TV SHOW POPULARITY RATING USING KNN

A PROJECT REPORT

Submitted by

S.ARAVINTH
(Reg. No: 19BCR004)

M.KAVIN PRAKASH
(Reg. No: 19BCR029)

D.K.NAREN
(Reg. No: 19BCR033)

in partial fulfillment of the
requirements for the award of the
degree of

BACHELOR OF SCIENCE

IN

COMPUTER SYSTEMS AND DESIGN

DEPARTMENT OF COMPUTER TECHNOLOGY-UG

KONGU ENGINEERING COLLEGE

(Autonomous)

PERUNDURAI ERODE – 638 060

Estd : 1984

NOVEMBER 2021

DEPARTMENT OF COMPUTER TECHNOLOGY-UG

KONGU ENGINEERING COLLEGE

(Autonomous)

PERUNDURAI ERODE – 638060

NOVEMBER 2021

BONAFIDE CERTIFICATE

This is to certify that the mini project report entitled **TV SHOW POPULARITY RATING USING KNN** is the bonafide record of project work done by **S.ARAVINTH** (REG.NO:19BCR004)**, M.KAVIN PRAKASH** (REG.NO:19BCR029) and **D.K.NAREN** (REG.NO:19BCR033) in partial fulfillment for the requirement for the award of Degree of Bachelor of Science in **COMPUTER SYSTEMS AND DESIGN** of Anna University, Chennai during the year 2021-2022.

SUPERVISOR   **HEAD OF THE DEPARTMENT**

                                                                                      (Signature with seal)

Date:

Submitted for the end semester viva-voce examination held on_____

INTERNAL EXAMINER   **EXTERNAL EXAMINER**

DEPARTMENT OF COMPUTER TECHNOLOGY-UG

KONGU ENGINEERING COLLEGE

(Autonomous)

PERUNDURAI ERODE – 638060

NOVEMBER 2021

DECLARATION

We affirm that the Project Report titled **TV SHOW POPULARITY RATING USING KNN** being submitted in partial fulfillment for the award of **B.Sc. Degree in Computer Systems and Design** is the original work carried out by us. It has not formed the part of any other project submitted for award of any degree, either in this or any other University.

S.ARAVINTH (Reg.No:19BCR004)

M.KAVIN PRAKASH (Reg.No:19BCR029)

D.K.NAREN (Reg.No:19BCR033)

DATE:

I certify that the declaration made above by the candidates is true to the best of my knowledge.

DATE: **Signature of the Supervisor**

(Name & Designation)

ABSTRACT

Nowadays, there are a lot and a lot of social media platforms, one of the most wanted and watched is Disney+hotstar. The user's subscription also increasing day by day. Whereas reviews and ratings on TRP are viewed and reviewed by users on public media's. Media users want to relax their time with social media like Hulu, YouTube, Amazon prime, Netflix, and so on, one of the most famous applications among them is the Disney+hotstar streaming service. There are some category such as horror, popular, comedy. Finding Television Rating Point (TRP) for each channel is difficult. By using K-Nearest Neighbors (KNN) and Singular value decomposition (SVD) are used for finding the TRP rating.

In this work, K-Nearest Neighbor (KNN) and Singular Value Decomposition (SVD) presents an approach for mining online user reviews to find the channels Target Rating Point (TRP)

K-Nearest Neighbor (KNN) is the supervised machine learning algorithm for classification and regression. Where Singular Value Decomposition (SVD) is a solution for matching the matrix and finding the linear combination of data.

Since the TRP for each channel is varied according to the user review. The outcome of each channel (TRP) will be a set of reviews and the number of viewers list organized by Hotstar streaming services based on each feature. To know which channel has obtained the maximum target point.

ACKNOWLEDGEMENT

We express our sincere thanks to our beloved Correspondent **Thiru.P.SACHITHANANDAN** and philanthropic trust members of the Kongu Vellalar Institute of Technology Trust for having provided with necessary resources to complete this project.

We are always grateful to our beloved visionary Principal, **Dr.V.BALUSAMY, B.E.(Hons), M.Tech., Ph.D.,** and thank him for his motivation and moral support.

We express our deep sense of gratitude and profound thanks to **Dr.P.NATESAN M.E., Ph.D.,** Head of the Department, Computer Technology- UG for his invaluable commitment and guidance for this project.

We are in immense pleasure to express our hearty thanks to our beloved Project Coordinator **Dr.S.PARVATHAVARTHINI MCA., PhD.,** and our guide **Ms.N.KANIMOZHI M.E.,** for providing valuable guidance and constant support throughout the course of our project. We also thank the teaching, non-teaching staff members, fellow students and our parents who stood with us to complete our project successfully.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBRIVATIONS

KNN	K-NEAREST NEIGHBOUR
SVD	SINGULAR VALUE DECOMPOSITION
TRP	TELEVISION RATING POINT

CHAPTER-1

INTRODUCTION

1.1 OVERVIEW:

Every day, television shows are increased more in number whereas people's opinions about reality shows are expressed, whereas people share their experiences and thoughts regarding various channels, products and services on the World Wide Web. These are called opinions that are valuable in the decision-making process. Therefore, the World Wide Web has become a huge repository containing a different kind of opinions and thoughts of the people. However, to get benefits from these accumulated opinions, at first raw data's are taken based upon peoples opinion in the repository and view count

Sentiment Analysis is an ongoing field of research in the text mining field. The process of identifying and detecting subjective information using natural language processing, text classification. In short, the aim of sentiment analysis is to extract information on the attitude of the writer or speaker towards a specific channel, product or service.

Sentiment analysis is extremely useful in social media monitoring as it allows the user to gain an overview of the wider public opinion behind certain topics. Collecting customer reviews from social networks or any other web and apply the machine learning algorithms to classify the text to predict the polarity which is very useful for new customers.

Hotstar video streaming and broadband services connection help user around a global download and watch the large video files from the comfort of their homes. For

today's audiences, it's all about immediacy and mobility, the content they are looking for must be just a click away to fit their needs. Now everything is possible. Maybe you want to watch an episode of your favorite show when you are traveling, or maybe each member of your family wants to watch something different in a separate room of the house.

All of these demands are being fulfilled with the help of video streaming as well as the proliferation of devices that gives the user access to it. Now if you want to play movies, music or watch an episode of your favorite TV show you can easily do it wherever you may be. A whole season for a TV show can be found on the web and binged watched in just one weekend.

In the process of opinion mining whereas Hotstar contains more than thousands of channel, within this by selecting a particular channel and taking the view count to calculate the Television Rating Point(TRP), depend upon the number of view count and rate the channel or shows accordingly, like if the view count is more than ten thousand then allot a rating score of 10 points to the particular shows. if new data is present then add it in the middle of the process then the whole process will be starting again. Based on the number of views we will rate that the particular Television shows accordingly with the highest Rating (Highest TRP).

