



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 11, November 2021

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 7.569

 9940 572 462

 6381 907 438

 [ijrset@gmail.com](mailto:ijrset@gmail.com)

 [www.ijrset.com](http://www.ijrset.com)



# Big Data Analytics for Program Popularity Prediction in Broadcast TV Industry

Rutuja S Waware<sup>1</sup>, Sandeep N Gite<sup>2</sup>

P.G. Student, Department of Computer Engineering, MSS's Engineering College, Jalna, Maharashtra, India<sup>1</sup>

Assistant Professor, Department of Computer Engineering, MSS's Engineering College, Jalna, Maharashtra, India<sup>2</sup>

**ABSTRACT:** The precise and timely prediction of program popularity is of great value for content providers, advertisers, and broadcast TV operators. This information can be beneficial for operators in TV program purchasing decisions and can help advertisers formulate reasonable advertisement investment plans. Moreover, in terms of technical matters, a precise program popularity prediction method can optimize the whole broadcasting system, such as the content delivery network strategy and cache strategy. Several prediction models have been proposed based on video-on-demand (VOD) data from YouKu, YouTube, and Twitter. However, existing prediction methods usually require a large quantity of samples and long training time, and the prediction accuracy is poor for programs that experience a high peak or sharp decrease in popularity. This paper presents our improved prediction approach based on trend detection. First, a dynamic time warping distance-based  $K$ -medoids algorithm is applied to group programs' popularity evolution into four trends. Then, four trend-specific prediction models are built separately using random forests regression. According to the features extracted from an electronic program guide and early viewing records, newly published programs are classified into the four trends by a gradient boosting decision tree. Finally, by combining forecasting values from the trend-specific models and the classification probability, our proposed approach achieves better prediction results. The experimental results on a massive set of real VOD data from the Jiangsu Broadcasting Corporation show that, compared with the existing prediction models, the prediction accuracy is increased by more than 20%, and the forecasting period is effectively shortened.

**KEYWORDS:** Dynamic time warping, random forests regression, decision tree, naïve bayes.

## I. INTRODUCTION

As the maturity and popularity of high-definition (HD) and 3D technology increase, IP video traffic will become a major part of all consumer Internet traffic. According to data published by Cisco Visual Networking Index in July 2016 [1], the consumption of Internet video streams of broadcast TV will continue to grow at a rapid pace and will constitute 26% of consumer Internet video traffic by 2020. However, user attention is not uniformly distributed among all programs. Only a few programs can attract massive user attention; the remaining programs are left without anybody to watch them. Take Tencent video, for instance [2]. There have been 45 billion cumulative requests for the top-50 programs, which is more than 80 percent of the total number of requests.

In this context, it is of great importance to predict the popularity of broadcast TV programs. First, using the program popularity prediction results, the audience will save much time when trying to find valuable TV programs among massive collections of video resources, which will improve user satisfaction and retention. Second, based on program popularity forecasting data, a company will be able to maximize its advertising effect by choosing the TV programs with highest potential. Finally, with the help of the popularity prediction model, a broadcast TV operator will be able to optimize the configuration of the network in advance by deploying enough transmission and storage resources to distribute popular programs. For example, by applying Auto Regressive and Moving Average (ARMA) models to real traces extracted from YouTube, accurate predictions can be obtained, as shown by Hassine [3]. Thus, an original solution that combines the predictions of several ARMA models is proposed to determine which contents should be cached. This will greatly help improve the ability of the Content Delivery Networks (CDNs) to promptly react to instant feedback regarding consumer demand in this new video era.

However, accurately predicting the popularity of broadcast TV programs is a challenging task. First, there are many factors influencing TV program popularity that are difficult to measure, such as the quality of the program and the interests of the audience. A hybrid-stream model is proposed to solve these problems for video analysis [4]. Moreover, the relationship between popular events in the real world and in TV programs cannot be easily introduced into the



prediction model. Last, there is a massive gap between the popularity evolutionary trends of different programs, which should be considered when designing the prediction model. In this paper, cooperating with a broadcast TV operator, we analyze massive user behavior data and present our improved method for predicting the popularity of broadcast TV programs.

