Computers, Materials & Continua CMC, vol.61, no.1, pp.69-80, 2019

Modeling and Predicting of News Popularity in Social Media Sources

Kemal Akyol[1, *] and Baha Şen[2]

Abstract: The popularity of news, which conveys newsworthy events which occur during day to people, is substantially important for the spectator or audience. People interact with news website and share news links or their opinions. This study uses supervised learning based machine learning techniques in order to predict news popularity in social media sources. These techniques consist of basically two phrases: a) the training data is sent as input to the classifier algorithm, b) the performance of pre-learned algorithm is tested on the testing data. And so, a knowledge discovery from the data is performed. In this context, firstly, twelve datasets from a set of data are obtained within the frame of four categories: Economic, Microsoft, Obama and Palestine. Second, news popularity prediction in social network services is carried out by utilizing Gradient Boosted Trees, Multi-Layer Perceptron and Random Forest learning algorithms. The prediction performances of all algorithms are examined by considering Mean Absolute Error, Root Mean Squared Error and the R-squared evaluation metrics. The results show that most of the models designed by using these algorithms are proved to be applicable for this subject. Consequently, a comprehensive study for the news prediction is presented, using different techniques, drawing conclusions about the performances of algorithms in this study.

Keywords: News popularity, sentiment scores, social network services, Gradient Boosted Machines, Multi-Layer Perceptron, Random Forest.

1 Introduction

News conveys newsworthy events occurring in the course of day to people. News popularity is substantially important so as to predict the spectator or audience for a particular news or journal in modern mining problems [Alswiti and Rodan (2017)]. It is measured through people's interaction with news website. They share links of news or their opinions [Lerman and Ghosh (2010)]. Further, social sharing websites and news websites are used in order to read the various news. Online news popularity examines diverse factors such as sharing count, commenting count and liking count etc. on social media. Online examination of news content, which is a large and still growing market for traditional printed media, has undergone major changes [Canneyt, Leroux, Dhoedt et al. (2018)].

[1] Faculty of Engineering and Architecture, Kastamonu University, 37100, Kastamonu, Turkey.

[2] Faculty of Engineering, Yıldırım Beyazıt University, 06500, Ankara, Turkey.

[*] Corresponding Author: Kemal Akyol. Email: kakyol@kastamonu.edu.tr.

Spread of news to large number of readers within a short period is very important for its popularity. Therefore, there exists a competition among different sources to produce content for a major subset of the population [Bandari, Asur and Huberman (2012)]. Since user behaviors in social media are a reflection of event in the real world, researchers have discovered that they can use it to predict social media and for predictions about the future. Social media data provides an advantage of information acquisition which may be difficult to collect from relatively large acquisitions, large quantities and other sources of data. That is, news popularity can be measured by means of it [Lawrence, Chase, Kyle et al. (2017)]. Evaluation of this subject is relatively novel for researchers.

