

PREDICTING POPULARITY OF KOREAN CONTENTS IN ARAB COUNTRIES USING A DATA MINING TECHNIQUE

Park Young Eun¹, Soumaya Chaffar², Kim Myoung Sook³ and Ko Hye Young⁴

¹College of Business Administration, Prince Sultan University

²Department of Computer Science,
Prince Sultan University, Riyadh, Saudi Arabia

³Department of Business Administration,
Future Convergence Industry College, Seoul Women's University

⁴Department of Digital Media, Future Convergence Industry College,
Seoul Women's University

ABSTRACT

Recently, many people in the Middle East and North Africa enjoy watching a variety of Korean contents such as Korean dramas, films, broadcasting programs and listening to Korean Pops. The Korean wave refers to the phenomenon of Korean entertainment and popular culture rolling over the world with TV dramas, films and pop music. Also it is known as "Hallyu" literally meaning 'flow from Korea' in Korean. This study examines the analysis of pattern on Arab countries (Middle East and North Africa) Consumers' consumption of the Korean Contents using social media, Facebook data. Then we focus on developing Predictive System using a Data Mining Technique.

KEYWORDS

Korean Wave (Hallyu in Korean), Social Media, Data Mining, Predictive Analysis

1. INTRODUCTION

The Korean Wave (K-wave), or Hallyu literally meaning 'flow from Korea' in Korean, referred as "the growing popularity of Korean pop culture, such as TV dramas, films, pop music, fashion, beauty, and online games being widely embraced and shared among the people of Asian countries in the late 1990s. In addition, with the rapid spread of social media like Facebook, YouTube and Twitter, K-wave has expanded its fandom outside of Asia to the West. That is, currently, this Korean wave has become the phenomenon of Korean popular culture rolling over the world not only Asian countries but also North and Latin America, Europe, even Middle Eastern countries and North Africa with Korean entertainment contents. The world-wide success of Korean pop culture contributed to improve the 'Korea' image and make a positive impact on Korean economy (Ahn et al., 2013).

In the late 1990s, a few Korean TV dramas (hereafter, K-dramas), such as *What is Love All About?* (1997) and *Stars in My Heart* (1997), became popular in East and Southeast Asia and provided a wide range of Asian audiences with glimpses of Korean pop culture. The initial Korean wave was followed by the megahits of three K-dramas—*Autumn Fairy Tale* (2000), *Winter Sonata* (2002), and *Dae Jang Geum* (2003)—in Japan, Thailand, Singapore, and Hong Kong between 2002 and 2006. In the early 2000s, Hallyu was also led by the success of K-pop artists, such as BoA, Big Bang, and Dong Bang Shin Ki(TVXQ), in several Asian countries. In recent years, K-pop fandom has been evident even outside of Asia (Hong Mercier, 2013; Lansky, 2012). Finally, Psy and his song *Gangnam Style* were not only a world-wide phenomenon, but also a great significant turning point in the history of K-Pop and Korean pop culture. This is explicit just from Psy's achievements in the year 2012 alone. *Gangnam Style* has reached up to 1.7 billion views on YouTube, becoming the most-liked video in the site's history and topped the charts in over forty-one countries. What is more, the mere fact that small towns in South America, Southeast Asia, and the Middle East know the "horse dance" of *Gangnam Style* is an unexplainable achievement—not done justice by words (Park, 2015). The global dissemination of Korean popular culture such as *Gangnam Style* would not have been possible without global social media or social network service (SNS) sites (Park, 2013).

However, while it has been more than 15 years since the Korean pop culture phenomenon has emerged, academic analyses have not sufficiently addressed its consumption of Middle Eastern and North African area from a global perspective. The existing literatures on the Korean Wave focus their attention on the Asian market and tend to still define it primarily as an intra-Asian flow of particular forms of content without sufficiently addressing its dimension of social media and its technology and effect on Korean Wave from a global perspective (Jin and Yoon, 2016). Thus, it look over how a wide range of Western, Middle Eastern and African fans of Korean pop culture engage with social media and are networked with other fans (Jin and Yoon, 2016).