## II. RELATED WORK

Due to the rich variety and timeliness of TV content, the semantic understanding of broadcast TV programs is more difficult than that of news, microblogging or other web content. An ideal prediction model for broadcast TV programs achieves not only high prediction accuracy but also good calculation performance, which means that the prediction result is available before audience interest fades. At present, there is little research on program popularity prediction for broadcast TV. The existing popularity prediction methods are for other media formats but can be used as references. Commonly used web content popularity prediction methods include cumulative growth, temporal analysis and evolutionary trends.

2.1 The development of Facebook and other social media websites encouraged more people to share their thoughts. Undertakings saw this and began their very own fan pages for the clients to connect with them and express their steadfastness. Yu-Hsuan Chang, Chen-Ming Wu, Tsun Ku and GwoDong Chen taken a shot at an anticipating model (2013) that utilized the substance created in the television program fan pages by watchers and Artificial Neural Network to perform conjectures on program evaluations. The exploration developed a program evaluations estimate module dependent on Back-engineering Network. It at that point utilizes trained Artificial Neural Network to play out an appraisals conjecture for up and coming projects.

2.2 In 2017, Chengang Zhu, Guang Cheng and Kun Wang with Big Data to foresee the prevalence of communicate television programs. With the assistance of the forecast model, a communicate television administrator will most likely upgrade the setup of the system ahead of time by conveying enough transmission and capacity assets to disperse well known projects. Auto Regressive and Moving Average (ARMA) models were connected to genuine follows extricated from YouTube and exact forecasts were acquired. They connected a dynamic time warping (DTW) remove based k-medoids calculation to group programs with comparable fame into four transformative patterns, which can catch the innate heterogeneity of program ubiquity.

2.3 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.

2.4. TITLE: Internet of things: Vision, applications and research challenges.

AUTHOR: D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac.

Internet of Things (IoT) is the convergence of the Internet with smart objects or “things belonging to Internet” to exchange the real-world information. Now the Internet is moving out from Internet of people to the Internet of everything. Billions of devices expected to connect to the Internet in future that will require a well manageable mechanism where the objects can sense and behave accordingly by themselves. This paper presents how the data/information from smart objects gives rise to the actual discoveries in IoT world. IoT vision is presented with a simple architecture of IoT. different IoT application domains with their use cases are discussed. Finally, a taxonomy of IoT research areas is presented with different research challenges.

2.5. TITLE: An architectural approach towards the future internet of things.

AUTHOR: D. Uckelmann, M. Harrison, and F. Michahelles.

Many of the initial developments towards the Internet of Things have focused on the combination of Auto-ID and networked infrastructures in business-to-business logistics and product life cycle applications. However, a future Internet of Things can provide a broader vision and also enable everyone to access and contribute rich information about things and locations. The success of social networks to share experience and personalised insights shows also great potential for integration with business-centric applications. The integration and interoperability with mainstream business software platforms can be enhanced and extended by real-time analytics, business intelligence





and agent-based autonomous services. Information sharing may be rewarded through incentives, thus transforming the Internet of Things from a cost-focused experiment to a revenue-generating infrastructure to enable trading of enriched information and accelerate business innovation. Mash-ups and end-user programming will enable people to contribute to the Internet of Things with data, presentation and functionality.

2.6.TITLE:Towards resource sharing in mobile device clouds: Power balancing across mobile devices.

AUTHOR:A. Mtibaa, A. Fahim, K. A. Harras, and M. H. Ammar.

