# TV SHOW POPULARITY ANALYSIS USING DATA MINING TECHNIQUES

1st Rachi Ashok Bandewar 2nd Prof Pushpa Tandekar 3rd Prof Ashish Deharkar

|  |  |  |
| --- | --- | --- |
| rachibandewar1@gmail.com | p.tandekar@yahoo.in | ashish.deharkar@gmail.com |
| Department of Computer Science | Department of Computer Science | Department of Computer Science |
| And Engineering | And Engineering | And Engineering |
| Shri Sai College of Engineering and | Shri Sai College of Engineering and | Shri Sai College of Engineering and |
| Technology,Chandrapur ,India | Technology,Chandrapur ,India | Technology,Chandrapur,India |

**ABSTRACT**: The television industry is a constantly evolving billion dollar industry. With online streaming services such as Netflix and Amazon Prime, people have access to thousands of TV shows. The rating and reviews that the audience provides is the biggest indication of whether the show is successful or not. With such data available, we can find out what features the most successful shows have in common and the shows of which genre are likely to be more successful with the help of various Machine Learning techniques such as classification and clustering. Algorithms such as k-NN, SVM, Naive Bayes, Decision Trees and Gradient Descent can be employed to build a model with high accuracy Based on the predictions made by the model we can also make favorable recommendations to different demographics based on their interests.

# Keywords— TV show popularity, Sentiment Analysis, Machine Learning

**INTRODUCTION**: In today's world internet has become an important part of our day-to-day life. Social TV is a novel paradigm that has received much attention in the last decade, researches have brought us new technologies to support interaction among users. TV entertainment has gone way beyond traditional broadcasts. The purpose of this sentiment analysis is to improve recommendations Video on demand by exploring social interactions among users. In TV show recommendation shows are recommended to viewers based on sentiment analysis. Users are provided to add ratings and reviews on the show of their interests. In this Social application, users can send friend requests, get friend requests and also view online friends and can also share shows with their friends. Easy predictions of trending show based on viewers' reviews or comments. Positive and negative comments can be obtained easily by applying some keyword method.

Data importing and presenting into a graph for better understanding. By performing sentiment analysis on review and apply emotions related to post. Emoji using data mining technique also users can check other users emotion-related show. To evaluate the performance of TV shows graphical results are presented in bar graphs to view overall results. To attract major audiences many efforts are to be taken into considerations. Different episodes are to be released on different days or weekends to know the popularity check. Viewers can also search TV shows of different categories. Review on particular shows will be extracted and comments are put together from different sources and social networks. Based on comment and the popularity of the TV show rating is given accordingly. Admin can add or remove new TV shows with category, show name, and images.

Supervised machine learning algorithm

At its most basic sense, machine learning uses programmed algorithms that learn and optimise their operations by analysing input data to make predictions within an acceptable range.. These three categories are: supervised, unsupervised and semi-supervised.

In supervised machine learning algorithms, a labelled training dataset is used first to train the underlying algorithm. This trained algorithm is then fed on the unlabelled test dataset to categorise them into similar groups.

Logistic regression

Logistic regression (LR) is a powerful and well-established method for supervised classification . LR helps in finding the probability that a new instance belongs to a certain class. Since it is a probability, the outcome lies between 0 and 1. Therefore, to use the LR as a binary classifier, a threshold needs to be assigned to differentiate two classes.

Support vector machine

Support vector machine (SVM) algorithm can classify both linear and non-linear data. It first maps each data item into an n-dimensional feature space where n is the number of features. It then identifies the hyperplane that separates the data items into two classes while maximising the marginal distance for both classes and minimising the classification errors

Decision tree

Decision tree (DT) is one of the earliest and prominent machine learning algorithms. A decision tree models the decision logics i.e., tests and corresponds outcomes for classifying data items into a tree-like structure.

Random forest

A random forest (RF) is an ensemble classifier and consisting of many DTs similar to the way a forest is a collection of many trees.

Naïve Bayes

Naïve Bayes (NB) is a classification technique based on the Bayes’ theorem This theorem can describe the probability of an event based on the prior knowledge of conditions related to that event.