In addition, there has no analytical research and forecasting system to find the key factors affecting consumers' demand in these regions. In this regard, drawing on consumption pattern with these regions' fans of the recent Korean wave, this study explores how the Hallyu phenomenon is integrated into Middle East and North Africa through a social media. Finally, this paper aims to develop the pattern modeling of consumption behavior, systemically, in Middle East and North Africa using Data Mining techniques to build Korean contents management system and enhance marketing performance expectations by applying the results to the company's prediction system.

2. RELATED WORK

Data mining is an essential tool in the process of knowledge discovery in databases in which intelligent methods are applied in order to extract patterns. Predicting is one of the most interesting and challenging tasks where to develop data mining applications. The use of computers with automated tools, large volumes of data are being collected and made available to the research groups (Shweta, 2012). As a result, many researches were carried out on various datasets using the data mining techniques to enhance forecasting in the business and medical fields.

Shweta (2012) has discussed various data mining approaches that have been utilized for breast cancer diagnosis and prognosis. The most effective way to reduce breast cancer deaths is to detect

it earlier. Early diagnosis needs an accurate and reliable diagnosis procedure that can be used by physicians to distinguish benign breast tumors from malignant ones without going for surgical biopsy (Shweta, 2012). As a result, data mining techniques has become a popular research tool for medical researchers to identify and exploit patterns and relationships among large number of variables, and made them able to predict the outcome of a disease using the historical datasets (Delen et al., 2005; Sarvestan et al., 2010 ; Shweta, 2012).

In addition, Der-Chiang Li et al.(2012) thought the overall electricity consumption, treated as a primary guideline for electricity system planning, is a major measurement to indicate the degree of a nation's development and The electricity consumption forecast is especially important with regard to policy making in developing countries. However, it is difficult to obtain accurate predictions using long-term data, and thus forecasting with limited (short-term) data is more effective and of considerable interest (Der-Chiang Li et al., 2012). Khana et al. (2013) also approached with similar idea. They explore three different data mining techniques for detecting abnormal lighting energy consumption using hourly recorded energy consumption and peak demand (maximum power) data. Two outliers' detection methods are applied to each class and cluster for detecting abnormal consumption in the same data set. In each class and cluster with anomalous consumption the amount of variation from normal is determined using modified standard scores (Khana, I., Capozzolia,A.,Corgnatia, S. P. and Cerquitellib,T.,2013).This study gives valuable implications for building energy management systems to reduce operating cost and time by not having to detect faults manually or diagnose false warnings. It means that will be useful for developing fault detection and diagnosis model for the whole building energy consumption. Besides, Fan et al.(2014) present data mining based approach to developing ensemble models for predicting next-day energy consumption and peak power demand, with the aim of improving the prediction accuracy(Fan, Xiao and Wang, 2014). Their ensemble model is developed and the weights of the eight predictive models are optimized using genetic algorithm. This approach shows that we can analyze the large consumption data and based on this analysis, we can apply the predictive models to know the consumption and peak demand for next season and develop strategies of fault detection, diagnosis, operation optimization and interactions between consumers and providers in the business field.

Predictive analytics has been useful in predicting the frequency of trading and the stock price for the next day, based on data from social media such as Twitter. This model was developed by Riverside and other researchers. A trading strategy supported by this model was created by three researcher, Ruiz et al.(2012). Their trading strategy performed 1.4%-11% better and other baseline strategies. In addition, during a 4-month simulation, this strategy outperformed the Dow Jones Industrial Average. Hristidis's study looked further than the effect of negative and positive sentiment on stock prices, it looked at the quantity of tweets and the interrelationship between tweets, users and topics.