Some of the studies addressed for this subject are as follows: Alswiti and Rodan examined the effectiveness of feature selection on popularity prediction, by using different features, classification models and attribute ranking models. According to their studies, Random Forest classifier accomplished the best accuracy for all features. J48 and AdaBoost classifiers showed variant sensitivities depending on feature selection [Alswiti and Rodan (2017)]. Canneyt et al. presented a model to predict online news popularity. By analyzing the capture view patterns of online news, they introduced suitable models via well-chosen based functions. By means of actual news dataset, they showed that the combination of the content, meta-data, and the temporal behavior features lead to significantly improved predictions. Gradient Tree Boosting algorithm proves to be more successful for news popularity predicting in their studies [Canneyt, Leroux, Dhoedt et al. (2018)]. Bandari et al. [Bandari, Asur, Huberman et al. (2012)] built a multi-dimensional feature space derived from attributes of articles and evaluated the effect of these features for online article popularity. By using both regression and classification algorithms, they obtained an overall 84% accuracy on Twitter despite randomness in human behavior. Fletcher and Park explored the influence of individual trust on sharing preferences and online news engagement behaviors in news media across eleven countries [Fletcher and Park (2017)]. Anil and Indiramma discussed the importance of recommendation systems, which is useful to find interesting items, different methodologies and social factors [Anil and Indiramma (2015)]. Kywe et al. aimed to analyze the massive information and the huge number of people interacted through Twitter system by utilizing taxonomy [Kywe, Lim and Zhu 2012)]. Keneshloo et al. dealt with the subject popularity, and built models using metadata, content, temporal, and social features. The study was applied to a real data at the Washington Post [Keneshloo, Wang, Han et al. (2016)]. Uddin et al. focused on online news popularity prediction based on sharing the news before publication by using the Gradient Boosting Machine algorithm [Uddin, Patwary, Ahsan et al. (2016)]. Lee et al. [Lee, Moon and Salamatian (2012)] proposed a framework for modelling and predicting the online contents popularity based on survival analysis. The framework infers the likelihood for which the content will be popular. A model was introduced by using a lifetime of content and the comment count popular metrics with a set of explanatory factors. Kümpel et al. reviewed the scientific, peer-reviewed 461 articles quantitatively and qualitatively. The articles dealt with the relationship between news sharing and social medias from the year 2004 to 2014 [Kümpel, Karnowski and Keyling (2015)]. Tatar et al. introduced a valuable study based on user comments. They analyzed the ranking effectiveness of the prediction models online news ranking automatically [Tatar, Antoniadis, Amorim et al. (2014)]. Fernandes et al. [Fernandes, Vinagre and

Cortez (2015)] introduced a proactive intelligent decision support system in order to detect earlier popularity of news information. Random Forest classifier gave the 73% best accuracy on the 39,000 articles which were taken from the Mashable website. Wu and Shen identified the properties of news propagation by tracing the data on Twitter. They implemented a news popularity prediction model that can predict the final number of retweets of a news tweet very quickly by utilizing these characteristics [Wu and Shen (2015)]. Liu and Zhang [Liu and Zhang (2017)] explored that the grammatical construction of titles may affect news popularity positively. They calculated a score of traditional category and author features using logarithmic conversion, and presented a novel methodology in order to predict online news popularity before publication. As it can be seen in these studies, diversified features as input data are used for regression or classification approaches. This study handles out sentiment scores (title and headline), and the number of views in 2 days by interval 20 minutes of news, and presents the news popularity prediction models in social media sources by utilizing the Gradient Boosted Machines (GBM), Multi-Layer Perceptron (MLP) and Random Forest (RF) machine learning algorithms. These algorithms are used in many research areas like medicine, social media and other daily life areas.

The main focus of this study is to carry out the modeling and predicting of news popularity in social media sources. In this context, this study consists of two modules. The first one is to apply the data pre-processing techniques on all datasets. The second one is to demonstrate the performance of boosting, neural networks and ensemble learning based machine learning algorithms. In this context, machine learning algorithms are implemented on the datasets and their performances are discussed in our study.

The rest of the paper is organized as follows. Section 2 presents the materials and methods. Section 3 gives experimental study and results. Finally, the paper ends with conclusions in Section 4.

2 Material and methods

2.1 Data

A set of the data consists of news items and their respective social feedback on multiple platforms: Facebook, Google+ and Linkedln. This set is collected from public end-points of the social media sources that are already anonymized and aggregated by the data owners. News data file concerns the description of news items and consists of 93239 instances and each news item is described by 11 attributes, which are explained in Tab. 1. The data descriptors are based on information obtained by querying the official media sources Google News and Yahoo News [Moniz and Tongo (2018)].

A set of data files so called Feedbacks is concerned with the evolution of news items' popularity in the social media sources, Facebook, Google+ and LinkedIn. News was collected during a two-year period, from January 7, 2013 to January 7 2015, for each of the four categories, Economy, Microsoft, Obama and Palestine. News popularity is measured as the number of views 2 days by interval 20 minutes upon publication simultaneously. This set is composed of 12 data files, for all combinations of these categories and social media sources.