The TRP calculation can be compared among the different channel in the Hotstar streaming services with a certain algorithm, to find which channel show has obtained maximum target point

1.2 MACHINE LEARNING:

Machine Learning is a type of artificial intelligence which allows software applications to predict outcomes accurately. Machine Learning algorithms uses historical data as input to predict the output. Many leading companies like Google, Facebook, Amazon, Hotstar and Netflix using Machine Learning to competitive with other companies.They are classified into Supervised Learning, Unsupervised Learning and Reinforcement Learning are known as three important types of Machine Learning Techniques.
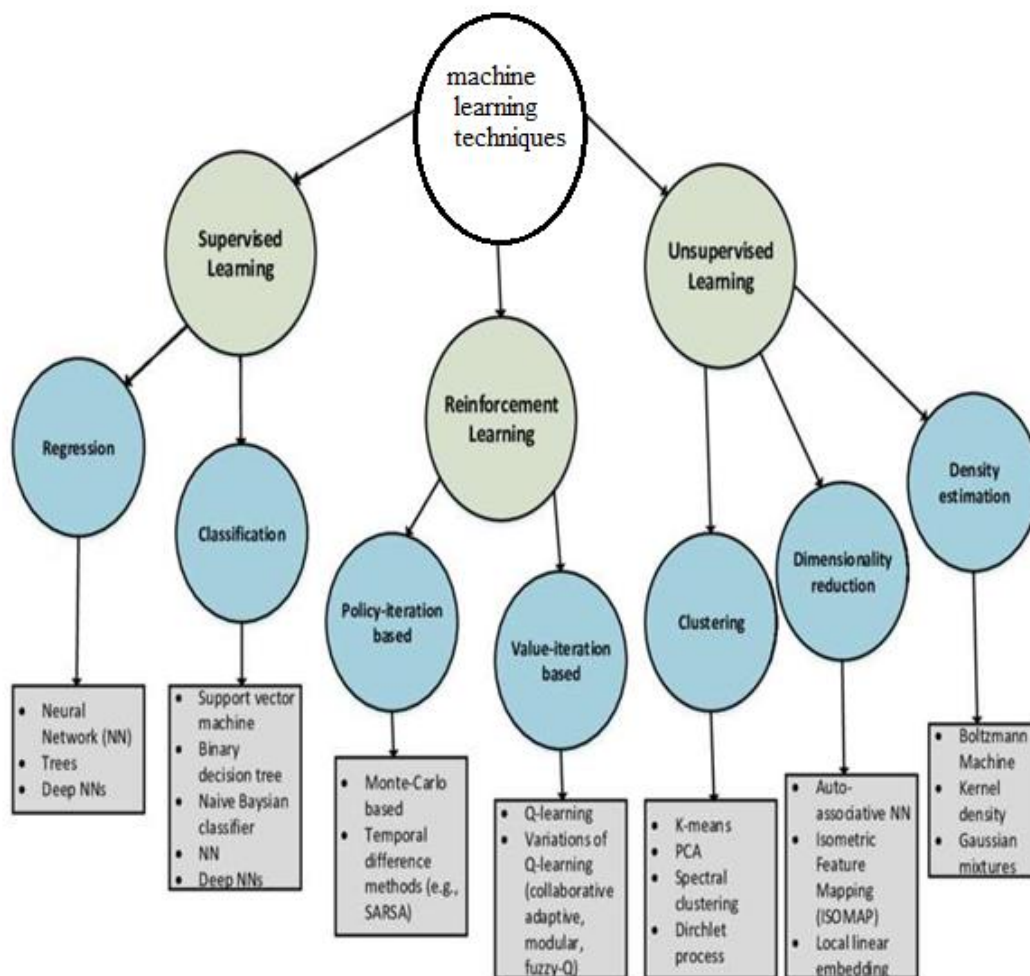
FIGURE 1.1 CLASSIFICATION OF MACHNE LEARNING

1.2.1 SUPERVISED LEARNING:

Supervised Learning is the most popular paradigm for performing machine learning operations. It is widely used for data where there is a precise mapping between input-output data. The dataset, in this case, is labeled, meaning that the algorithm identifies the features explicitly and carries out predictions or classification accordingly. As the training period progresses, the algorithm is able to identify the relationships between the two variables such that we can predict a new outcome. Resulting Supervised learning algorithms are task-oriented. As we provide it with more and more examples, it is able to learn more properly so that it can undertake the task and yield us the output more accurately. Some of the algorithms that come under supervised learning are as follows: Linear regression, random forest, support vector machine, artificial intelligence

1.2.2 UNSUPERVISED LEARNING:

Unsupervised machine learning holds the advantage of being able to work with unlabeled data. This means that human labor is not required to make the dataset machine-readable, allowing much larger datasets to be worked on by the program. The model learns through observation and finds structures in the data. Once the model is given a dataset, it automatically finds patterns and relationships in the dataset by creating clusters in it. In supervised learning, the labels allow the algorithm to find the exact nature of the relationship between any two data points. However, unsupervised learning does not have labels to work off of, resulting in the creation of hidden structures. Relationships between data points are perceived by the algorithm in an abstract manner, with no input required from human beings. The creation of these hidden structures is what makes unsupervised learning algorithms versatile. Instead of a defined and set problem statement, unsupervised learning algorithms can adapt to the data by dynamically changing hidden structures.This offers more post-deployment development than supervised learning algorithms. What it cannot do is add labels to the cluster, like it cannot say this a group of apples or mangoes, but it will separate all the apples from mangoes. Suppose we presented images of apples, bananas and mangoes to the model, so what it does, based on some patterns and relationships it creates clusters and divides the dataset into those clusters. Now if a new data is fed to the model, it adds it to one of the created clusters. The example of unsupervised learning is k-mean clustering, principle component analysis, SVD, FP-growth etc.

1.2.3 REINFORCEMEN LEARNING:

Reinforcement learning directly takes inspiration from how human beings learn from data in their lives. It features an algorithm that improves upon itself and learns from new situations using a trial-and-error method. Favorable outputs are encouraged or 'reinforced', and non-favorable outputs are discouraged or 'punished'. Based on the psychological concept of conditioning, reinforcement learning works by putting the algorithm in a work environment with an interpreter and a reward system. In every iteration of the algorithm, the output result is given to the interpreter, which decides whether the outcome is favorable or not.

1.3 CLASSIFICATION OF ALGORITHM:

The nature of the data determines which classification algorithm will provide the best solution to a given problem. Machine Learning has number of algorithms we going to use specific algorithm namely K-Nearest Neighbour (KNN) and Singular value decomposition (SVD).