Recently, O'Connor (2015) has found the relationship between Facebook popularity and consumer brand stock prices. The premise of his study aimed at finding out whether popularity (measured in Facebook likes) affected performance (share prices). He identified 30 brands that had the maximum number of followers, and tracked the likes these companies got for a period of one year as well as their share price on a daily basis. He found that 99.95% of the changes in share prices could be explained by the change in the number of fans. He did not see a direct relationship between likes and an increase or decrease in share prices, however, stock market trends were affected by the appreciation a company got on social media. The majority of a change in a particular company's stock price was linked with the likes that brand received for that period

or day. Finally, the data of social media sentiment has become most applicable to produce other predictive models.

Jin and Yoon(2014) examine how Hallyu fans engage with a social media-saturated environment, drawing on qualitative interviews with North American fans of Korean pop culture. In comparison to the existing cultural analyses of the Korean wave, which focus, at best, on the content of particular genres or texts and their consumption, they map out transnational pop cultural flows with reference to the media environment through which the participatory culture of media users is spread (Jin& Yoon, 2014).

However, most studies tend to focus on either the role of digital technology of social media or predictive analytics using data mining techniques without fully consideration their conjunction between social media and data mining technique. Based on these existing studies and our awareness of these limitations, we approach to explore predictive models with social media using data mining technique for application of this concept into the consumption of Korean pop cultures. From this study, we, therefore, engage with the notion of spreadable social media in disseminating Korean pop cultures, and developing predictive models of K-contents' consumption and demand for the future, especially in the Middle East and North Africa.

3. METHODOLOGY

3.1. Data Collection

In order to understand popularity growth of Korean contents in Arabic countries, we collected data from two popular Facebook pages: the first one about 'Korean movies and drama' and the second one about 'K-pop'. Numbers of likes recorded on a daily basis by different countries over a period of two years (October 10th 2014 to October 10th 2016) have been collected. Then, we selected only those of Middle East and North Africa. Four North African (Algeria, Egypt, Morocco and Tunisia) and three Middle Eastern countries (Iraq, Saudi Arabia and United Arab Emirates) liked the Korean Movies and Drama. These countries except the United Arab Emirates also liked the K-pop page (see table 1).

Table 1. Number of likes of Korean pages by North African and Middle Eastern Countries over a Period of 2 Years

		Korean Movies & Drama	K-Pop
North Africa	Algeria	7155335	260879
	Egypt	10475598	198190
	Morocco	7152532	269912
	Tunisia	5059424	386222
Middle East	Iraq	3421466	166695
	Saudi Arabia	1286017	105315
	United Arab Emirates	633538	0

During the last two years and compared to other Arab countries (Algeria, Egypt, Morocco, Iraq and Saudi Arabia), Tunisia has the smallest population (around 11.317 million people), nevertheless it has the largest number of fans of K-pop (about386.222 likes). On the other hand,

Egypt (about 10.475.598likes) has the largest numbers of likes of Korean movies and drama compared to other Arab countries.

3.2. Predictive Data Mining

Aiming to analyze the popularity of Korean contents in Middle Eastern and North African countries, we adopted a data-driven approach based on Data Mining techniques. The collected data as described above represent a time series composed of a chronological sequence of observations on number of likes by country associated to public Korean pages (movies/drama and k-pop). Predict future likes trends of Korean contents for Middle Eastern and North African countries can look ahead to future consumption behavior in order to maximize the success and profitability of Korean contents in these regions. This task falls under time series forecasting which is performed with different techniques including statistical and machine learning ones. The latest are often more powerful than the classical statistical techniques such as ARMA and ARIMA(Saigal and Mehrotra, 2012).Different Machine Learning algorithms have been used for time series forecasting such as Linear Regression, Robust Regression, Gaussian Processes, Support Vector Machine (SVM), etc. SVM algorithms have been used with a considerable success and often outperformed other methods (Ristanoski, Liu and Bailey, 2013). In this research work we will use SVM for regression (namely SMOreg in Weka Software) in order to predict future likes trends of Korean contents for different Middle Eastern (Iraq, Saudi Arabia and United Arab Emirates) and North African countries (Algeria, Egypt, Morocco and Tunisia). We will use the Mean Absolute Error in order to evaluate our model. The Mean Absolute Error (MAE) is used to measure how close predictions are to the eventual outcomes. MAE is given by the following equation:

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i|$$

where f_i is the predicted value and y_i is the true value.

