DATA VISUALIZATION DALAM MEMPROYEKSIKAN PERFILMAN HOLLYWOOD DENGAN METODE K-MEANS++, RETURN ON INVESTMENT DAN ASSOCIATION (Studi Kasus: Data Film dari Website Internet Movie Data Base dan The Numbers dari tahun 1997 -2016)

Baron Setyo Utomo

Program Studi Statistika Fakultas MIPA

Universitas Islam Indonesia

INTISARI

Website Internet Movie Data Base (IMDB) merupakan penyedia layanan database dan informasi film milik Amazon. IMDB me*review* sebuah film dengan cukup lengkap. *The Numbers* adalah situs *web* yang melacak pendapatan *box office* dan proyeksi penjualan film secara sistematis dan algoritmik. situs ini telah menjadi *data base* informasi industri film terbesar yang tersedia secara gratis di web dengan lebih dari 20.000 judul film. Data dari kedua *website* tersebut digabungkan sehingga diperoleh dataset film yang kemudian di analisis menggunakan metode *k-means++* serta *association*, selain itu dilakukan analisis *Return on Investment* guna mencari tahu karakteristik film yang menguntungkan untuk investasi. Berdasarkan hasil analisa *cluster k-means++* diperoleh 3 *cluster* yang sebelumnya telah dianalisa menggunakan metode *silhouette coefficient* untuk validasi jumlah *cluster*. Berdasarkan analisis *Return on Investment diketahui bahwa genre film yang memiliki nilai rata-rata ROI tertinggi adalah genre drama, family, fantasy. Terdapat hubungan asosiasi antar atribut seperti genre, companies, director, artis dan rating.*

Kata Kunci : *IMDB, The Numbers, Film, K-means++, Silhouette Coefficient, ROI, Association.*

DATA VISUALIZATION FOR PROJECTING HOLLYWOOD MOVIES USING K-MEANS++ METHOD , RETURN ON INVESTMENT AND ASSOCIATION
(Case Study: Movie Data from Internet Movie Data Base Website and The Numbers from year 1997 to 2016)

Baron Setyo Utomo

Department of Statistics Faculty of Mathematics and Natural Science

Islamic University of Indonesia

ABSTRACT

Internet Movie Database (IMDb) is a movie informations and services database provider owned by Amazon. IMDb reviews a movie quite complete. The Numbers is a website that tracking the earnings of box office and the forecast sales of movies with systematic and algorithmic ways. This site has become the biggest database informations of movies nowadays that can be accessed for free with more than 20.000 of movie titles. Combination of those websites will gain movie dataset that will be analyzed using k-means++ and association methods. In addition, analyzing return of investment is needed to know the characteristics of movie that gain profit for investment.Based on cluster analyzing using k-means++ method, it obtained 3 clusters that has been analyzed using silhouette coefficient method for validating the amount of cluster. Based on return on investment known that the genres which have the highest ROI average number are drama, family and fantasy. There is association between attributes like genre, company, director, actor and rating.

Keywords: IMDB, The Numbers, Film, K-means ++, Silhouette Coefficient, ROI, Association.