1.3.1 K NEAREST NEIGHBOURS:

K-Nearest Neighbour is an classification Algorithm and it is a simplest machine learning algorithm based on Supervised Learning Technique. KNN classifies the data with help of reference data. It solves both classification and regression problems. KNN stores the entire training dataset which it uses as it representation and it does not learn any model.
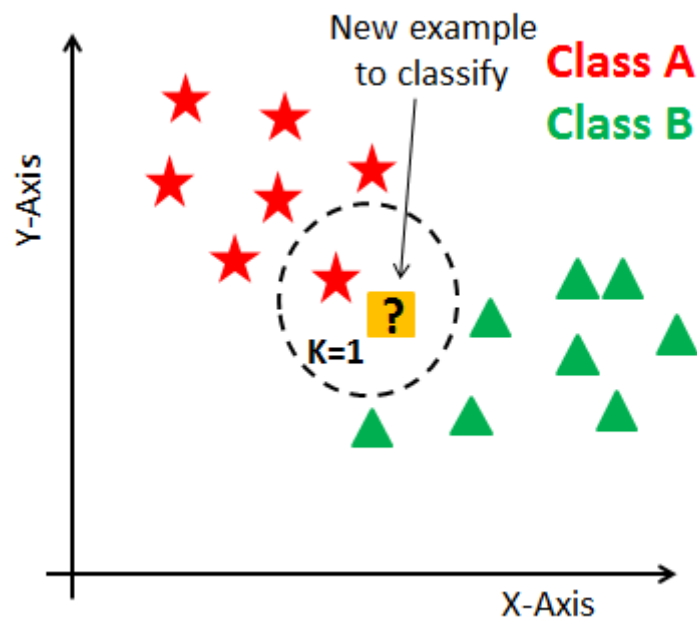
FIGURE 1.2 K-NEAREST NEIGHBOUR SPECIFICATION

1.3.2 SINGULAR VALUE DECOMPOSITION:

Singular value decomposition (SVD) provides another way to factorize a matrix, into singular vectors and singular values. Singular value decomposition has numerous applications in statistics, machine learning, and computer science. Where svd will be more stable than other algorithms.

Singular value decomposition (SVD) is otherwise known as matrix decomposition method for reducing a matrix into a constituent part in order to make certain subsequent matrix The SVD is used widely both in the calculation of other matrix operations, such as matrix inverse, but also as a data reduction method in machine learning. SVD can also be used in the least squares linear regression, image compression, and denoising data

CHAPTER-2

LITERATURE SURVEY

Dipak Gaikar, Riddhi Solanki, Harshada shinde, Pooja Phapale and Ishan Pandey [1] (2019) proposed a work on Movie Success Prediction using Popularity Factor from Social Media using Machine Learning. They have used K-nearest neighbors (KNN) Algorithm. This Algorithm is used for classification the Popularity factor of various movie parameters like actor, actress, director, writer, budget that is collected which helps in the movie success prediction. The objective of this paper is to find the movie success rating using social media. Their future enhancement is to obtain input from the audience which can be added to dataset to improve the result. Also extend and incorporate our project into mobile app.

Ajinkya Ambadkar, Rahul Jojare, Rushal Wankhade, Vinayak Gupta, Dr.P.K. Agrawal [2] (2020) proposed a work on TRP rating using Artifical Intelligence approach. They have used Apriori Algorithm. Rating of 1000 Netflix TV shows were used as dataset taken from Kaggle. They have performed Natural language processing in which text on a webpage changing its meaning into another language using NLP machine learning algorithm. Also performed Sentimental analysis which classifies and gives result in two types, that is Positive analysis and Negative analysis. Apriori Algorithm which classifies the Boolean association rules. Bottom up searching is the technique used apriori. This algorithm is used to operate on database containing transactions.

Rajeswari, Prasad and Kiran presented a paper An Advanced Neighbourhood approach of recommending movies on Netflix data by the combination of KNN and XGBoost [3](2020). In that Collaborative Filtering technique, recommendation, nearest neighbour and XGBoost were used by using the netflix dataset found that 99% of ratings

are not filled or missing because the number of movies watched by the user are typically low. By using neighbour model becomes the factor on the developed models KNN-XGBoost, the work was gradually constructed with various components of the model by formulations. On using KNN-XGBoost model can be applied to predict rating. The precision of the proposed system is approx. 82 % while using svd most accurate will be find was the future enhancement.

Sidhu, Attwal and r. Gaurav presented about To Evaluate & Predict the Television Serials" TRP [4] (2019). The number of TV reality shows are increasing day-by-day. firstly viewers' opinions are examined. By examining the viwer's opinion after that six factors have been constructed ie ..random tree, This random tree is a part of classification. an iterative learning revelation process like Data cleaning, Data coordination, Data selection, Data change and Data mining was made. Techniques used Classification: technique that assumes items in a collection to target categories or classes and Prediction: to identify data points purely on the description of another related data value. In these 6 major factors were considered so more factors can be added. WEKA tool was used for classification; In future other data mining tools can also be tried was the future enhancement.

Author name Saura Sambit Acharya, Ashvin Gupta and Prabu Shankar K.C [5] (2019) proposed a work on TV Show Popularity Analysis using social media, Data Mining.In this paper, author used K-nearest neighbors' algorithm (KNN),Support Vector Clustering (SVM).These Algorithm is used for Classification the Popularity factors of various types of movies like thriller, drama. That helps in the movie success prediction. The datasets used in this project to train machine learning models is obtained from IMDB site. The object of this paper is to find the popularity of the tv show using social media and data maining. Their future enhancement is to obtain input from the audience which can be added to the dataset to improve the result.

Tejaswi Kadam, Gaurav Saraf, Vikas Dewadkar and P.J Chate [6] (2019) proposed a work on Tv Show Popularity using Sentiment Analysis. The objective of this paper is to predict the performance of Tv show popularity of particular show and actors performance in that show using text reviews. Reviews are collected with help of social media like Twitter. They have used Porter Stemmer Algorithm . This algorithm is used in the process of removing the commoner morphological. They have performed Natural language processing in which text on a webpage changing its meaning into another language using NLP machine learning algorithm. Hash tagged dataset is used with tweets, to create hashtagged dataset duplicate tweet, non-English tweet and non-hashtag tweet are filtered. Sentimental Analysis is primary use case of NLP. Using sentimental analysis comments on social media are assessed by data scientists to see the performance of their show business to perform better.