4. RESULTS & DISCUSSION

In order to forecast the future trend for the number of likes of Korean music, movies and dramas, we employed WEKA forecasting plugin (Saigal and Mehrotra, 2012), which is a time series analysis and forecasting model. The target variable to be predicted is the number of likes that will be collected in 1 day until 10 days into the future.

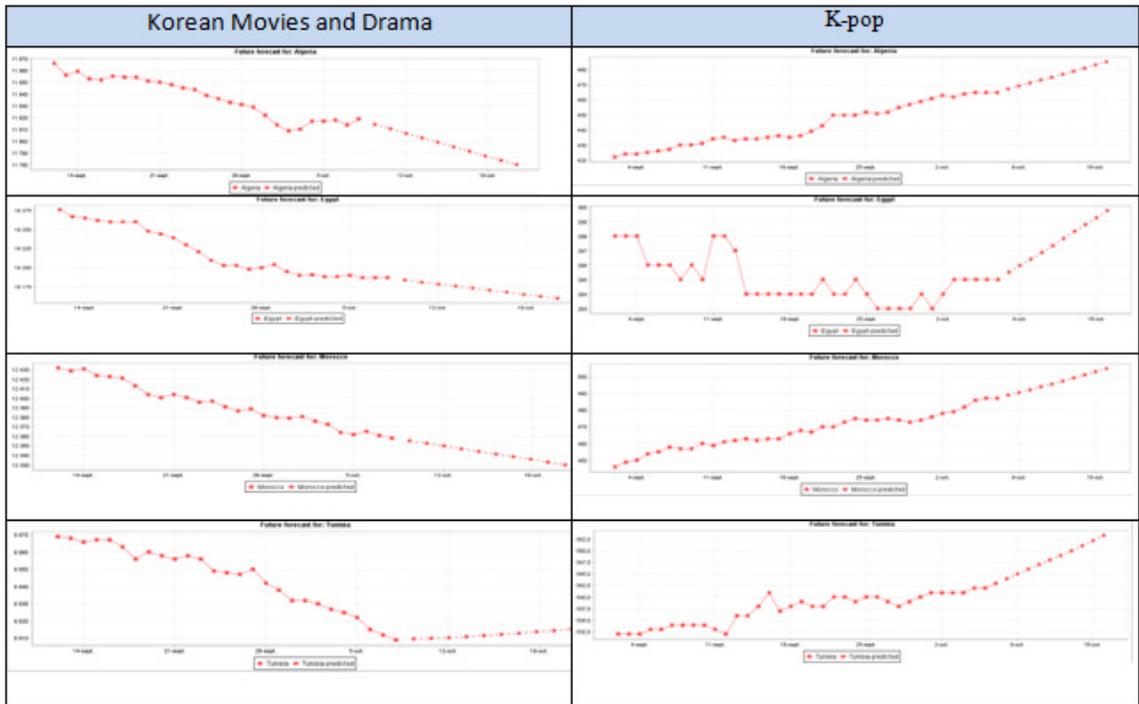
The collected data was divided into training data (used to estimate the model) and test data (used to evaluate the forecasts).The size of the test data set depends on the size of the whole data and number of time units to forecast and should ideally be at least as large as the maximum forecast horizon required (Hyndman and Khandakar, 2008). In this research 5% of the total sample used to evaluate the forecasts for a period of 10 days. The performance results of the algorithms are based on the Mean Absolute Error (MAE) and are reported in Table 2.