Table 1: Descriptions of attributes in news data file

Variable	Type	Description
IDLink	numeric	Unique descriptor for news
Title	string	News title
Headline	string	Headline of the news
Source	string	Original news
Topic	string	Query topic
PublishDate	timestamp	Date and time for published news
SentimentTitle	numeric	Sentiment score for news items' title
SentimentHeadline	numeric	Sentiment score for news items' headline
Facebook	numeric	News items' popularity value according to Facebook
GooglePlus	numeric	News items' popularity value according to Google+
Linkedln	numeric	News items' popularity value according to Linkedln

The dataset, which includes enormous data, is a pre-processed and re-structured by discarding the instances which include N/A (null) value(s) from datasets. After pre-processing steps, the number of news in these categories is presented in Tab. 2.

Table 2: The number of instances in social media sources

	Economy	Microsoft	Obama	Palestine
Facebook	29928	18531	27015	7690
Google+	32022	19978	27110	7731
Linkedln	32022	19979	27110	7732

2.2 Methods

In this study, modeling and prediction of news popularity in social media sources is performed by using GBM, MLP and RF which are among the popular evolutionary algorithms and experimental results were compared.

Briefly, GBM conducts new models in repeatedly during learning to better predict the target variable. The goal is to create new basic learning models that will have maximum correlation with the negative gradient of the loss function associated with the whole ensemble [Friedman (2001)].

The back-propagated MLP is feed-forward networks updating the weights based on differences between the predicted and actual values for the target variable. The main idea is to minimize the mean square error between the actual and predicted values iteratively [Alpaydin (2010)].

The RF introduced by Breiman is an ensemble learning algorithm created by random decision trees. The main difference of this algorithm from the decision tree is that the RF investigates the best attribute during the division of node while Decision tree investigates

the best feature among the random subsets. Therefore, this algorithm gives better results considering better modeling [Breiman (2001)]. Internal parameters of algorithms and their values were assigned as given in Tab. 3.

Table 3: Internal parameters for algorithms

	Parameter
GBM	Alpha=0.95
	Attribute sampling=all columns
	Mid-point splits=True
	Binary splits=True
	Static random seed=True
MLP	Maximum number of iterations=100
	Number of hidden layers=2
	Number of hidden neurons per layer=100
RF	Tree depth=100

3 Experiments and results

The proposed study consists of two main modules: data processing and machine learning. The first module carries out the prepared steps mentioned Pseudo Code 1 for machine learning module. In addition to the original data retrieved from the social media sources, the pre-processed dataset consists of the sentiment scores information of both the title and headline of the news items. Therefore, the pre-processed datasets are described by 147 attributes (2 sentiment values, title and headline, 144 measurements and outcome variable, the new items' popularity). Flowchart of the proposed study is introduced in Fig. 1.