Author name Krishnamoorthy, K.S Ramya and K.Pavithra [7] (2020) proposed a work on TV Show Popularity and Performance Predication. Using CNN Algorithm.In this paper, author used CNN Algorithm and K-nearest neighbours algorithm(KNN) with SVD.It used to identify the TRP rating and then find the well-liked shows with the help of very best rating movie. The datasets used in this paper is to get the information about popular shoes in IMDb.it is taken from IMDb page. The object of this paper is to find the TV Shows popularity and Performance Predication using CNN Algorithm.

Author name D. Anand, A.V. Satyavani, Raveena and M. Poojitha [8] (2018) proposed a work on Analysis and Prediction of Television Show Popularity Rating Using Incremental K-Mean Algorithm.In this paper, author used K-Mean algorithm and Incremental K-Means algorithm to compare the TRP. It make easy to predicting TV show analysis. The main concept of this paper is to provide a detailed and comprehensive description of important clustering algorithms. For TRP Predicting they have taken different TV programs telecasted like comedy ,drama ,reality ,education. Etc. Their future enhancement is to obtain input from the audience which can be added to the dataset to improve the result.

Author name Bane and Sheetlani proposed [9] a paper Success of Bollywood Movie Using Machine Techniques.used attribute namely actor and actress. Algorithm used are knn,svm.knn comes under the techniques of Supervised Learning.it gives accuracy of 84.1% for pre-released features and 89.27% for all features, divided a classes like hit, flop and neutral of the movies. The model developed for the accuracy of hit movies up to 80%.

Ramya Akula, Zachary Wieselthier, Laura Martin and Ivan Garibay [10] (2019) proposed a work on Forecasting the Success of Television Series using Machine Learning. In this paper, author used K-Nearest Neighbours(KNN).It make easy to predicting TV series with all the episodes. In this paper they extrapolate the data on title,directors,writers,original air date and ratings from Internet Movie Database[IMDb].Their future enhancement is to obtain input from the audience which can be added to the dataset to improve the result.

CHAPTER-3

SYSTEM SPECIFICATION

3.1 HARDWARE REQUIREMENTS

Processor: Intel dual core processor

RAM: 2 GB SD RAM

Hard disk: 500 GB

SOFTWARE REQUIREMENTS

Operating System: Windows 10 Professional

Simulation Tool: NetBeans IDE 8.1

Language: Java

3.2 TOOL DESCRIPTION:

3.2.1 JAVA

The software requirement specification is created at the end of the analysis task. The function and performance allocated to software as part of system engineering are developed by establishing a complete information report as functional representation, a representation of system behavior, an indication of performance requirements and design constraints, appropriate validation criteria.

FEATURES OF JAVA:

Java platform has two components:

> ➢ The Java Virtual Machine (Java VM)
>
> ➢ The Java Application Programming Interface (Java API)

The Java API is a large collection of ready-made software components that provide many useful capabilities, such as graphical user interface (GUI) widgets. The Java API is grouped into libraries (packages) of related components.

The following figure depicts a Java program, such as an application or applet, that's running on the Java platform. As the figure shows, the Java API and Virtual Machine insulate the Java program from hardware dependencies.
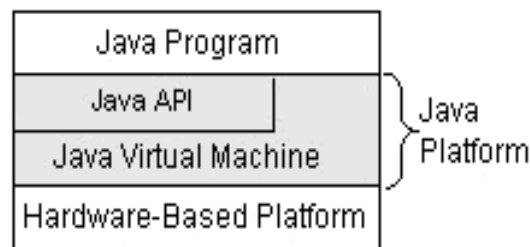
FIGURE 3.1 FEATURES OF JAVA

As a platform-independent environment, Java can be a bit slower than native code. However, smart compilers, well-tuned interpreters, and just-in-time byte code compilers can bring Java's performance close to that of native code without threatening portability.

3.2.2 NETBEANS IDE 8.1

NetBeans IDE lets you quickly and easily develop Java desktop, mobile, and web applications, as well as HTML5 applications with HTML, JavaScript, and CSS. The IDE also provides a great set of tools for PHP and C/C++ developers. It is free and open-source and has a large community of users and developers around the world.

NetBeans IDE is the official IDE for Java 8. With its editors, code analyzers, and converters, you can quickly and smoothly upgrade your applications to use new Java 8 language constructs, such as lambdas, functional operations, and method references.

Batch analyzers and converters are provided to search through multiple applications at the same time, matching patterns for conversion to new Java 8 language constructs.

With its constantly improving Java Editor, many rich features and an extensive range of tools, templates, and samples, NetBeans IDE sets the standard for developing with cutting edge technologies out of the box.

3.3 DATA SET DESCRIPTION:

The dataset used in this project is kaggle.com one of the popular website among the machine learning engineers. Kaggle may allow the user to search and publish data set and models build in web-based data-science environment. Our dataset is numeric datatype; it's based Hotstar streaming service dataset. It consist of 1000 of movie reviews with twofold names positive and negative with attribute as ratingLevel, user rating score, user rating size.

TABLE 3.1 DATA SET

Title	Rating	Rating Level	Rating Description	release year	user rating score	user rating size
White Chicks	PG-13	crude and sexual humor, language and some drug content	80	2004	82	80
Lucky Number Slevin	R	strong violence, sexual content and adult language	100	2006	NA	82

Grey's Anatomy	TV-14	Parents strongly cautioned. May be unsuitable for children ages 14.	90	2016	98	80
Prison Break	TV-14	Parents strongly cautioned. May be unsuitable for children ages 14 and under.	90	2008	98	80
How I Met Your Mother	TV-PG	Parental guidance. May not be suitable for all children.	70	2014	94	80
Supernatural	TV-14	Parents strongly cautioned. May be unsuitable for children ages 14 and under.	90	2016	95	80
Breaking Bad	TV-MA	For mature audiences. May not be suitable for children 17 and under.	110	2013	97	80
The Vampire Diaries	TV-14	Parents strongly cautioned. May be unsuitable for children ages 14 and under.	90	2017	91	80
The Walking Dead	TV-MA	For mature audiences. May not be suitable for children 17 and under.	110	2015	98	80

CHAPTER-4

METHODOLOGY

4.1 OBJECTIVE OF THE SYSTEM:

To find the trp for a channel, by calculating the number of users level, user rating and user rating size. Classifier algorithm KNN is used for classifying the dataset of Hotstar and to calculate and predict the Ranking TRP we use SVD

4.2 INTRODUCTION TO KNN:

K-Nearest Neighbour is an classification Algorithm and it is a simplest machine learning algorithm based on Supervised Learning Technique. KNN classifies the data with help of reference data. It solves both classification and regression problems. KNN stores the entire training dataset which it uses as it representation and it does not learn any model.