Table 2. Evaluation of SMOreg for predicting the number of likes of Korean pages within a period of 10 days using MAE

MAE	Korean Movies and Drama										K-pop									
	1 day	2 days	3 days	4 days	5 days	6 days	7 days	8 days	9 days	10 days	1 day	2 days	3 days	4 days	5 days	6 days	7 days	8 days	9 days	10 days
Algeria	3.65	5.87	8.11	10.19	12.02	13.68	15.13	16.58	18.03	19.97	1.29	2.05	2.51	3.01	3.28	3.42	3.64	3.8	3.95	4.06
Egypt	3.82	6.03	8.03	9.84	11.43	12.73	15.38	17.94	20.48	23.24	0.82	1.25	1.68	2.04	2.36	3.24	2.76	3.66	4.14	4.59
Morocco	2.76	3.75	3.82	4.21	4.35	4.53	5.35	5.77	5.19	5.78	1.34	1.74	2.11	2.7	2.95	3.45	3.65	3.95	4.46	4.84
Tunisia	3.35	5.61	8.11	10.7	13.37	16.16	18.88	21.40	24.37	27.15	1.12	1.61	2.08	2.58	2.94	3.34	3.83	4.24	4.77	5.39
Iraq	1.98	2.56	3.74	4.86	6.21	7.57	9.22	10.79	12.51	14	0.66	0.97	1.18	1.34	1.61	1.83	2.12	2.4	2.71	3.09
Saudi Arabia	2.89	4.80	6.52	7.94	9.06	10.91	12.95	15.08	16.85	18.78	0.54	0.82	1.13	1.37	1.6	1.72	1.85	1.9	2.05	2.03
UAE	2.16	4.06	5.96	7.75	9.49	11.44	13.23	14.86	16.47	17.88										

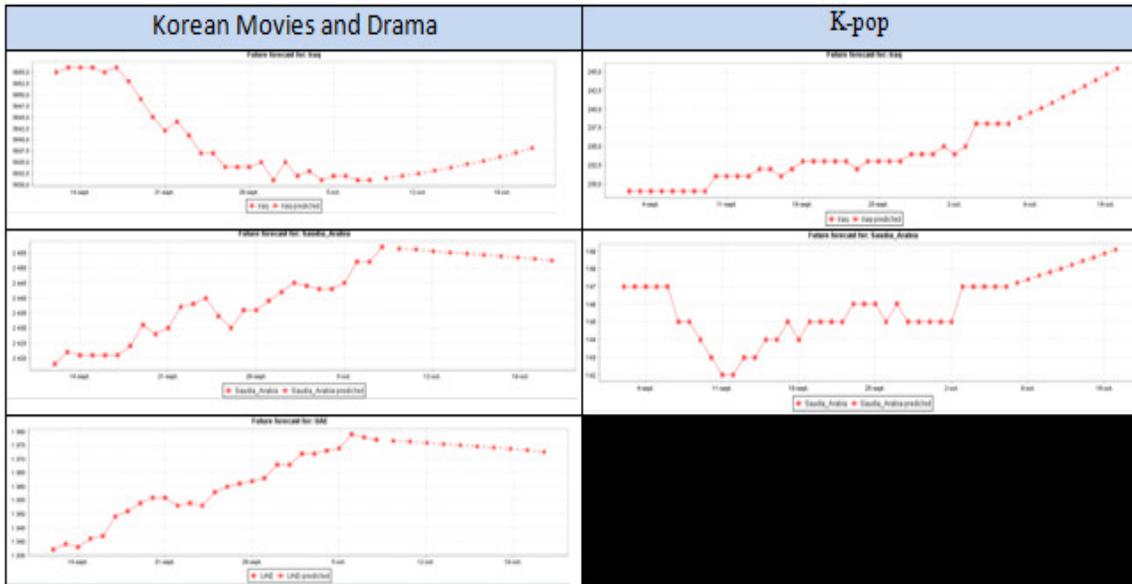
Popularity trends of Korean Movies and K-pop in the North African and Middle Eastern countries are respectively presented in table 3 and 4. On the Y axis are the number of likes and on the X axis are the weekly dates in the last month.