Despite the increased capabilities of mobile devices, mobile application resource requirements can often transcend what can be accomplished on a single device. This has been addressed through several proposals for efficient computation offloading from mobile devices to remote cloud resources or closely located computing resources known as cloudlets. In this paper we consider an environment in which computational offloading is performed among a set of mobile devices. We call this environment a Mobile Device Cloud (MDC). We are interested in MDCs where nodes are { $\setminus$ em highly collaborative}. We develop computational offloading schemes that of the ensemble of mobile devices where we consider the network to be alive as long as no device has depleted its battery. As a secondary contribution in this work, we develop and use an experimentation platform that allows us to evaluate a range of computational models and profiles derived from a realistic testbed. We use this platform as a first step in an evaluation exercise that demonstrates the effectiveness of our computation offloading algorithms in extending the lifetime of an MDC.

2.7.TITLE:Towards mobile opportunistic computing.

AUTHOR:A. Mtibaa, K. A. Harras, K. Habak, M. Ammar, and E. W. Zegura.

With the advent of wearable computing and the resulting growth in mobile application market, we investigate mobile opportunistic cloud computing where mobile devices leverage nearby computational resources in order to save execution time and consumed energy. Our goal is to enable generic computation offloading to heterogeneous devices that include Cloud, mobile devices, and cloudlets. We propose a generic and flexible architecture that maximizes the computation gain with respect to various objective functions such as, minimizing the response time, reducing the overall energy consumption, and increasing the network lifetime. This novel architecture is designed to automate computation offloading to numerous compute resources over disrupted network connections.

2.8. Yusuke Fukushima, Toshihiko Yamasaki and KiyoharuAizwa of the University of Tokyo in 2016 taken a shot at an examination to gauge the appraisals of shows before their communicate without utilizing video substance or group of onlookers' reaction to them. They considered the ubiquity of performing artists included, every day perspectives onscreen characters' Wikipedia and notices of them on Twitter were utilized as markers of prevalence. It was clear that a show gets a higher rating if various famous on-screen characters are in it. Anticipate the gathering of people evaluations with high exactness by considering the performers and staff associated with expansion to the traditional data, for example, broadcast appointments and the stations in which they play.

### III. METHODOLOGY

#### 3.1 Existing System:

The precise and timely prediction of program popularity is of great value for content providers, advertisers, and broadcast TV operators. This information can be beneficial for operators in TV program purchasing decisions and can help advertisers formulate reasonable advertisement investment plans. Moreover, in terms of technical matters, a precise program popularity prediction method can optimize the whole broadcasting system, such as the content delivery network strategy and cache strategy. Several prediction models have been proposed based on video-on-demand (VOD) data from YouKu, YouTube, and Twitter. However, existing prediction methods usually require a large quantity of samples and long training time, and the prediction accuracy is poor for programs that experience a high peak or sharp decrease in popularity

#### 3.2 Proposed System

This paper presents our improved prediction approach based on trend detection. First, a dynamic time warping distance-based  $K$ -medoids algorithm is applied to group programs' popularity evolution into four trends. Then, four trend-specific prediction models are built separately using random forests regression. According to the features extracted from an electronic program guide and early viewing records, newly published programs are classified into the four trends by a gradient boosting decision tree. Finally, by combining forecasting values from the trend-specific models and the classification probability, our proposed approach achieves better prediction results. The experimental results on a massive set of real VOD data from the Jiangsu Broadcasting Corporation show that, compared with the existing



prediction models, the prediction accuracy is increased by more than 20%, and the forecasting period is effectively shortened.