FIGURE 4.1 K NEAREST NEIGHBOUR

K-NEAREST NEIGHBORS:

Example of k-NN classification. The test sample (inside a circle) should be classified either to the first class of blue squares or to the second class of red triangles. If k = 3 (outside circle) it is assigned to the second class because there are 1 square and 2 triangles inside the inner circle. If, for example, k = 5 it is assigned to the first class.

KNN can be used for classification - the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors.

WORKING OF KNN:

We can understand its working with the help of following steps −

Step 1 − For implementing any algorithm, we need a dataset. So during the first step of KNN, we must load the training as well as test data.

Step 2 − Next, we need to choose the value of K i.e. the nearest data points. K can be any integer.

Step 3 − For each point in the test data do the following −

- Calculate the distance between test data and each row of training data with the help of any of the method namely: Euclidean, Manhattan or Hamming distance. The most commonly used method to calculate distance is Euclidean.

- Now, based on the distance value, sort them in ascending order.

- Next, it will choose the top K rows from the sorted array.

- Now, it will assign a class to the test point based on the most frequent class of these rows.

Step 4 − End

4.3 SINGULAR VALUE DECOMPOSITION (SVD):

Let A be an m $\times$ n matrix. It will be a decomposing a matrix into another matrix:

$$M=U\Sigma V^t$$

Property 1: A factorization $M=U\Sigma V^t$ is the Singular Value Decomposition for M, we consider this matrix of the value multiplying matrix **M**, which the **M** gets decomposed into other three different matrices:

$$M\ m \times n = U\ m \times m\ \Sigma\ m \times n\ Vt\ n \times n$$

Where U is an m $\times$ m orthogonal matrix, Σ is a pseudo diagonal matrix with non-negative elements and V is an n $\times$ n orthogonal matrix. The diagonal elements of the matrix Σ are called the singular value Decomposition of M.

Where

$$M=U\Sigma V \qquad\qquad 1$$

It will become

$$M^* = V\Sigma^*U^* \qquad\qquad 2$$

Because of U and V are real or complex unitary matrix their transpose will be their inverse, whereas Σ is a diagonal matrix so their transpose of the diagonal matrix will be the same when we multiple the 2nd equation with 1st equation we get:

$$R=M^* M= V\ \Sigma^*\ \Sigma V^* \qquad\qquad 3$$

Where in this equation 3 which implies Σ^2 =Eigen value of R, arrange the Eigenvalues in decreasing order because singular values are in decreasing order and find the Eigenvectors. To calculate SVD, we consider finding eigenvalues and eigenvectors, Eigenvalues and the eigenvectors are:

$$AA^T \text{ and } A^TA.$$

The eigenvectors of A^tA to be considered as columns of V, the eigenvectors of AA^t to be considered as columns of U. Also, the singular values in Σ are a square root of eigenvalues is from AA^t or A^tA. The singular values are the diagonal entries of the Σ matrix and are arranged in descending order. The singular values are always real numbers. If the matrix M is real, then U and V is also considered as real matrix.

Property 2: Matrix M is symmetric if and only if there exist a diagonal matrix D and an orthogonal matrix P.

$$M=PDP^t$$

Then, if the property is equal to the m×n matrix, then the formula A*A can be obtained a factorization according to property 1, and property 2

$$A^t A=PDP^t$$

If D is a diagonal matrix with entries that are computed from the eigenvalues of A^tA, then P is the eigenvector that corresponds to the eigenvalues of diagonal matrix D. According to property [1] if we give a matrix A then $M=U\Sigma V^t$ is the Singular Value Decomposition for M, where Σ is a pseudo diagonal matrix and U and V is an orthogonal matrix, Where $\sigma 1 > \sigma 2 > \sigma 3 > \ldots$ By arranging and normalizing it in decreasing order such that the value was in equal to 1

$$\mathbf{A}=\mathbf{U}\,\Sigma\,\mathbf{V}^{\mathrm{T}} \text{ and } \mathbf{A}^{\mathrm{T}}=\mathbf{V}\,\Sigma\,\mathbf{U}^{\mathrm{T}}$$

$$\mathbf{A}^{\mathrm{T}}\mathbf{A} = \mathbf{V}\,\Sigma\,\mathbf{U}^{\mathrm{T}}\mathbf{U}\,\Sigma\,\mathbf{V}^{\mathrm{T}}$$

$$\mathbf{A}^{\mathrm{T}}\mathbf{A} = \mathbf{V}\,\Sigma^{2}\mathbf{V}^{\mathrm{T}}$$

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{V} = \mathbf{V}\,\Sigma^{2}$$

Singular value decomposition brought a matrix on Σ are sorted from the largest to the smallest i.e of decreasing order then the best way to get the best possible result from the matrix A is by taking the first p rows and columns of the matrix Σ. Doing so eliminates the zero vector and also removes some small values.

CHAPTER 5

RESULT AND DISCUSSION

5.1 RESULT:

Research work on the HOTSTAR dataset for analyzing the Television Rating Point (TRP) by various attributes. After selecting the particular attributes of the HOTSTAR data set, which is applied to K Nearest Neighbour to classify the data then the TRP rate, is calculated by the Singular Value Decomposition (SVD) Algorithm, finally, the results are shown.

TABLE 5.1 TRP SCORE

TITLE	TRP
The 100	100.33333333333333
13 Reasons Why	96.33333333333333
Orange Is the New Black	96.0
Shameless (U.S.)	95.66666666666667
Ingobernable	95.33333333333333
Dexter	95.0
Dave Chappelle	94.66666666666667
Chewing Gum	94.33333333333333
House of Cards	93.66666666666667
Grace and Frankie	93.33333333333333
The Magicians	93.0
Kevin Hart: What Now?	92.33333333333333

5.2 DISCUSSION:

From the above table , television rating point for a Hotstar streaming service has been calculated. Hereby using two algoriths namely k-nearest neighbour(k-nn) and singular value decomposition(svd), k-nn is mainly used for grouping purpose and svd is for calculation purpose. First similar movie names are grouped together, then trp for each movie is calculated and last the movie name and trp value are separated from the group and arranged in which movie obtained higher trp.

CHAPTER 6

CONCLUSION AND FUTURE ENHANCEMENT

6.1 CONCLUSION:

Opinion mining is an important field of data mining that has been emerging with a boom. The proposed methodology is supported by natural language processing methods to Hotstar Streaming Service, then the features are classified to know the channels TRP(Television Target Point). Various features of the Hotstar streaming services are manipulated by users' rating scores by using the Singular Value Decomposition Algorithm to predict accurate results. The outcome of the system is a set of rating organized by the attributes of user rating and user rating description. This system helps to reduce the manual effort of evaluating reviews according to features in which the user is interested.