Table 3. Actual and predicted values of likes by North African Countries for Korean Movies and K-pop using SMOreg



For the next ten days, it is obvious that the number of likes for K-pop will increase for all North African countries, however concerning Korean Movies and Drama except Tunisia it is decreasing for Algeria, Egypt and Morocco. Popularity trends of Korean Movies and K-pop in the Middle Eastern countries are presented in the table 4 below

Table 1. Actual and predicted values of likes by Middle Eastern Countries for Korean Movies and K-pop using SMOreg



Similarly to North African, Middle Eastern countries recognize an increase of the number of likes for K-pop. However, concerning Saudi Arabia and United Arab Emirates, the number of likes will decrease for Korean Movies and Drama which is not the case for Iraq.

From the results, we found some interesting findings as follows. First of all, the trend for the number of likes of Korean music, movies and drama shows differently between genres. The number of likes for Korean movies and dramas is decreasing while the number of likes for K-pop is increasing. This result can be explained with time difference. In the beginning, people in Arab countries started to enjoy Korean contents with drama and movie. And then over time they also enjoyed Korean drama's songs (OST : Open Source Track) after they got used to Korean contents by dramas. Second, from the actual and predictive trend, we found that the trend and also speed for the number of likes of Korean contents is totally decreasing while Korean contents became more popular in Arab countries. This result can be explained with the change in consumer behavior. At the beginning with the emergence of Social Media, users were interested to share their experiences about Korean drama and movie through Facebook, but more and more after many websites and applications (such as 'www.myasianTV.se'; www.viki.com ; www.kissasian.com ; www.baykorean.net ; www.dramafever.com; www.dramayou.com ; www.dardarkom.com ; http://kshowonline.com ; http://aradrama.com) appeared to make people enjoy Korean contents directly, users access several websites related to Korean contents straightly without using social networks like Facebook. Lastly, as we noticed in the results, the number of likes of Korean products by Gulf countries are very small comparing to other countries, this is because we believe that Twitter is more popular than Facebook in these countries. In addition, some countries started to broadcast Korean drama through main public channels, for example MBC, main channel of Saudi Arabia spread Korean dramas with Arabic subtitles. This also affects the future trend of Korean contents in Arab countries.

5. CONCLUSION & FUTURE WORK

Based on our analysis and results on predicting popularity of Korean contents in Arab countries, we can find some valuable implications as follows. First, K-contents such as drama, movie and music are sometimes a gateway to a wider interest in Korean culture, food and brands. Korean brands can be inserted into dramas and music videos. South Korea's government long ago embraced pop cultures as a way to transform itself into global market's trendsetter and fuel its economy. Product placement is huge in K-drama. Korean companies' products such as Samsung phones and Hyundai cars make frequent appearances. In 2016, market observers who forecast that one Korean drama "*Descendants of the Sun*" alone will boost the Korean economy by \$261 million, partly by driving demand for tourism and products. Considering this huge effect, prediction on K-contents' consumption in the attractive, emerging markets such as Middle east and North Africa will be various. Increasing the awareness of Korean brand by promoting Korean cultural content will remove entry barrier and create some opportunities for Korean companies who seek to operate a business in Arab countries. Second, Korea entertainment companies can take into consideration the cultural characteristics of the Middle East. There was a long queue to taste Halal certified traditional Korean food, such as bibimbap, and bulgogi and beauty or fashion tips for Arabic women. Considering and depicting these unique points, contents companies can produce distinctive cultural contents to integrate with Arabic culture. Third, this study gives valuable implications for building Korean contents management systems to reduce operating cost and time by not having to detect faults manually or diagnose false warnings. It will be useful for developing predictive model for the whole building Korean contents consumption and also for overcoming liability of foreignness in the global marketplace.

