

TV SHOW POPULARITY ANALYSIS USING DATA MINING

Dr Vijay Gupta

*(Associate Professor, MCA Department,
International School of Informatics and Management, Jaipur)*

Ms. Khushboo Choudhary

*(Student, MCA Department International School of Informatics
and Management, Jaipur)*

Abstract

Television plays a crucial role in today's era with various similar categories of television shows being telecasted on daily basis. As an outcome, it produces opposition amongst makers and it is actually tough for the end users to make a selection amongst diverse Television plays. To pattern the approval for these shows Television Rating Point (TRP) is computed with data mining methodologies. Accurate and appropriate prediction of TV show popularity is of great significance for content providers, promoters, and broadcast television operators. This information may be valuable for operators in TV program procuring choices and might help promoters to articulate low-cost advertising venture suggestions. This paper presents enhanced TV show popularity prediction analysis with K-Means Clustering and R studio. The motive of this paper is to evaluate the recognition of TV Plays and also examine the people enjoying a specific show and predicting popularity of that shows, depending upon the user reviews using R-studio.

Keywords: Data Mining, Prediction Analysis, K-Means Clustering.

Introduction

Television community is fast expanding with the dissolute extension of hypermedia and network broadcast proficiency. Moreover, abundant television insides with additional consumption practices granted to consumers has also increased viewership.

Data Mining can be defined as a methodical process intended for inspecting data for expedition of reliable frameworks and/or competent links amongst variables, and to authorize the outcomes by realizing the differentiated frameworks to innovative subclasses of data. Data mining can also be defined as a practice of involuntarily realizing treasured data in massive data warehouses. It's core objective is forecasting and predictive data mining is the most common kind of data mining. The purpose of establishing data mining procedures is defining substantial frameworks from datasets and also propose capabilities and abilities to estimate the outcome of a forthcoming learning.

Cluster Analysis is the process of isolating data into groups of equivalent objects. It patterns data through its group of clusters. Data modelling places clustering in an ancient outlook engrained in computation, statistics, and arithmetical assessment. K-means clustering as the greatest spontaneous and widespread clustering practice, which is segregating a dataset into K number of clusters in the zone of its commencement like that an unprejudiced role well-defined in situations of the whole within-group sum-of-squares is diminished.

K-Means Clustering Algorithm

The K-means clustering is a known segregating method. The objects are categorized into any of the K-groups. Outcomes of Partitioning approaches are a set of K clusters, every single object of data set is going to a single cluster.

Every cluster contains a cluster representative. As stated earlier, clustering can be defined as a separation of a dataset into number of groups so that similar items fall to similar groups. To assemble the database, algorithm uses a repetitive pattern. The input required is the number of anticipated clusters and the preliminary means and creates concluding means as output. If the algorithm need is to generate K clusters then there will be K primary means and final means too.

To collect the data items in dataset in figures of groups, the following algorithm is used. To accomplish this mission, it creates certain repetition till certain meeting standards are encountered. After repetition of individual lately computed means are changed so that they come nearer to the concluding means. And at last, the algorithm unites and then halts executing repetitions.

k-means clustering involves subsequent two inputs:

k = number of clusters

Training set(i) = {a1, a2, a3....., ai}

Tool Used for K-Means Clustering

Algorithm Implementation

The instrument that is used for the execution of k-means clustering procedure is the R Studio.

Essentially, R software is mostly inscribed in C, FORTRAN and R. It is upheld in Mac OS, Windows and Linux. R has objects that can be installed with C, C++, java, .net, Python. R studio is the frequently suggested Graphical User Interface for R. To envision data in R, certain packages might necessary to be installed. Installing these packages may produce extra supplementary memory in the main system. Most of the packages in RStudio are completely reachable in each data mining apparatus presented at CRAN and GitHub, which construct it diverse in comparison to others. Approximately entire machine learning measures are available in R.

The leading usage of R is in finances, genetics, machine learning fields.