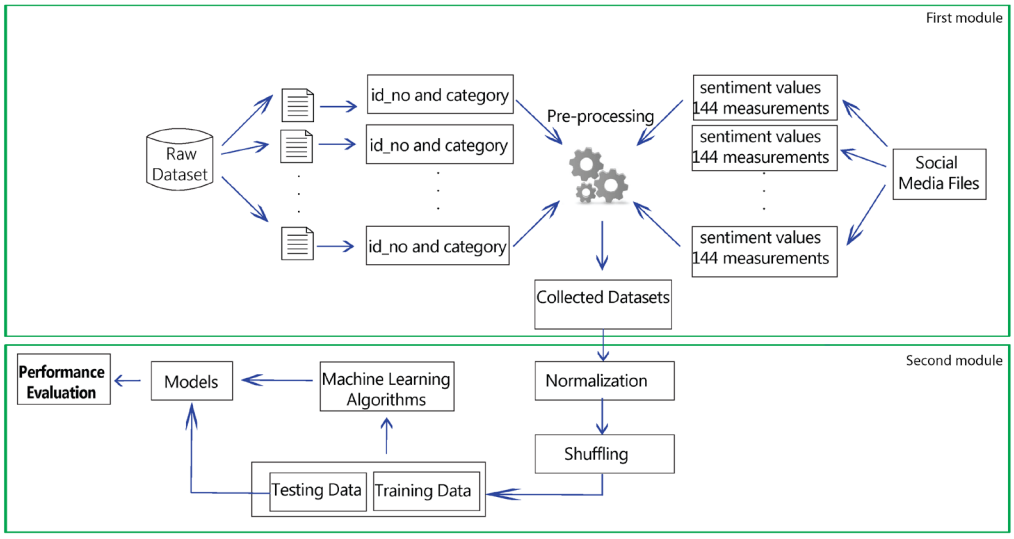

Figure 1: The flow chart of the study

Pseudo Code I. The preparation of datasets.

```
begin
    while (until end of News file)
        read news item
        fetch unique identifier and category attributes
        for (each category)
            for each social media files
                if unique identifier and category in social media file:
                    add to social media warehouse (merge (unique identifier, sentiment values and level of popularity in 20 minute intervals according to social media))
                end
            end
        end
    end
    save as category_social_media file ← social media warehouse
end
```

The information of attributes for these datasets is presented in Tab. 4. All data collection and processing procedures mentioned in these steps are implemented in Python 2.7 on Anaconda platform.

Table 4: The information of attributes for these datasets

Attribute	Explanation
Title	Categoric_social_media (For example, economi_facebook)
SentimentTitle (numeric)	Sentiment score for news items' title
SentimentHeadline (numeric)	Sentiment score for news items' headline
Facebook/GooglePlus/LinkedIn (numeric)	News items' popularity final value according to the social media source.
TS1 .. TS144 (numeric)	Level of popularity between time slice 1 (0-20 minutes upon publication) and 144 (Final level of popularity after 2 days upon publication) are evaluated for the study.

The second module, news popularity prediction, receives the processed data and splits it into training and test sets in order to evaluate the performance of prediction models, GBM, MLP and RF. This module steps mentioned Pseudo Code 2 are executed on 'Knime' platform by integrated Python programming imported from the 'protobuf' library. Python codes could run in a node on this platform. The 'numpy' and 'pandas' libraries are benefited during the build-up of both modules for practicing of the enormous data.

Pseudo Code II. The steps of learning, prediction and evaluation of each model.

```
begin
    for (each social media warehouse as DataFrame)
        DataFrame ←pre-processing (DataFrame)
        DataFrame ←normalization (DataFrame)
        DataFrame ←shuffling (DataFrame)
        Sub-DataFrame 1 and 2 ←partitioning for training and testing (DataFrame)
        model ← learning and prediction (Sub-DataFrame 1 and 2)
        evaluation (model)
    end
end
```

In our study, the performances of the models are evaluated using measures such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and the R-squared coefficient (R^2) to consider how well they are for predictions that match the actual results. These metrics are given by the following equations respectively.

$$MAE = [n^{-1} \sum_{i=1}^{n} |e_i|] \quad (1)$$

$$RMSE = [n^{-1} \sum_{i=1}^{n} |e_i|^2]^{1/2} \quad (2)$$

MAE and RMSE metrics are based on statistical summaries of e_i (i=1,2,...,n). e_i=P_i-O_i is described as individual model prediction error usually. n is the number of data instances, P_i and O_i are the predicted and observed values respectively [Willmott and Matsuura (2005)].

$$R^2 = 1 - \frac{\Sigma(y-\hat{y})^2}{\Sigma(y-\bar{y})^2} \quad (3)$$

where y is the observed response variable, $\bar{y}$ its mean and $\hat{y}$ the corresponding predicted values. R^2 coefficient measures the degree of variation in the target variable. This coefficient is a value between 0 and 1, where 1 equates to a perfect fit of the model [Alexander, Tropsha and Winkler (2015)].

