

# Big Data Analytics for Program Popularity Prediction in Broadcast TV Industries.

1<sup>st</sup> Nikhil Sunil Kapadne, 2<sup>nd</sup> Ajay Mansing Jadhav 3<sup>rd</sup> Nikhil Mahendra Jadhav , 4<sup>th</sup> Nikhil Mahendra Jadhav  
Pillai Hoc College of Engineering and Technology

**Abstract:** TV Show Popularity prediction using sentiment analysis is one of the most interesting and challenging tasks. A critical demand along this line is to predict the popularity of online serials, which can enable a wide range of applications, such as online advertising, and serial recommendation. The problem motivation stated above suggest is that it is only the viewer of a program who is responsible for its popularity or failure and if we anyhow can identify the most common features of a program which, the viewers want most, and through some effective scientific methodology could insert these requirements in the proposed TV program well at the time of production. The purpose of this work is to evaluate the performance of TV Show and also calculate how many people are liked to a particular show or actors of that show and predicting Popularity of that shows, based on the text reviews. We are getting reviews on social networking websites like Twitter.

**IndexTerms:** Broadcast TV, Popularity Prediction, Dynamic time wrapping, Random forests regression Gradient boosting decision tree

Mining the existing data within social media applications is promising to produce the required information to meet the relevant objective. However, social media data are vast, noisy, unstructured, and dynamic in nature, and thus mining the data is not easy to do. In this report we focus on the popular social networking service, Twitter. Through a case study we explore whether data from Twitter could benefit by giving it business intelligence for products.

## 1.2 Relevance

Social media is a platform where people create content, share their opinions, views and ideas. Since social media can also be understood as a form of collective wisdom, it can be used to predict real-world outcomes. Twitter is the place where we all gather to express particular points of view and feelings about specific topics. Twitter has millions of users that spread millions of personal posts on a daily basis. And this gives us the opportunity to study social human subjectivity.

## 1.3 Organization of Project Report

The proposed project report is structured in following chapters. We discuss the existing system and problem present in pervious system in Chapter 2<sup>nd</sup>. In Chapter 3<sup>rd</sup>, we illustrate the Requirement Gathering which will be included both Software Requirement and hardware Requirement. In next Chapter 4<sup>th</sup>, we will be discussing plan of the project. Plan of the project with Gantt chart and system architecture. In Chapter 5<sup>th</sup> Project analysis will be done using various UML diagrams. In Chapter 6<sup>th</sup> we will be discussing Project Design by including Class diagram, DFD, Component diagram etc. Finally, we conclude the paper in Chapter 8<sup>th</sup> with the future scope of proposed project

## I. INTRODUCTION

### 1.1 Background

Social networking sites nowadays are contributing a lot towards data. In order to find the interesting patterns or trends from this huge data, data scientists need to clean, integrate, aggregate and analyze the data. Twitter is a platform where users post short messages (called tweets) to their followers and read tweets of people whom they follow. Twitter has grown tremendously in popularity and is now used in many areas ranging from political activism to spam. Twitter has been a source of activity not just for users but also for third-party interest groups. In this context, there are increasing number of applications on Twitter which enhance its usability. An example is a user-friendly application that lets people streamline tweets from their mobile (example: Tweet Deck). The purpose of this project is to find out trends by aggregating the data in social networking site such as Twitter. Analysis of Twitter Data to evaluate sentiment to the user tweets, find the hash-tag used by the users and to find the most popular user for the twitter data related to Commonwealth Games Eid Festival India Vs England Test Match Gaza attack Launch of Xiaomi Smartphone on Flipkart Every day, we create 2.5 quintillion bytes of data (source:IBM) — so much that 90% of the data in the world today has been created in the last two years alone. This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few. This data is big data.

## II. LITERATURE SURVEY.

### 2.1 Existing System

The social network service, is a rich source of information on customer. Examples include Facebook, MySpace, Dig, Twitter and JISC listservs on the academic side, because of its ease of use, speed and reach, social media is fast changing the public chat in society and setting trends and agendas in topics that range from the environment and politics to technology and the entertainment industry. Since social media can also be understood as a form of collective wisdom, it can be used to predict real-world outcomes. Twitter has millions of users that spread millions of personal posts on a daily basis. And this gives us the opportunity to

study social human subjectivity. Manual classification of thousands of posts for opinion mining task is unfeasible for a human being.

### 2.2 Problem Statement

The problem in sentiment analysis is classifying the polarity of a given text at the document, sentence, or feature/aspect level. Whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral.

## III. DESCRIPTION OF TV SHOW POPULARITY PREDICTION USING BIG DATA ANALYTIC

### 3.1 Block Diagram

As discussed in the introduction the block diagram of Tv Show Popularity Prediction using Big Data Analytic is illustrated in the fig 3.1. The overview of the block diagram is described below.