Experiments and Observations

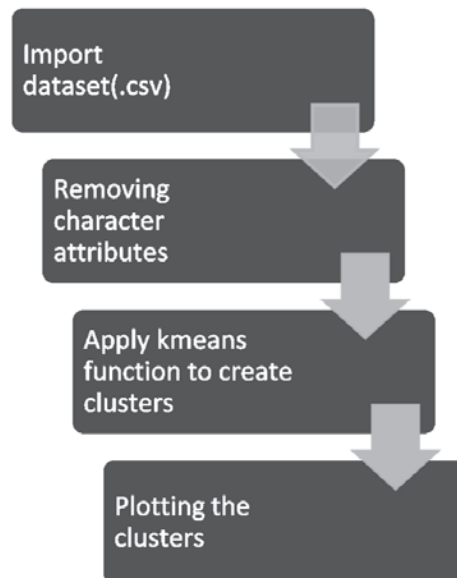
Dataset

The experimental data originate from BARC (Broadcast Audience Research Council India), which is one of the largest broadcast TV platforms in India. The data set is summarised in Table 1. It is a sample dataset that covers broadcast TV period over 7 days between February 1st and February 7th, 2019. The dataset contains 4 attributes and 57 instances. We have taken prime-time shows from topmost evaluated channel in India. The primary attribute is the Telecast By channel which comprises 4 most widespread channels network namely STAR Plus, Zee TV, Sony Entertainment Television, Colors. The next attribute is the TV Serial which contains the popular shows of these channels. The third attribute is the Genres which designates the category of these shows like Drama, Fiction etc. The last attribute is the View Counts of the viewers.

Telecast by Channel	TV Serial	Genres	View Count
Zee TV	Aap Ke Aa Jane Se	Soap Opera	4896
Sony Entertainment Television	C.I.D. (1998-)	Crime Fiction	3770
Sony Entertainment Television	Chandragupta Maurya (2018-)	Drama	7145
Colors	Court Room- Sachchai Hazir Ho	Drama	4883
Star Plus	Dil Toh Happy Hai Ji	Soap opera	7740

TABLE 1. Summary of the data set.

Methodology



Results

I. Visualizing Shows Popularity of Discrete Channel

The graph shown below in figure 1 depicts the top watched show of STAR Plus channel. The parameter that is reserved for predicting the popularity of the show is the View Counts. Here are the screenshots of graphs produced from the sample dataset.

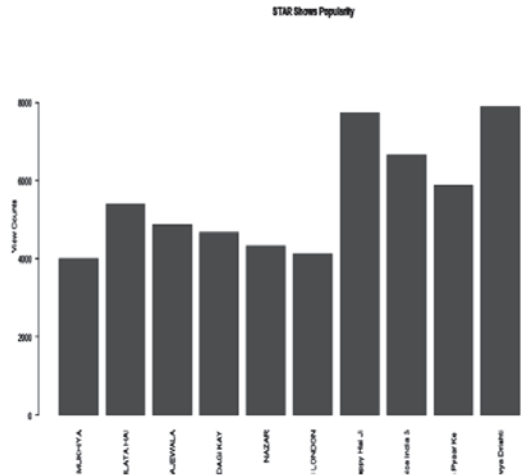


Figure 1. STAR Plus shows popularity

According to the above graph we can state that the show “Yeh Hai Mohabbatein” is the top viewed show among all the shows that are running on channel STAR Plus.