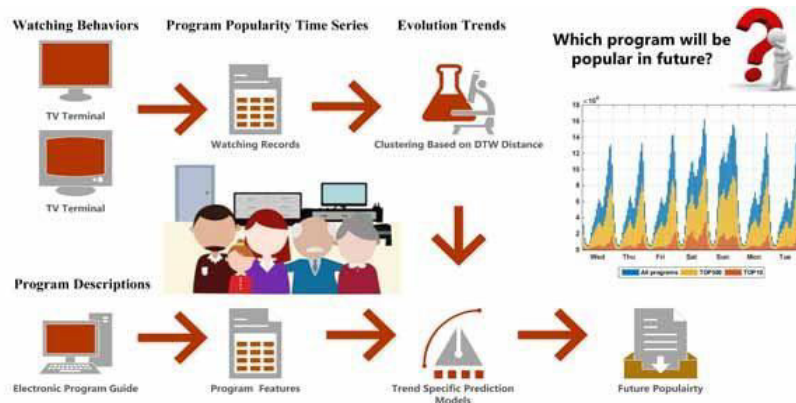


Fig3.1. Big Data Analytics for Program Popularity Prediction

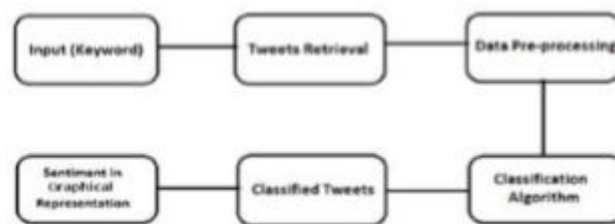


Fig.3.2Block diagram of TV Show Popularity Prediction using Big Data Analytic

### 3.3 Input Design

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

### OBJECTIVES

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

2.It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.



