

# Final Project Data Mining dan Business Intelligence (Reguler) dan Natural Language Processing (IUP) - Studi Kasus Cineplex21

**Link Video :** <https://www.youtube.com/watch?v=-pnSOD5HtE0>

**Link PDF Presentasi, Dataset hasil, Notebook pengolahan data:**

<https://github.com/vincentmichael089/Twitter-Sentiment-Analysis>

## **Nama Kelompok :**

1. Vincent Michael Sutanto - 16/398531/PA/17492 (Datamining Reguler)
2. Rimba Erlangga - 16/398526/PA/17487 (Datamining Reguler)
3. Septian Wijaya - 16/398528/PA/17489 (Datamining Reguler)
4. Emerald Makerti - 16/398501/PA/17462 (Datamining Reguler)
5. M. Azmi Aris - 16/39/PA/17 (Datamining Reguler)
6. Fahmi Zufarian S - 16/392766/PA/17070 (NLP IUP)
7. Rani Rizkiani Ilyas - 16/398525/PA/17486 (NLP IUP)S

## **Analisis SWOT Perusahaan Berdasarkan Data:**

### **-Strength :**

Cineplex 21 memiliki total 1.140 layar, yang tersebar di 49 kota di 207 lokasi di seluruh Indonesia (website resmi cineplex21). Angka-angka tersebut terpaut jauh merk bioskop lainnya. Menunjukkan bahwa jumlah layar dan jumlah lokasi bioskop cineplex 21 merupakan kekuatan atau strength dari perusahaan. Dipastikan bahwa cineplex 21 masih sangat mampu untuk menjaring penonton secara masif dan memiliki kekuatan untuk melawan pesaing-pesaingnya.

### **-Weakness :**

Engagement Rate yang rendah merupakan salah satu weakness dari cineplex 21. Engagement Rate adalah metrik yang menilai bagaimana suatu konten dapat diterima dengan baik oleh audiensnya. Engagement Rate sendiri sering digunakan untuk mengevaluasi dan menganalisis pemasaran barang atau produk. Skor Engagement Rate yang dimiliki akun sosial media hanya mencetak angka 0.38%, sedangkan idealnya engagement rate dengan akun setara adalah sebesar 2.98%. Skor ini menunjukkan kurangnya interaksi mendalam antara cineplex21 dengan pengikutnya.

### **-Opportunity :**

Menurut data yang didapat dari website katadata.co.id, dari tahun 2015 sampai dengan tahun 2018, terus terjadi kenaikan angka jumlah penonton bioskop. Hal ini menandakan bahwa Supply dan Demand berjalan seimbang. Dalam hal ini supply adalah strength atau kekuatan dari cineplex21 yaitu jumlah layar bioskop yang sangat banyak dan demand adalah kenaikan jumlah penonton. Badan Ekonomi kreatif juga menargetkan peningkatan jumlah penonton hingga 60 juta penonton di akhir tahun 2019, yaitu 15.4% lebih tinggi dari tahun 2018.

Era Online Ticketing juga menjadi salah satu kesempatan yang menarik. Selain memudahkan penonton dalam membeli tiket, adanya promo yang sering diadakan oleh penyedia jasa online ticketing juga berpotensi untuk menaikkan jumlah angka penonton.

### **-Threat :**

Layanan streaming online saat ini belum menjadi ancaman yang serius bagi industri bioskop (survey PostTrak. PostTrak adalah sebuah layanan survey audiens bagi studio film). Hasil survey menunjukkan bahwa penonton yang melakukan streaming, juga suka menonton bioskop. Namun tidak menutup kemungkinan bahwa online streaming dapat menjadi ancaman yang serius jika perusahaan bioskop lengah dan tidak mampu bersaing dengan online streaming di masa depan.

## Strategi Mengatasi Weakness dan Threat:



Disusun strategi untuk mengatasi Weakness dan Threat dari analisis SWOT yang dilakukan.

Dari sisi kelemahan, disusun strategi marketing media sosial yang baik dengan harapan terjadi perbaikan angka engagement rate dan juga terjadi peningkatan jumlah penonton. Peningkatan jumlah penonton juga harus ditangani dengan baik salah satunya terdapat pada strategi penanganan ancaman.