6.2 FUTURE ENHANCEMENT:

The opinions matter a lot while mining the sentiments from social media, any forums or websites and so on. The proposed system helps to give a result of Television Rating Point(TRP) and can find the users favorite channel. In the future, extend feature-based opinion mining focus on the Users emoji review. Also, I like to extend the work to find out the strength of various features which help to increase the sentiment emoji scores

APPENDICES

7.1 SAMPLE CODE

```java
package trp;

import java.io.File;

import java.io.FileInputStream;

import java.util.ArrayList;

import java.util.Arrays;

import java.util.List;

import java.util.HashMap;

import java.util.Map;

import java.util.TreeMap;

import java.util.Iterator;

import java.util.Set;

import java.util.Collections;

public class Trp {

    public static void main(String[] args) {

        // TODO code application logic here

        System.out.println("*************HOTSTAR*****************");
```

```java
String dataset="";


try

{

   File f =new File("hotstar.csv");

   FileInputStream fis=new FileInputStream(f);

   byte data[]=new byte[fis.available()];

   fis.read(data);

   fis.close();

   dataset=new String(data);

   System.out.println(dataset.trim());

   System.out.println();


   List<String> aa = new ArrayList();

}

catch(Exception e)

{

   e.printStackTrace();

}

 System.out.println("*********K-NearestNeighbour Algorithm*****************");

try
```

```java
{

    ArrayList title=new ArrayList();

    String st[]=dataset.trim().split("\n");

    for(int i=1;i<st.length;i++)

        {

            String row[]=st[i].trim().split("\t");

            if(!(title.contains(row[0].trim())))

            {

                title.add(row[0].trim());

            }

        }

    int k=0;

        for(int j=0;j<title.size();j++)

        {

            String t=title.get(j).toString().trim();


            int count=0;

            for(int i=1;i<st.length;i++)

            {


                String row[]=st[i].trim().split("\t");
```

```java
            if(t.trim().equals(row[0].trim()))

            {

                count++;

            }

        }

        if(j==0)

        {

            k=count;

        }

        else

        {

            if(count>k)

            {

                k=count;

            }

        }

    }

    for(int j=0;j<title.size();j++)

    {

        String t=title. get(j).toString().trim();

        for(int i=1;i<st.length;i++)
```

```
        {

            String row[]=st[i].trim().split("\t");

            if(t.trim().equals(row[0].trim()))

            {

                System.out.println(st[i].trim());

            }

        }

        System.out.println();

    }

System.out.println("*****************Singular Value Decomposition
Algorithm*****************");

System.out.println();

System.out.println(st[0].trim()+"\tTRP");

for(int j=0;j<title.size();j++)

    {

        String t=title.get(j).toString().trim();


        int viewCount=0;

        String svd="";

        double trp=0;

        for(int i=1;i<st.length;i++)
```

```java
            {

                String ro[]=st[i].trim().split("\t");

                if(t.trim().equals(ro[0].trim()))

                {

                    double
fortrp=(Double.parseDouble(ro[3].trim())+Double.parseDouble(ro[5].trim())+Double.parseDo
uble(ro[6].trim()))/(double)3;

                    trp=trp+fortrp;

                    viewCount++;

                    if(viewCount==1)

                    {

                        svd=st[i].trim();

                    }

                }

            }

            System.out.println();

            System.out.println(svd.trim()+"\t"+trp);

        }
    System.out.println("**************TRP FINDED ****************");

        System.out.println();

        HashMap<Double, String> rank = new HashMap<Double, String>();
```

```java
    for(int j=0;j<title.size();j++)

    {

        String t=title.get(j).toString().trim();

        int viewCount=0;

        String svd="";

        double trp=0;

        for(int i=1;i<st.length;i++)

        {

            String ro[]=st[i].trim().split("\t");

            if(t.trim().equals(ro[0].trim()))

            {

                double
fortrp=(Double.parseDouble(ro[3].trim())+Double.parseDouble(ro[5].trim())+Double.parseDo
uble(ro[6].trim()))/(double)3;

                trp=trp+fortrp;

                viewCount++;

                if(viewCount==1)

                {

                    svd=st[i].trim();

                    System.out.println(t+"\t"+fortrp);

                    rank.put(fortrp, t);
```

```
                }

              }

            }

        }

        System.out.println();

        System.out.println("*****************TRP RANKING FOR
CHANNEL*********************");

        System.out.println();

        Map<Double, String> map = new TreeMap<Double, String>(rank);

        Map<Double, String> reverseSortedMap = new TreeMap<Double,
String>(Collections.reverseOrder());

        reverseSortedMap.putAll(map);

        Set set = reverseSortedMap.entrySet();

        Iterator iterator = set.iterator();

        while(iterator.hasNext())

        {

          Map.Entry me = (Map.Entry)iterator.next();

          System.out.println(me.getValue()+ " : "+me.getKey());

        }

    }

    catch(Exception e)
```

```
    {

        e.printStackTrace();

    }

    }

}
```