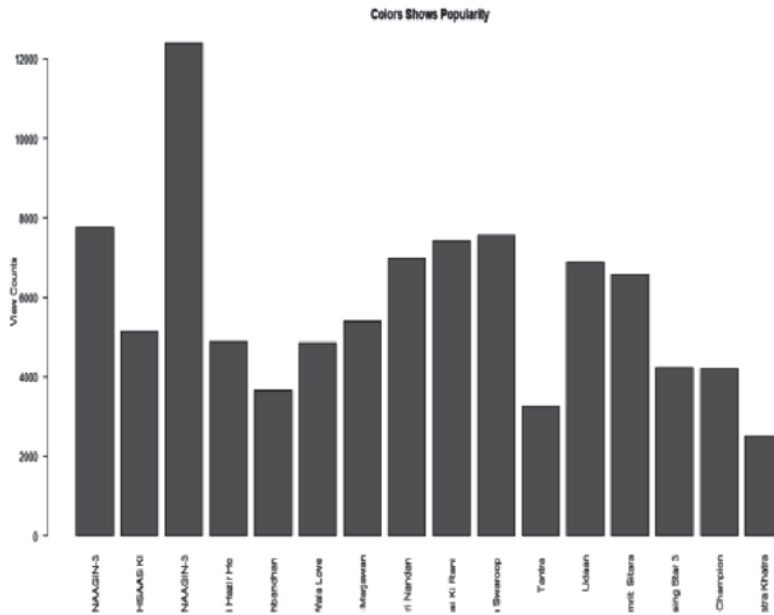


Figure 2. Colors shows popularity

There are numerous TV shows that are telecasted on channel Colors. Every show has its individual fascinating matter to larger budding audiences. From the plotted graph we can say that maximum of the individuals watch the show “Naagin 3”.

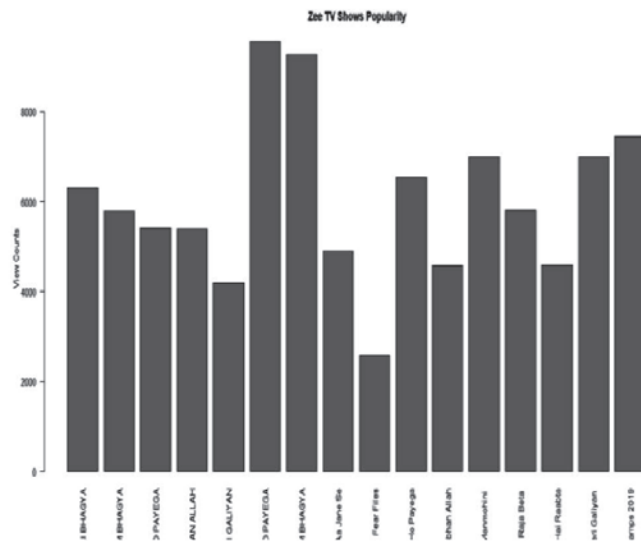


Figure 3. Zee TV shows popularity

Coming to Zee TV, that is a renowned channel in Indian Television because of its extended consecutively running shows. The top viewed show on this channel is “Tumse Na Ho Payega”. Running back “Kumkum Bhageya” can also be taken as an opponent of “Tumse Na Ho Payega”. There might be chances that “Kumkum Bhageya” will beat “Tumse Na Ho Payega” in the future.

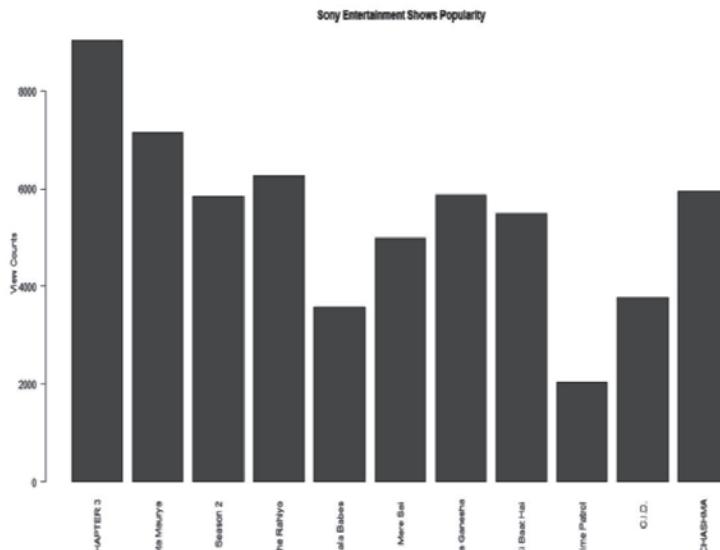


Figure 4. Sony Entertainment shows popularity

There are several shows that are telecasted on this channel but we can see that the show named “Super Dancer Chapter 3” stands first position in the popularity of this channel. “Chandragupta Maurya” is also standing sound and holds some level of fame which is a handy opponent of the show “Super Dancer Chapter 3”. The show “Crime Petrol” is not as popular as other shows of this channel.

I. Visualizing K-Means Clustering Algorithm

The graph below predicts the consequence of K-Means Clustering. This graph contains 3 clusters in which the first cluster depicts the less popular shows, second cluster illustrates the average shows and the third cluster illustrates the least popular shows.