Dari sisi ancaman, dapat dibentuk suatu model optimasi penjadwalan. Optimasi model penjadwalan dimaksudkan agar penayangan film di bioskop selalu up to date, dalam konteks selalu rilis tepat waktu dan pergi tepat waktu juga. Optimasi model ini ditujukan agar penonton tidak menjadi jenuh akibat film yang ditayangkan terlalu lama sehingga mereka malah beralih ke online streaming. **Pembahasan seterusnya akan menitikberatkan dalam membuat model untuk mengatasi ancaman**

### Batasan Masalah Strategi:

Strategi yang diusulkan hanya berlaku untuk diterapkan pada bioskop yang segmen penontonya adalah yang memiliki preferensi untuk menonton film barat.

### Harapan Penerapan Strategi:

1. terjadi kenaikan jumlah penonton dan juga pendapatan.
2. penjadwalan film menjadi lebih optimal dan sebisa mungkin tidak membosankan penonton.
3. kenaikan dan hasil pengoptimalan diharapkan dapat dimanfaatkan sebagai portofolio advertising

### Pembentukan Model *Up-to-date* : Rilis tepat waktu dan pergi tepat waktu

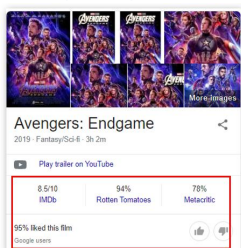
**Dataset** : 90 Judul Film barat (2016-2019) dengan atribut:

1. Skor Metacritic, Rotten Tomatoes, IMDb, dan Google User scrap dari Google
2. Genre, Bulan Rilis, Rumah produksi dan distributor scrap dari Google
3. Lama penayangan (diambil dari google trends dengan threshold 20, penjelasan pada file pdf presentasi)
4. Sentiment Analysis twitter diambil kelas persentase positif dan negatif untuk tiap judul film

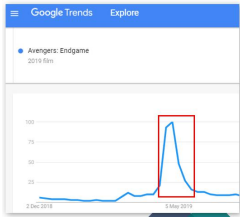
**Model** : Decision Tree dengan alasan pemilihan model:

1. Sederhana
2. Menangani data kategorikal dan numerik
3. White-box , mudah untuk dijelaskan

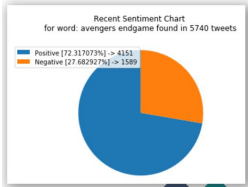
### Pengumpulan Dataset



1. skor audiens diharapkan dapat menjadi parameter yang mempengaruhi model, karena biasanya film akan rilis lebih dahulu di luar negeri dari pada di Indonesia. Sehingga diharapkan skor ini merepresentasikan tingkat kepuasan penonton luar negeri yang menonton film tersebut lebih dulu. Selain itu diambil juga data kategorikal yaitu rumah produksi, genre, bulan rilis, dan distributor film



2. Pengambilan data lama penayangan diambil dari Google Trends, hal ini dikarenakan tidak ditemukannya layanan yang men-archive rata-rata suatu judul film ditayangkan di bioskop. Data Google Trend dibentuk dengan format timeseries, yaitu interest penonton terhadap suatu film dalam dimensi waktu. Biasanya ketika film rilis, indikator akan menyentuh angka 100, dan akan menurun setiap minggunya. Penurunan interest per minggu tersebut akan dicatat sampai indikator menyentuh angka 20. Hal ini dikarenakan ketika indikator menyentuh angka 20, maka minggu-minggu sesudahnya tidak akan naik atau turun dari angka 20. Kelas penayangan yang didapat dari Google Trends dibagi ke dalam kelas 2,3,4,5, dan 6 minggu,



3. Sentimen analysis ini digunakan untuk mengetahui tingkat semangat netizen terhadap suatu film yang akan dirilis. Untuk setiap judul film diambil sebanyak 10.000 tweet dari rentang waktu 2 bulan sebelum film tersebut dirilis. Tweet yang telah diambil kemudian akan diproses untuk diketahui sentiment analysis untuk masing-masing tweet. Sentiment Analysis kemudian akan direpresentasikan dalam bentuk persentase positif dan persentase negatif. Untuk catatan, total tweet yang digunakan untuk representasi sentiment analysis karena terdapat kelas netral yang dibuang. kelas netral menunjukkan bahwa tweet tidak bersentimen positif, ataupun negatif.

