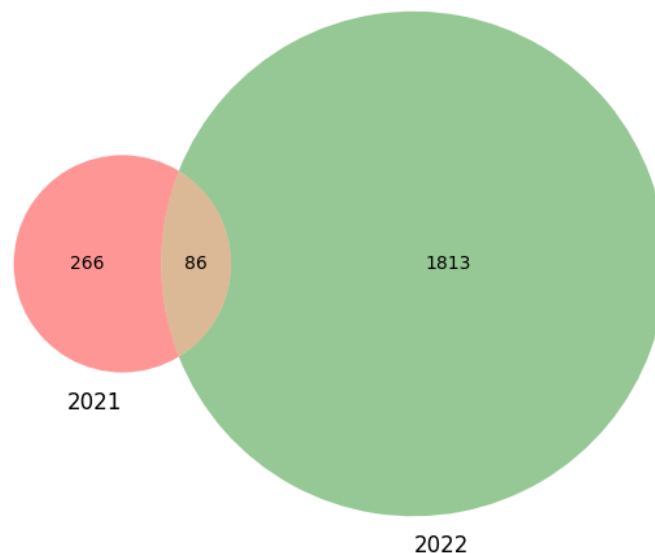
$A = \{\text{Ariana Grande \& Kid Cudi - Just Look Up; The Vivo Songbook; Black Island Hilda and the Mountain King; Arlo the Alligator Boy;; Squid Game}\}$
 $|A| = 352$

$B = \{\text{Adib - Netflix; The Last Movie Ever Made: The Don't Look Up podcast; CheNetflix: El podcast; Un domingo de Ramos. Un podcast de Netflix;; The Privilege}\}$

$|B| = 1899$

$|A \cap B| = 86$

Venn Diagram for Sets 2021 and 2022



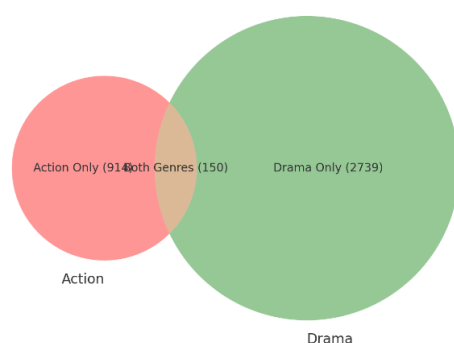
Dari data yang saya sebutkan terlihat bahwa kardinalitas himpunan A adalah 352 film dan kardinalitas himpunan B adalah 1899. Kardinalitas Irisan himpunan A dan B adalah 86. Langkah selanjutnya kita akan mencari selisih, dengan cara melakukan selisih antara A dan B, dengan rumus $A - B$ dan $B - A$. Karena film yang masuk himpunan sangat banyak dan tidak mungkin untuk menuliskan semua film yang ada maka akan lebih mudah jika kita mencari selisih kardinalitasnya saja. Selisih $A - B$ berarti himpunan A yang tidak masuk di Himpunan B dan irisan antara Himpunan A, Himpunan B. Selisih $B - A$ berarti himpunan B yang tidak masuk di Himpunan A dan Irisan Himpunan A, Himpunan B.

(i) $|A - B| = 266$

(ii) $|B - A| = 1813$

E. Beda Setangkup

Diagram Venn dengan Irisan Antara Genre Action dan Drama



Notasi: $A \triangle B = \{x | x \in A \text{ atau } x \in B, \text{ tetapi } x \notin (A \cap B)\}$

Beda Setangkup antara Action (A) dan Drama (B) mencakup film-film yang hanya terdapat dalam salah satu genre, yaitu Action saja atau Drama saja, tanpa adanya film yang memiliki kedua genre tersebut sekaligus.

Variabel dalam Diagram Venn

1. Set A (Action): Berisi film-film yang memiliki genre Action.
2. Set B (Drama): Berisi film-film yang memiliki genre Drama.
3. Area Only Action: Film yang hanya bergenre Action, dengan total 914 film.
4. Area Only Drama: Film yang hanya bergenre Drama, dengan total 2739 film.

Contoh Film dalam Beda Setangkup

1. Film Hanya Genre Action (beberapa contoh dari total 914):
"Toy Boy", "The Idhun Chronicles", "Untitled Netflix Ghostbusters Animated Series", "Johan Falk: Organizatsija Karayan", "The Expendables 2"
2. Film Hanya Genre Drama (beberapa contoh dari total 2739):
"CT Fletcher: My Magnificent Obsession", "Night Light", "The Iron Bridge", "The Intent", "Anatomy of a Scandal"

Dalam diagram Venn di atas:

- 1) Lingkaran kiri mewakili genre Action, yang berisi semua film bergenre Action.
- 2) Lingkaran kanan mewakili genre Drama, yang berisi semua film bergenre Drama.
- 3) Area di luar perpotongan menunjukkan film yang eksklusif hanya dalam satu genre saja, baik Action maupun Drama, sehingga sesuai dengan konsep Beda Setangkup.

Beda Setangkup antara genre Action dan Drama memberikan gambaran bahwa ada 914 film yang hanya bergenre Action dan 2739 film yang hanya bergenre Drama, tanpa ada irisan antara kedua himpunan ini dalam konteks symmetric difference.

F. Perkalian Kartesian

Perkalian Kartesian, atau produk Cartesian, adalah operasi yang menghasilkan himpunan pasangan terurut dari dua himpunan yang berbeda. Jika A dan B adalah dua himpunan, maka perkalian Kartesian $A \times B$ adalah himpunan yang terdiri dari semua pasangan terurut (a, b) di mana $a \in A$ dan $b \in B$.

Secara formal, produk Cartesian antara dua himpunan A dan B didefinisikan sebagai:

$$A \times B = \{ (a, b) \mid a \in A, b \in B \}$$

Sebagai contoh, jika $A = \{1, 2\}$ dan $B = \{x, y\}$, maka produk Cartesien $A \times B$ adalah:
 $A \times B = \{ (1, x), (1, y), (2, x), (2, y) \}$

Operasi ini sangat penting dalam berbagai cabang matematika, terutama dalam teori himpunan, relasi, dan geometri, di mana konsep ruang berdimensi lebih tinggi sering kali dibangun melalui produk Cartesien dari himpunan-himpunan satu dimensi. Perkalian Kartesian juga dapat diperluas untuk lebih dari dua himpunan, menghasilkan pasangan terurut atau tupel yang lebih panjang, tergantung pada jumlah himpunan yang terlibat.

Pada Dataset ini kami mencoba menggunakan kolom Certificate dan Jumlah Penonton. pada hal ini kami coba menggunakan 4 item

Himpunan Certificate (C) = { TV-Y, TV-PG, TV-MA }

Himpunan Jumlah Penonton (J) = { 5.125.931; 3.959.899; 2.898.989 }

$$C \times J = \{(TV-Y;5.125.931);(TV-PG; 3.959.899); (TV-MA;2.898.989)\}$$

3. REFERENSI

- Kaggle : <https://www.kaggle.com/datasets/payamamanat/imdb-dataset>

4. BUKTI Pengerjaan