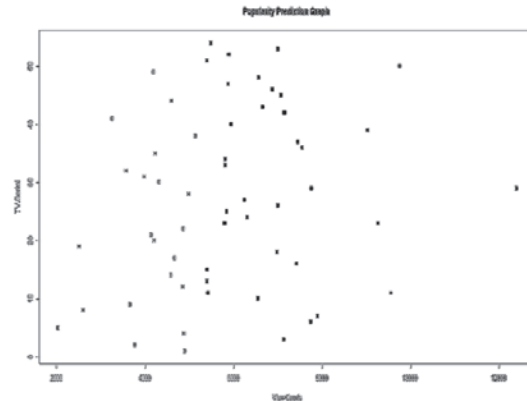


Figure 5. Clusters in K-Means Clustering Algorithm

As revealed in the graph, clusters are formed with correctness and are well distinguished from each other as cluster1 is symbolized with green colour, cluster 2 is symbolized with black colour and cluster 3 is symbolized with red colour.

These clusters are formed on the basis of view counts of the shows telecasted.

Clustermeans:	
View Count	
1	6515.500
2	4086.818
3	10002.800

Figure 6. Showing Means of Clusters

Figure 6 demonstrates the cluster means that are formed subsequently applying K-Means clustering.

These means are computed for each cluster on the basis of view counts of the shows. The data points whose remoteness is least from the given cluster is allocated to that precise cluster.

	1	2	3
Comedy	1	1	0
Crime Fiction	0	1	0
Drama	16	11	2
Horror	0	1	0
Mythological	1	0	0
Reality	2	3	1
Religious	1	0	0
Romantic Thriller	1	0	0
Sitcom	1	0	0
Soap Opera	4	2	1
Spiritual	0	1	0
Supernatural	3	2	1

Table2. Cluster Analysis on the Basis of Genres

Table 2 above shows clusters that contain various shows belonging to different genres.

Conclusion

We have applied k-means clustering by which three clusters are formed. According to these clusters the TV shows have been separated into three categories according to their popularity.

- Cluster 1: 30 shows (Less Popular Shows)
- Cluster 2: 22 shows (Average Shows)
- Cluster 3: 5 shows (Popular Shows)

Works Cited

- Simon Clippingdale, Makoto Okuda, Masaki Takahashi, Massahide Naemura and Mahito Fijii “Level-of-Interest Estimation for Personalized TV Program Recommendation”2013 IEEE International Conference on Consumer Electronics (ICCE)
 - Zhenzhensong, Zhiqiang Wei, Dongning Jia “The Algorithms of Intelligent TV Programs Recommendation Based on Data Mining”, IEEE, 2012 DOI 10.1109/CSSSS.2012.403
 - Mengyi Zhang, Minyong Shi, Zhiguo Hong, Songya Shang and Menghan Yan “A TV Program Recommendation System based on Big Data”,IEEE ICIS June 26-29, 2016, Okayama, Japan
 - Jiagshan Xu, Liang-jie Zhang, Haiming Lu and Yanda Li university P.R “The Development and Prospect of Personalized TV program Recommendation Systems”, IEEE,2002114351/02, 2002.
 - NikolaosPolatidis,ChristosK.Georgiadis “A Multi-level collaborative filtering method that improves recommendations” 0957-4174/© 2015 Elsevier Ltd. All rights reserved.
 - Shinjee Pyo, Eunhui Kim, and Munchurl Kim “LDA-Based Unified Topic Modeling for Similar TV User Grouping and TV Program Recommendation” IEEE VOL. 45, NO. 8, AUGUST 2015 2168-2267
 - Dorr Aimme (1986). Television And Children: A Special Medium For Special Audience, Sage, Thousand Oaks.4.
 - Fatima, Nabiha (2000). Effects of Satellite Channels (ZEE TV) on Middle Class of Lahore. M.A. Thesis, University of the Punjab, Lahore, 605.
 - Giacomo Corneo, (2002).Work and Television, CESifo Working Paper Series No. 829; IZA Discussion Paper No. 3766.
 - <https://www.rstudio.com/>
 - Alex Berson, Stephen J. Smith, Data Warehousing, Data Mining and OLAP, McGrawHill, 2004
 - D. Hand, H. Mannila, and P. Smyth, Principles of Data Mining, MIT Press, 2011
-