3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow

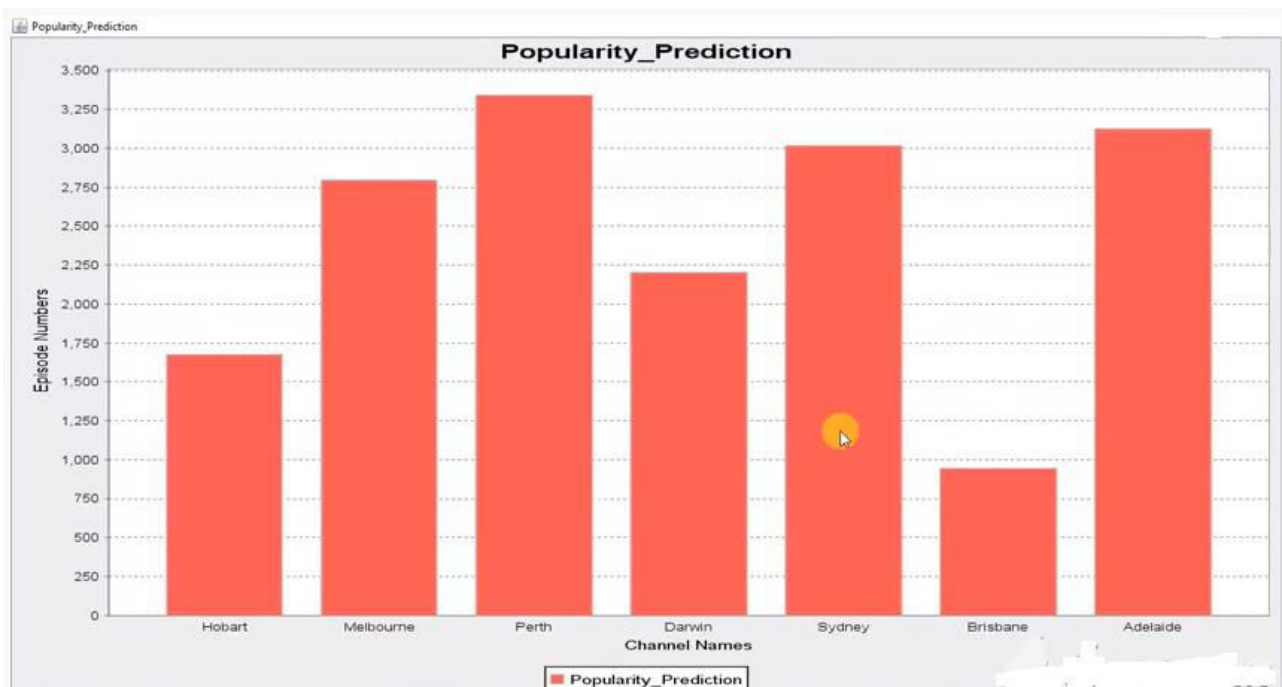
#### IV.EXPERIMENTAL RESULTS

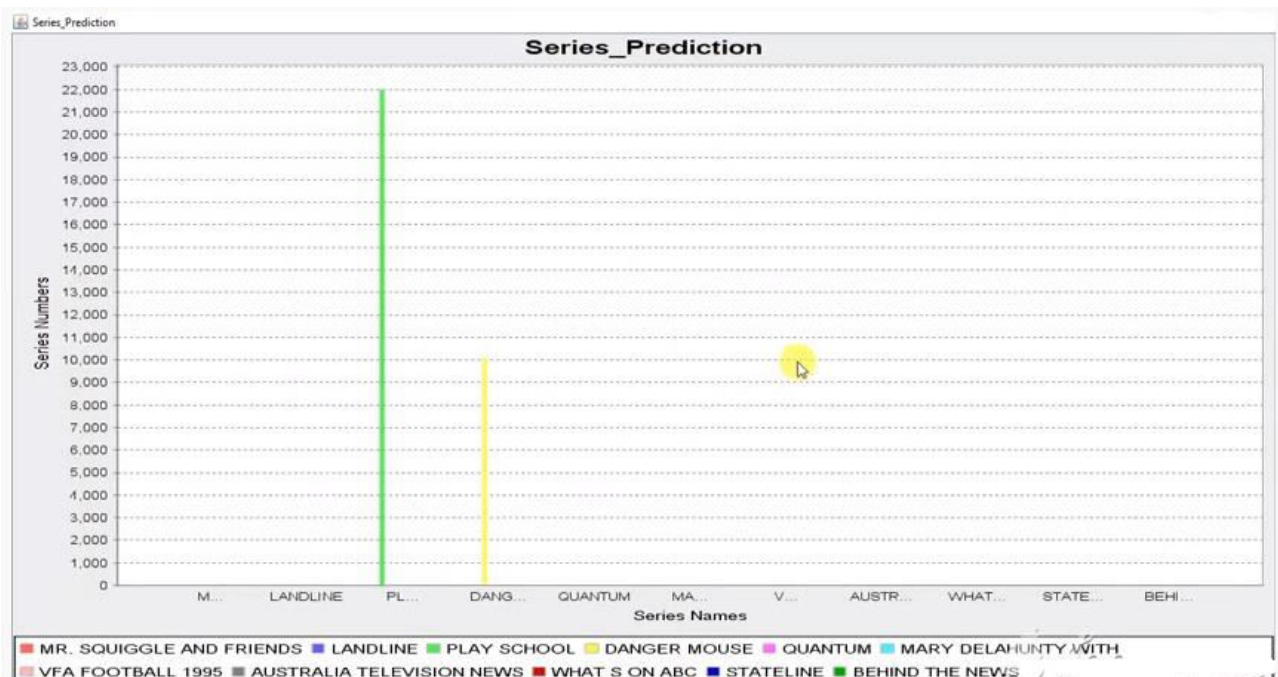
A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.
2. Select methods for presenting information.
3. Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

- ❖ Convey information about past activities, current status or projections of the
- ❖ Future.
- ❖ Signal important events, opportunities, problems, or warnings.
- ❖ Trigger an action.
- ❖ Confirm an action.





## V. CONCLUSION

In this paper, we have analyzed massive user behavior data and presented our improved method to predict the popularity of broadcast TV programs. To the best of our knowledge, this is the first work to tackle the problem of predicting program popularity in the broadcast TV platform. We applied a dynamic time warping (DTW)-distance-based K-Medoids algorithm to group programs with similar popularity into 4 evolutionary trends, which has the ability to capture the inherent heterogeneity of program popularity. Moreover, we built trend-specific prediction models using random forests regression, which have better overall predictive performance than a single model trained on the entire data set. We performed an extensive experimental evaluation of our method, in which we compared it with 3 representative methods. Our method outperforms the others, with a gain in accuracy of at least 20%, and can give reliable prediction results much faster.