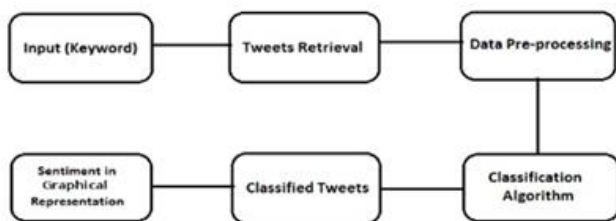


fig. 4. Block diagram of TV Show Popularity Prediction using Big Data Analytic

### 3.2 Description of block diagram

#### 3.2.1 Input (Keyword)

We search for a particular keyword which is regarding a TV series or Web Series or Movies and then extract the tweets related to Keyword.

#### 3.2.2 Tweets Retrieval

To ensure that we obtained all tweets referring to a TV show, we used hash tags of TV show as search arguments. Tweets were extracted over frequent intervals using the Twitter Search API 1.1

#### 3.2.3 Data Pre-Processing

Twitter users can simply create a hashtag by prefixing a word with a hash symbol “#hashtag.” For instance,

various hashtags, such as #balikavadhu, #BALH, #CID had been created for different TV shows.

### 3.2.4 Classified Tweets

This clustering analysis allows an object not to be part of a cluster, or strictly belong to it, calling this type of grouping hard partitioning. On the other hand, soft partitioning states that every object belongs to a cluster in a determined degree.

### 3.2.5 Sentiment in Graphical Representation

The analysis of Sentimental comment and predicting whether it is good or bad comments. Easy prediction of TV Show trending based on people rating. Good or Bad comments based on peoples reviews or comments. Easy importing of data and exporting it into the graph. Graphical data in the printable format. The visitor will get to know the show popularity.

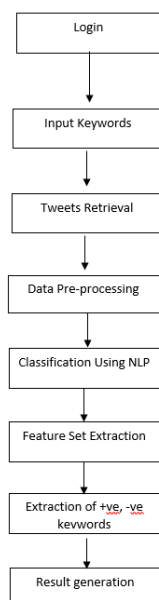
## IV. IMPLEMENTATION.

Implementation of the main screen giving the created project specific domain name and the host for 24hr availability of the system, screen that follows the welcome page giving various options, screens for each of the options During this milestone period, it would be a good idea for the team (or one person from the team) to start working on a test-plan for the entire system. This test-plan can be updated as and when new scenarios come to mind.

## V. RESULTS.



## VI. FLOW CHART.



## VII. RESEARCH AND METHODOLOGY.

### 6.1 Algorithm for TV Show Popularity Prediction Using Big Data Analytic

#### PORTER STEMMER ALGORITHM

- Step1. Gets clear of plurals and -ed or -ing suffixes
- Step2. Turns terminal y to i when there is another vowel in the stem
- Step3 Maps double suffixes to single ones: -ization, -ational, etc.
- Step4. Deals with suffixes, -full, -ness etc.
- Step5. Takes off -ant, -ence, etc. Removes a final -e

## VIII. ADVANTAGES, FUTURE SCOPE AND APPLICATIONS.

### 7.1 Introduction

Every prototype designed has its own advantages, applications and future scope. In this chapter we have discussed about the advantages, applications and future scope of TV Show Popularity Prediction Using Big Data Analytic.

### 7.2 Future Scope

Using hashtags to collect training data prove useful, as did using data collected based on positive and negative emotions. However, which method produces the better training data and whether the two sources of training data are complementary may depend on the type of features used. Next focus of research would be to work on how Text Mining assists in extracting actionable intelligence from voluminous amounts of unstructured information sources of any social data within a domain system.

## IX. REFERENCES.

- [1] Akshay Java, Xiaodan Song, Tim Finin and Belle Tseng. Why we twitter: understanding microblogging usage and communities. Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, pages 56–65, 2007
- [2] Amrapali Mhaisgawali, Dr. Nupur Giri Twitter based TV Rating Analysis with TV -viewing and Sentiment Analysis”, Current Trends in Information Technology ISSN: 2249-4707 Volume 4, Issue 1 [www.stmjournals.com](http://www.stmjournals.com)
- [3] Bernardo A. Huberman, Daniel M. Romero, and Fang Wu. Social networks that matter: Twitter under the microscope. First Monday, 14(1), Jan 2009.
- [4] B. Jansen, M. Zhang, K. Sobel, and A. Chowdury. Twitter power: Tweets as electronic word of mouth. Journal of the American Society for Information Science and Technology, 2009
- [5] Daniel Gruhl, R. Guha, Ravi Kumar, Jasmine Novak and Andrew Tomkins. The predictive power of online chatter. SIGKDD Conference on Knowledge Discovery and Data Mining, 2005.
- [6] G. Mishne and N. Glance. Predicting movie sales from blogger sentiment. In AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs, 2006.
- [7] Mahesh Joshi, Dipanjan Das, Kevin Gimpel and Noah A. Smith. Movie Reviews and Revenues: An Experiment in Text Regression NAACL-HLT, 2010.
- [8] Ramesh Sharda and Dursun Delen. Predicting box office success of motion pictures with neural networks. Expert Systems with Applications, vol 30, pp 243– 254, 2006.
- [9] Sitaram Asur, Bernardo A. Huberman “Predicting the Future With Social Media” 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology
- [10] Twitter: <http://twitter.com>.