Even though our findings were significant, this study has some limitations. Such limitations and the future direction of the research are as follows. First, we need to study many other features such as demography (age, sex, gender, etc.) comments in order to make more individualized, customized, localized marketing strategies. Second we need to collect comments entered by users and apply sentiment analysis techniques (Chaffar & Inkpen, 2016) in order to analyze their satisfaction about some features of Korean products. Companies, through a business intelligence process, aim to analyze customers' feelings about the products, services, agents and organization. This can lead to the development of new strategies for customers' satisfaction and can provide the company with a competitive advantage in the market. Technologies that automatically recognize unhappy customers can be extremely useful to companies. Third, in this research only Facebook is used to collect the data, we need also to use another social network and specifically Twitter. Facebook is more popular in North African countries however Twitter is more popular in the gulf region. Given this market difference, this study cannot describe the general market situation, and this should be considered in future studies.

ACKNOWLEDGEMENTS

This research was supported by the MISP (Ministry of Science, ICT & Future Planning), Korea, under the National Program for Excellence in SW) (R7719-16-1002) supervised by the IITP (Institute for Information & communications Technology Promotion).

REFERENCES

- [1] Ahn, J., Oh, S. and Kim, H. (2013). Korean pop takes off! Social media strategy of Korean entertainment industry. In: Proceedings on the 10th international conference on service systems and service management, Hong Kong, 17–19 July, pp. 774–777. New York: IEEE.
- [2] Chaffar, S., Inkpen, D. (2016). Using a Generic Text-based Approach for Emotion Prediction. Accepted in International Conference on Computer and Applications, Dubai, UAE.
- [3] Dursun, D. Glenn, W. and Amit, K. (2005) “Predicting breast cancer survivability: a comparison of three data mining methods,” *Artificial Intelligence in Medicine*, vol. 34, pp. 113-127.
- [4] Der-Chiang, L., Che-Jung, C., Chien-Chih, C., Wen-Chih, C. (2012). Forecasting short-term electricity consumption using the adaptive grey-based approach—An Asian case, *Special Issue on Forecasting in Management Science*, Volume 40, Issue 6, pp. 767–773.
- [5] Fan, C., Xiao, F. and Wang S. (2014), Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques, *Applied Energy*, 127 (2014), 1-10.
- [6] Ruiz, E.J., Hristidis, V, Castillo, C., Gionis, A., Jaimes, A. (2012) Correlating Financial Time Series with Micro-Blogging Activity. *ACM International Conference on Web Search and Data Mining (WSDM)*.
- [7] Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting : the forecast package for R. *Journal of Statistical Software* 26(3), 1–22.
- [8] Jin, D.Y. and Yoon, K. (2016), The social mediascape of transnational Korean pop culture : Hallyu 2.0 as spreadable media practice, *New media & society*, 18(7), 2016.
- [9] Khana, I., Capozzolia, A., Corgnata, S. P., Cerquitellib, T. (2013), Fault Detection Analysis of Building Energy Consumption Using Data Mining Technique, *Energy Procedia* 42, pp. 557 – 566.
- [10] Park, G.S. (2013), Manufacturing Creativity : Production, Performance, and Dissemination of K-pop, *Korean Journal* 53(4) : 14-33.
- [11] Park, Y.E. (2015), YG Family, We are One and Number One : based on Eclectic Paradigm (OLI) by J. Dunning, *Korean Academy of International Business*.
- [12] Saigal, S., Mehrotra, D., (2012). Performance comparison of time series data using predictive data mining techniques. *Advances in Information Mining*. 4 (1), pp. 57–66.
- [13] Ristanoski, G., Liu, W. and Bailey, J. (2013) A time-dependent enhanced support vector machine for time series regression, *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. pp. 946-954.
- [14] Saigal, S., Mehrotra, D., (2012). Performance comparison of time series data using predictive data mining techniques. *Advances in Information Mining*. 4 (1), pp. 57–66.
- [15] Sarvestan Soltani A. , Safavi A. A., Parandeh M. N. and Salehi M. (2010) Predicting Breast Cancer Survivability using data mining technique, *Software Technology and Engineering (ICSTE)*, 2nd International Conference, vol.2, pp.227-231.

- [16] Shweta, K. (2012). Using Data Mining Techniques for Diagnosis and Prognosis of Cancer Disease, International Journal of Computer Science, Engineering and Information Technology (IJCEIT), Vol. 2, No. 2.