## REFERENCES

- [1] "Cisco visual networking index: Forecast and methodology, 2016\_2021," Cisco, San Jose, CA, USA, Tech. Rep., 2016.
- [2] Tencent Video. (2016). *Top Video Lists*. [Online]. Available: [http://v.qq.com/rank/detail/2\\_-1\\_-1\\_-1\\_1\\_1.html](http://v.qq.com/rank/detail/2_-1_-1_-1_1_1.html)
- [3] N. B. Hassine, R. Milocco, and P. Minet, "ARMA based popularity prediction for caching in content delivery networks," in *Proc. Wireless Days*, Mar. 2017, pp. 113\_120.
- [4] K. Wang, J. Mi, C. Xu, Q. Zhu, L. Shu, and D.-J. Deng, "Real-time load reduction in multimedia big data for mobile Internet," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 12, no. 5s, p. 76, 2016.
- [5] F. Figueiredo, J. M. Almeida, M. A. Gonçalves, and F. Benevenuto, "TrendLearner: Early prediction of popularity trends of user generated content," *Inf. Sci.*, vols. 349\_350, pp. 172\_187, Jul. 2016.
- [6] M. Tsagkias, W. Weerkamp, and M. de Rijke, "News comments: Exploring, modeling, and online prediction," in *Advances in Information Retrieval*. Cham, Switzerland: Springer, 2010, pp. 191\_203.
- [7] A. Tatar, P. Antoniadis, M. D. de Amorim, and S. Fdida, "From popularity prediction to ranking online news," *Social Netw. Anal. Mining*, vol. 4, pp. 174\_183, Dec. 2014.
- [8] H. Pinto, J. M. Almeida, and M. A. Gonçalves, "Using early view patterns to predict the popularity of youtube videos," in *Proc. 6th ACM Int. Conf. Web Search Data Mining*, 2013, pp. 365\_374.
- [9] A. Kaltenbrunner, V. Gómez, and V. López, "Description and prediction of slashdot activity," in *Proc. Latin Amer. Web Conf. (LA-WEB)*, 2007, pp. 57\_66.
- [10] G. Szabo and B. A. Huberman, "Predicting the popularity of online content," *Commun. ACM*, vol. 53, no. 8, pp. 80\_88, 2010.



- [11] K. Wang, Y. Wang, Y. Sun, S. Guo, and J. Wu, "Green industrial Internet of Things architecture: An energy-efficient perspective," *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 48\_54, Dec. 2016.
- [12] J. G. Lee, S. Moon, and K. Salamatian, "Modeling and predicting the popularity of online contents with Cox proportional hazard regression model," *Neurocomputing*, vol. 76, no. 1, pp. 134\_145, 2012.
- [13] K. Wang, H. Lu, L. Shu, and J. J. P. C. Rodrigues, "A context-aware system architecture for leak point detection in the large-scale petrochemical industry," *IEEE Commun. Mag.*, vol. 52, no. 6, pp. 62\_69, Jun. 2014.
- [14] A. Tatar *et al.*, "Predicting the popularity of online articles based on user comments," presented at the Proc. Int. Conf. Web Intell., Mining Semantics, Sogndal, Norway, 2011.
- [15] S.-D. Kim, S.-H. Kim, and H.-G. Cho, "Predicting the virtual temperature of Web-blog articles as a measurement tool for online popularity," in *Proc. IEEE 11th Int. Conf. Comput. Inf. Technol.*, Sep. 2011, pp. 449\_454.
- [16] S. Jamali and H. Rangwala, "Digging digg: Comment mining, popularity prediction, and social network analysis," in *Proc. Int. Conf. Web Inf. Syst. Mining*, 2009, pp. 32\_38.





**Impact Factor:  
7.569**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details