K-nearest neighbour

The K-nearest neighbour (KNN) algorithm is one of the simplest and earliest classification algorithms It can be thought a simpler version of an NB classifier. Unlike the NB technique, the KNN algorithm does not require to consider probability values.

Artificial neural network

Artificial neural networks (ANNs) are a set of machine learning algorithms which are inspired by the functioning of the neural networks of human brain.. Likewise, ANN algorithms can be represented as an interconnected group of nodes. The output of one node goes as input to another node for subsequent processing according to the interconnection.

Stochastic Gradient Descent—it is used to build a predictive models. It is used to find optimal solution to a linear regression problem. It involves “loss function”, ”weak learner”, ”additive model”. F1 score is a formula to compute the score of precision and recall the higher the f1 score is the better prediction will be. Precision tells us what fraction of your outcome is relevant And recall tells us the fraction of total relevant results correctly predicted by your model. Stemming is a process in which different forms of word are converted to their root word

# IMPLEMENTATION:

Dataset: The data is collected and this study of diseases and their corresponding symptoms is available on Kaggle.

Training Data: Training data is also known as training datasets, training sets, and training sets. It is an important aspect of the machine learning model which helps us to make accurate predictions and perform the tasks we want. Simply put, training data forms a machine learning model and tells you what the awaited result looks like. The model iteratively analyzes the dataset to understand its attributes precisely and make appropriate changes to enhance the performance.

Testing Data: The test dataset is a subset of the training dataset used to make an objective evaluation of the final model.

Balanced Data The observation of the dataset and its visualization leads us to the conclusion that the data is balanced and there’s no imbalance in the data, which means that training and testing will give real exactness.



**Fig No.1 Implementation**

# DATA MINING PROCESS

The data mining process involves a number of steps from data collection to visualization to extract valuable information from large data sets. As mentioned above, data mining techniques are used to generate descriptions and predictions about a target data set. Data scientists describe data through their observations of patterns, associations, and correlations. They also classify and cluster data through classification and regression methods, and identify outliers for use cases, like spam detection.

Data mining usually consists of four main steps: setting objectives, data gathering and preparation, applying data mining algorithms, and evaluating results.

1. **Set the business objectives:** This can be the hardest part of the data mining process, and many organizations spend too little time on this important step. Data scientists and business stakeholders need to work together to define the business problem, which helps inform the data questions and parameters for a given project. Analysts may also need to do additional research to understand the business context appropriately.
2. **Data preparation:** Once the scope of the problem is defined, it is easier for data scientists to identify which set of data will help answer the pertinent questions to the business. Once they collect the relevant data, the data will be cleaned, removing any noise, such as duplicates, missing values, and outliers. Depending on the dataset, an additional step may be taken to reduce the number of dimensions as too many features can slow down any subsequent computation. Data scientists will look to retain the most important predictors to ensure optimal accuracy within any models.
3. **Model building and pattern mining:** Depending on the type of analysis, data scientists may investigate any interesting data relationships, such as sequential patterns, association rules, or correlations. While high frequency patterns have broader applications, sometimes the deviations in the data can be more interesting, highlighting areas of potential fraud.

# DATA MINING PROCESS:



**Fig No.2 Data Mining Process**

# Data Mining Process are as follows:

1. **Business Understanding:** In this step, the goals of the businesses are set and the important factors that will help in achieving the goal are discovered.
2. **Data Understanding:** This step will collect the whole data and populate the data in the tool (if using any tool). The data is listed with its data source, location, how it is acquired and if any issue encountered. Data is visualized and queried to check its completeness.
3. **Data Preparation:** This step involves selecting the appropriate data, cleaning, constructing attributes from data, integrating data from multiple databases.
4. **Modeling:** Selection of the data mining technique such as decision-tree, generate test design for evaluating the selected model, building models from the dataset and assessing the built model with experts to discuss the result is done in this step.
5. **Evaluation:** This step will determine the degree to which the resulting model meets the business requirements. Evaluation can be done by testing the model on real applications. The model is reviewed for any mistakes or steps that should be repeated.