7.2 SAMPLE OUTPUT

```
run:
*************HOTSTAR*******************
title   rating  rating_Level    rating_Description      release_Year    user_Rating_Score       user_Rating_Size
White Chicks    PG-13   crude and sexual humor, language and some drug content  80      2004    82      80
Lucky Number Slevin      R      strong violence, sexual content and adult language      100     2006    0       82
Grey's Anatomy  TV-14   Parents strongly cautioned. May be unsuitable for children ages 14 and under.  90      2016    98      80
Prison Break    TV-14   Parents strongly cautioned. May be unsuitable for children ages 14 and under.  90      2008    98      80
How I Met Your Mother   TV-PG   Parental guidance suggested. May not be suitable for all children.      70      2014    94      80
Supernatural    TV-14   Parents strongly cautioned. May be unsuitable for children ages 14 and under.  90      2016    95      80
Breaking Bad    TV-MA   For mature audiences.  May not be suitable for children 17 and under.   110     2013    97      80
The Vampire Diaries     TV-14   Parents strongly cautioned. May be unsuitable for children ages 14 and under.   90      2017    91      80
The Walking Dead        TV-MA   For mature audiences.  May not be suitable for children 17 and under.   110     2015    98      80
Pretty Little Liars     TV-14   Parents strongly cautioned. May be unsuitable for children ages 14 and under.   90      2016    96      80
Once Upon a Time        TV-PG   Parental guidance suggested. May not be suitable for all children.      70      2016    98      80
Sherlock        TV-14   Parents strongly cautioned. May be unsuitable for children ages 14 and under.  90      2016    95      80
Death Note      TV-14   Parents strongly cautioned. May be unsuitable for children ages 14 and under.  90      2006    77      80
Naruto  TV-PG   Parental guidance suggested. May not be suitable for all children.      70      2008    88      80
The Hunter      R       language and brief violence     100     2011    0       82
Lottie Dottie Chicken   TV-Y    Suitable for all ages.  10      2009    0       82
Arrow   TV-14   Parents strongly cautioned. May be unsuitable for children ages 14 and under.  90      2015    96      80
Black Mirror    TV-MA   For mature audiences.  May not be suitable for children 17 and under.   110     2016    80      80
The Originals   TV-14   Parents strongly cautioned. May be unsuitable for children ages 14 and under.  90      2016    74      80
The 100 NR      This movie has not been rated.  124     2016    97      80
Masha and the Bear      TV-Y    Suitable for all ages.  10      2013    81      80
Ã" Pai, Ã" - Look at This        TV-MA   For mature audiences.  May not be suitable for children 17 and under.   110     2007    0       82
Nymphomaniac: Volume 1  NR      This movie has not been rated.  124     2013    0       82
Hunter X Hunter (2011)  NR      This movie has not been rated.  124     2013    57      80
Marvel's Luke Cage      TV-MA   For mature audiences.  May not be suitable for children 17 and under.   110     2016    95      80
Marvel's Iron Fist      TV-MA           110     2017    98      80
Narcos  TV-MA   For mature audiences.  May not be suitable for children 17 and under.   110     2016    94      80
The Flash       TV-PG   Parental guidance suggested. May not be suitable for all children.      70      2016    98      80
The Seven Deadly Sins   TV-14   Parents strongly cautioned. May be unsuitable for children ages 14 and under.   90      2016    0       82
Scream  TV-14   Parents strongly cautioned. May be unsuitable for children ages 14 and under.  90      2016    0       82
The Do-Over     TV-MA   For mature audiences.  May not be suitable for children 17 and under.   110     2016    84      80
```

FIGURE 7.1 DATASET CLASSIFICATION

```
*********K-Neeearest Neighbour Algorithm*****************
White Chicks      PG-13    crude and sexual humor, language and some drug content  80         2004     82       80

Lucky Number Slevin    R        strong violence, sexual content and adult language        100       2006    0        82
Lucky Number Slevin    R        strong violence, sexual content and adult language        100       2006    0        82
Lucky Number Slevin    R        strong violence, sexual content and adult language        100       2006    0        82

Grey's Anatomy   TV-14    Parents strongly cautioned. May be unsuitable for children ages 14 and under.    90       2016     98       80
Grey's Anatomy   TV-14    Parents strongly cautioned. May be unsuitable for children ages 14 and under.    90       2016     98       80
Grey's Anatomy   TV-14    Parents strongly cautioned. May be unsuitable for children ages 14 and under.    90       2016     98       80
Grey's Anatomy   TV-14    Parents strongly cautioned. May be unsuitable for children ages 14 and under.    90       2016     98       80
Grey's Anatomy   TV-14    Parents strongly cautioned. May be unsuitable for children ages 14 and under.    90       2016     98       80
Grey's Anatomy   TV-14    Parents strongly cautioned. May be unsuitable for children ages 14 and under.    90       2016     98       80

Prison Break     TV-14    Parents strongly cautioned. May be unsuitable for children ages 14 and under.    90       2008     98       80
Prison Break     TV-14    Parents strongly cautioned. May be unsuitable for children ages 14 and under.    90       2008     98       80
Prison Break     TV-14    Parents strongly cautioned. May be unsuitable for children ages 14 and under.    90       2008     98       80
Prison Break     TV-14    Parents strongly cautioned. May be unsuitable for children ages 14 and under.    90       2008     98       80
Prison Break     TV-14    Parents strongly cautioned. May be unsuitable for children ages 14 and under.    90       2008     98       80
Prison Break     TV-14    Parents strongly cautioned. May be unsuitable for children ages 14 and under.    90       2008     98       80
Prison Break     TV-14    Parents strongly cautioned. May be unsuitable for children ages 14 and under.    90       2008     98       80

How I Met Your Mother   TV-PG    Parental guidance suggested. May not be suitable for all children.    70       2014     94       80
How I Met Your Mother   TV-PG    Parental guidance suggested. May not be suitable for all children.    70       2014     94       80
How I Met Your Mother   TV-PG    Parental guidance suggested. May not be suitable for all children.    70       2014     94       80
How I Met Your Mother   TV-PG    Parental guidance suggested. May not be suitable for all children.    70       2014     94       80
How I Met Your Mother   TV-PG    Parental guidance suggested. May not be suitable for all children.    70       2014     94       80

Supernatural     TV-14    Parents strongly cautioned. May be unsuitable for children ages 14 and under.    90       2016     95       80
Supernatural     TV-14    Parents strongly cautioned. May be unsuitable for children ages 14 and under.    90       2016     95       80
Supernatural     TV-14    Parents strongly cautioned. May be unsuitable for children ages 14 and under.    90       2016     95       80

Breaking Bad     TV-MA    For mature audiences.  May not be suitable for children 17 and under.    110      2013     97       80
Breaking Bad     TV-MA    For mature audiences.  May not be suitable for children 17 and under.    110      2013     97       80
```
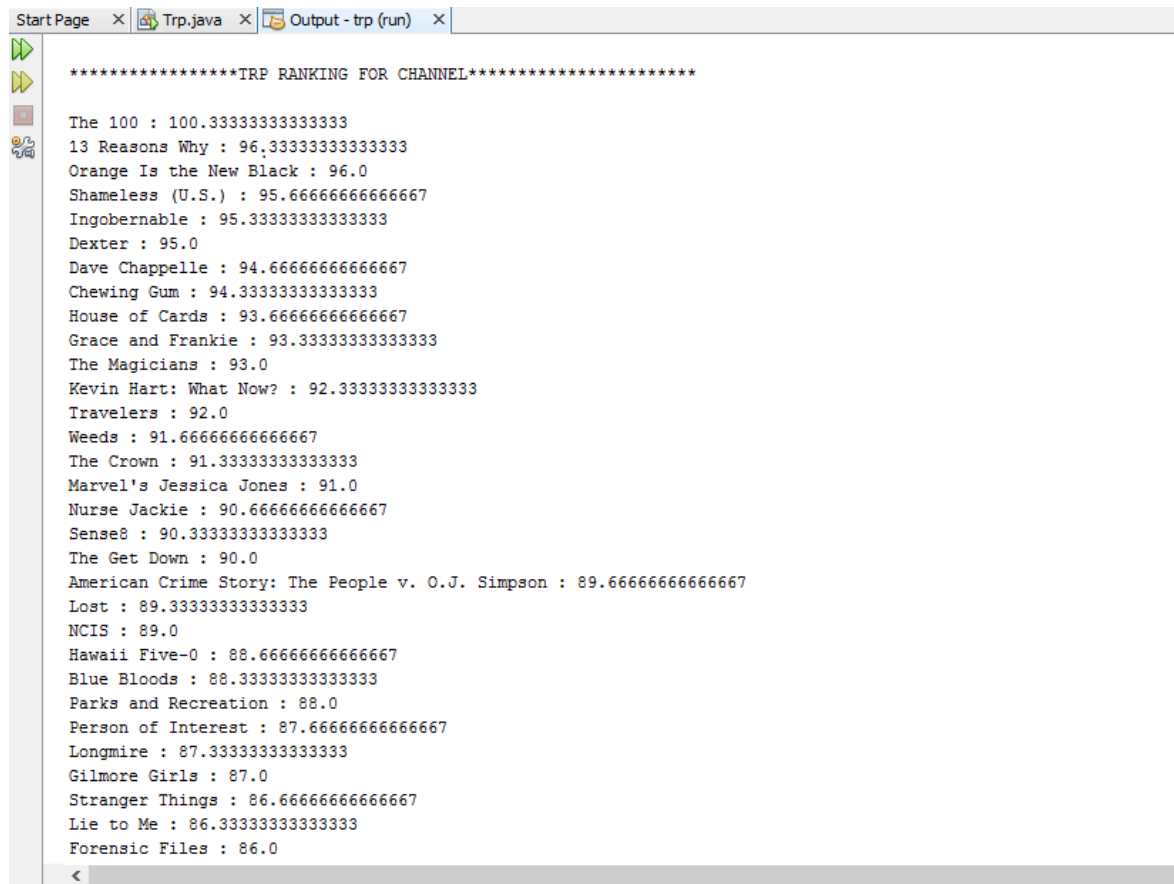
FIGURE 7.2 KNN CLASSIFICATION

```
******************Singular Value Decomposition Algorithm******************

title    rating   rating_Level    rating_Description        release_Year    user_Rating_Score      user_Rating_Size      TRP

White Chicks    PG-13    crude and sexual humor, language and some drug content  80        2004     82       80      80.66666666666667

Lucky Number Slevin     R        strong violence, sexual content and adult language        100       2006    0        82      182.0

Grey's Anatomy   TV-14    Parents strongly cautioned. May be unsuitable for children ages 14 and under.    90       2016     98       80      536.0

Prison Break    TV-14    Parents strongly cautioned. May be unsuitable for children ages 14 and under.    90       2008     98       80      625.3333333333334

How I Met Your Mother   TV-PG    Parental guidance suggested. May not be suitable for all children.    70       2014     94       80      406.66666666666663

Supernatural    TV-14    Parents strongly cautioned. May be unsuitable for children ages 14 and under.    90       2016     95       80      265.0

Breaking Bad    TV-MA    For mature audiences.  May not be suitable for children 17 and under.    110      2013     97       80      287.0

The Vampire Diaries     TV-14    Parents strongly cautioned. May be unsuitable for children ages 14 and under.    90       2017     91       80      522.0

The Walking Dead    TV-MA    For mature audiences.  May not be suitable for children 17 and under.    110      2015     98       80      192.0

Pretty Little Liars     TV-14    Parents strongly cautioned. May be unsuitable for children ages 14 and under.    90       2016     96       80      532.0

Once Upon a Time    TV-PG    Parental guidance suggested. May not be suitable for all children.    70       2016     98       80      248.0

Sherlock    TV-14    Parents strongly cautioned. May be unsuitable for children ages 14 and under.    90       2016     95       80      265.0

Death Note    TV-14    Parents strongly cautioned. May be unsuitable for children ages 14 and under.    90       2006     77       80      82.33333333333333

Naruto  TV-PG    Parental guidance suggested. May not be suitable for all children.    70       2008     88       80      79.33333333333333

The Hunter    R        language and brief violence    100       2011     0        82      60.666666666666664
```