## Insight dari Dataset

nama
Joker
Spider-Man: Far From Home
Aladdin
Black Panther
Incredibles 2
Spider-Man: Into the Spider-Verse
Bohemian Rhapsody
Aquaman
Ralph Breaks the Internet
Bumblebee
War for the Planet of the Apes
Coco
Kong: Skull Island
Power Rangers
Jumanji: Welcome to the Jungle
The Fate of the Furious
Despicable Me 3
Cars 3
Annal
Suicide Squad
Moana
Findino Dori
Central Intelligence
Train to Busan
Kung Fu Panda 3

16 dari 25 film dengan lama tayang 5-6 minggu adalah film keluarga

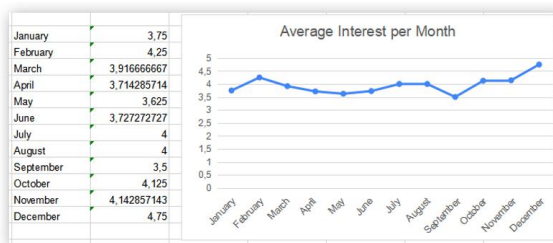
Film keluarga memiliki pasar yang luas sehingga mampu bertahan lama

Bioskop masih menjadi pilihan untuk *Quality-time* bersama keluarga

Online Streaming tidak bisa menjadi sarana rekreasi seperti bioskop

1. Film keluarga memiliki pasar yang luas, dari anak-anak sampai dengan orang tua, sehingga tidak heran film keluarga bertahan lama di bioskop

2. Bioskop masih menjadi pilihan untuk Quality Time bersama keluarga, atau dikatakan bioskop sebagai sarana rekreasi keluarga. Hal ini merupakan insight yang bisa mengalahkan ancaman atau threat dari online streaming, dimana online streaming tidak bisa digunakan sebagai sarana rekreasi seperti bioskop



Bulan Desember memiliki rerata penayangan tertinggi (4.75 minggu)

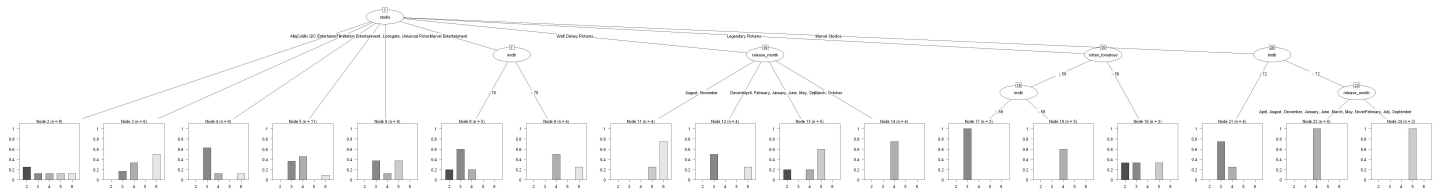
Sarana Rekreasi keluarga saat libur panjang

3. pada bulan desember rerata penayangan menyentuh angka 4.75 minggu. Angka ini lebih tinggi jika dibandingkan bulan-bulan lainnya. Kemungkinan yang mungkin terjadi adalah bulan Desember merupakan bulan yang identik dengan bulan liburan keluarga, sehingga bioskop dipilih sebagai salah satu sarana rekreasi keluarga terutama keluarga perkotaan.

## Penggunaan Model

Menggunakan bahasa pemrograman R dengan library G5.0 untuk model Decision Tree yang tingkat akurasi mencapai 85.0% dengan input dataset yang telah dibuat. File model dapat dilihat pada link github yang tertera diatas.

## Visualisasi Model



## Insight dari Model

Dari berkali-kali percobaan yang dilakukan, didapatkan fakta pada mayoritas percobaan, atribut sentimen negatif, nilai audiens metacritics, dan genre film tidak mempengaruhi model decision tree.

Sedangkan untuk atribut lainnya masih sering muncul ketika seed dari decision tree diubah. Untuk model yang mencapai akurasi 85% didapatkan bahwa atribut studio produksi sangat mempengaruhi klasifikasi lama minggu penayangan. Skor audiens imdb dan rotten tomatoes juga selalu hadir dalam atribut yang digunakan. Bulan rilis dan juga sentimen positif sering muncul juga namun dalam persentase penggunaan yang lebih sedikit.