# MODULES:

The system comprises of 2 major modules with sub-modules as follows:

**Admin**

**Login:** Admin need to login into the system by inputting the login credentials.

**Add Pages:** Admin will add page details such as page name and page link.

**Add Entry:** Admin can add entry for a page by selecting page name and providing other details.

**View Graph:** Admin can generate 5 graph (Pie Chart & Bar Chart) based on Age, Gender, Location, Comment's Sentiment based on peoples review on each TV Show. Admin will provide the data in excel format and if that data does not have any field then admin will enter it randomly.

**Print Graph:** Can print all the 4 graphs which is generated by the system.

**Visitor:**

**View Graph:** Visitor can view 5 graph (Pie Chart & Bar Chart) based on Age, Gender, Location, Comment's Sentiment based on peoples review on each TV Show. Admin will provide the data in excel format and if that data does not have any field then admin will enter it randomly. Visitor can also view the popularity of every show.

**Print Graph:** Can print all the 4 graphs which is generated by the system

# ADVANTAGES

* More efficient.
* View TV Shows Based on sentiment analysis.
* Sentimental comment analyzing and predicting good or bad comments.
* Easy prediction of TV Show trending based on people rating.
* Graphical representation of TV Show popularity.
* Sorting of graphs by Age, Gender, Location and Good or Bad comments based on peoples reviews or comments.
* Easy importing of data and exporting it into graph

. • Graphical data in printable format.

* The visitor will get to know the show's popularity.

# CONCLUSION :

In this paper we have presented a predictive model to predict the popularity of tv shows based on user comments from social media. We are able to obtain significant results over the provided datasets. The model uses sentiments of the viewers and can be implemented for any genre of tv show. Results are highly accurate based on the values we’ve obtained using data mining and machine learning.

# REFERENCES:

1. Rashmi Singh, Srushti Nemade, Atharv Pillai, Binoy Vijaykumar, and Prof. Gayatri Hegde, “TV Show Popularity Analysis “,International Journal of Creative Research and Thoughts Volume 9
2. Ruchika Honawale ,Neha Moyal, Shital Talekar, Vaishnavi Sadar Prof. Sunita Patil, “SURVEY PAPER ON TV SHOW POPULARITY PREDICTION USING SENTIMENT ANALYSIS IN SOCIAL NETWORK”, © 2021 JETIR June 2021, Volume 8, Issue 6
3. D. Anand,”ANALYSIS AND PREDICTION OF TELEVISIONSHOW POPULARITY RATING USING INCREMENTAL K-MEANS ALGORITHM”, International Journal of Mechanical Engineering and Technology (IJMET) Volume 9, Issue 1, January 2018,
4. Saura Sambit Acharya, Ashvin Gupta, Prabu Shankar K.C ,”TV Show Popularity Analysis using Social Media, Data Mining”,International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8, Issue-7S, May 2019
5. Ms. Medha Pradeep Phanshikar, Mr.D.A.Patil ,”Research Paper on TV Show Popularity Analysis,” International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 08 Issue: 12 | Dec 2021
6. Mr. Shivam Jog “Prediction and Sentimental Analysis of TRP rating Using Artificial Intelligence Approach”, International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET) | e-ISSN: 2319-8753, p-ISSN: 2320-6710| Impact Factor: 7.089|

||Volume 9, Issue 4, April 2020

1. Mr. Bhargav Siras, Mr. Akash Fender, Mr. Parikshit Mandurkar, Mr. Yash Nikalje, Dr. Sharda Chhabria, “VIDEO POPULARITY PREDICTION USING MACHINE LEARNING

“International Research Journal of Modernization in Engineering Technology and Science Volume :03

1. Ayushi Chahal, Preeti Gulia,” Machine Learning and Deep Learning”, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075 (Online),

Volume-8 Issue-12, October 2019

1. Anjali Jagwani,”A REVIEW OF MACHINE LEARNING IN EDUCATION”, JETIR May 2019, Volume 6, Issue 5
2. S.A.R. NIHA,” STUDY OF DATA MINING METHODS AND ITS APPLICATIONS” International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 11

| Nov -2017