FIGURE 7.3 CALCULATION USING SVD

```
Start Page  ×  Trp.java  ×  Output - trp (run)  ×

******************TRP RANKING FOR CHANNEL***********************

The 100 : 100.33333333333333
13 Reasons Why : 96.33333333333333
Orange Is the New Black : 96.0
Shameless (U.S.) : 95.66666666666667
Ingobernable : 95.33333333333333
Dexter : 95.0
Dave Chappelle : 94.66666666666667
Chewing Gum : 94.33333333333333
House of Cards : 93.66666666666667
Grace and Frankie : 93.33333333333333
The Magicians : 93.0
Kevin Hart: What Now? : 92.33333333333333
Travelers : 92.0
Weeds : 91.66666666666667
The Crown : 91.33333333333333
Marvel's Jessica Jones : 91.0
Nurse Jackie : 90.66666666666667
Sense8 : 90.33333333333333
The Get Down : 90.0
American Crime Story: The People v. O.J. Simpson : 89.66666666666667
Lost : 89.33333333333333
NCIS : 89.0
Hawaii Five-0 : 88.66666666666667
Blue Bloods : 88.33333333333333
Parks and Recreation : 88.0
Person of Interest : 87.66666666666667
Longmire : 87.33333333333333
Gilmore Girls : 87.0
Stranger Things : 86.66666666666667
Lie to Me : 86.33333333333333
Forensic Files : 86.0
<
```

FIGURE 7.4 TRP

REFERENCE

1. Dipak Gaikar, Riddhi Solanki. "Movie Success Prediction using Popularity Factor from Social Media using Machine Learning"[2019].

2. Ajinkya Ambadkar, Rahul Jojare. "TRP rating using Artifical Intelligence approach"[2020].

3. Rajeswari, Prasad. "Advanced Neighbourhood approach of recommending movies on Netflix data by the combination of KNN and XGBoost"[2020].

4. Sidhu, Attwal. "To Evaluate & Predict the Television Serials TRP"[2019].

5. Saura Sambit Acharya,Ashvin Gupta. "TV Show Popularity Analysis using Social Media,Data Mining"[2019].

6. Tejaswi Kadam, Gaurav Saraf. "Tv Show Popularity using Sentiment Analysis"[2016].

7. N.Krishnamoorthy,K.S Ramya. "TV Show Popularity and Performance Predication. Using CNN Algorithm"[2020].

8. D.Anand,A.V.Satyavani. "Analysis And Prediction Of Television Show Popularity Rating Using Incremental K-Mean Algorithm"[2018].

9. Bane and Sheetlani. "Success of Bollywood Movie Using Machine Techniques"[2019].

10. Ramya Akula, Zachary Wieselthier. "Forecasting the Success of Television Series using Machine Learning"[2019].