This study focuses on the analysis for the attributes of news data in social media sources and evaluates the performances of RF, GBM and MLP algorithms for news popularity prediction. %70 of data is used as a training set randomly, and remain is considered as the test set. Therefore, firstly the models are trained using the training sets and then tested on the test sets. R^2, MAE and RMSE measures are used so as to evaluate the performances of the models in all experiments. Tabs. 5-8 compares the performance of the models obtained according to Pseudo Code 2 algorithm on the datasets. This module also indicates that sentiment scores of news, and final value of the news items' popularity highly are influential in order to predict news popularity. Sentiment score also known as opinion mining is a field of text mining which examines people' opinions, judgments and ideas about entities [Liu and Zhang (2012)]. The *qdap R* package [Rinker (2013)] is used in order to obtain this score.

Tab. 5 shows the performances of the models on social media sources for Economy dataset. As shown in this table;

a) All algorithms have satisfactory performance on Facebook source for Economy

dataset. Further, MAE measures are same for all models. The maximum R^2 and minimum RMSE measures are obtained with MLP based model on this source.

b) All algorithms have satisfactory performance on Google+ source for Economy dataset. Further, MAE measures are same for all models. The maximum R^2 and minimum RMSE measures are obtained with RF based model on this source.

c) All algorithms have satisfactory performance on Linkedln source for Economy dataset. Further, MAE measure is same for all models. The maximum R^2 and minimum RMSE measures are obtained with RF based model on this source.

Table 5: The performances of the models for Economy dataset

Source media	Performance metrics	Models		
		GBM	MLP	RF
Facebook	R^2	0.585	**0.972**	0.792
	MAE	0	**0**	0
	RMSE	0.005	**0.001**	0.004
Google+	R^2	0.542	0.826	**0.971**
	MAE	0	0	**0**
	RMSE	0.006	0.004	**0.002**
Linkedln	R^2	0.898	0.976	**0.981**
	MAE	0	0	**0**
	RMSE	0.007	0.003	**0.003**

Table 6: The performances of the models for Microsoft dataset

Source media	Performance metrics	Models		
		GBM	MLP	RF
Facebook	R^2	0.793	0.875	**0.947**
	MAE	0	0	**0**
	RMSE	0.006	0.005	**0.003**
Google+	R^2	0.931	**0.994**	0.989
	MAE	0	**0.001**	0
	RMSE	0.005	**0.002**	0.002
Linkedln	R^2	0.567	**0.833**	0.667
	MAE	0	**0**	0
	RMSE	0.01	**0.006**	0.009

Tab. 6 shows the performances of the models on social media sources for Microsoft dataset. As shown in this table;

a) All algorithms have satisfactory performance on Facebook source for Microsoft dataset. Further, MAE measures are same for all models. The maximum R^2 and minimum RMSE measures are obtained with RF based model on this source.

b) All algorithms have satisfactory performance on Google+ source for Microsoft dataset. Further, MAE measures are same for all models. The maximum R^2 and minimum RMSE measures are obtained with MLP based model on this source.

c) All algorithms have satisfactory performance on Linkedln source for Microsoft dataset. Further, MAE measure is same for all models. The maximum R^2 and minimum RMSE measures are obtained with MLP based model on this source.

Tab. 7 shows the performances of the models on social media sources for Obama dataset. As shown in this table; all algorithms have satisfactory performance on Facebook, Google+ and Linkedln sources for Obama dataset. Further, MAE measures are same for all models. The maximum R^2 and minimum RMSE measures are obtained with RF based model on for all sources.

Table 7: The performances of the models for Obama dataset

Source media	Performance metrics	Models		
		GBM	MLP	RF
Facebook	R^2	0.916	0.958	**0.972**
	MAE	0	0	**0**
	RMSE	0.007	0.005	**0.004**
Google+	R^2	0.959	0.955	**0.989**
	MAE	0	0.001	**0**
	RMSE	0.007	0.007	**0.003**
Linkedln	R^2	0.954	0.797	**0.975**
	MAE	0	0	**0**
	RMSE	0.001	0.003	**0.001**

Table 8: The performances of the models for Palestine dataset

Source media	Performance metrics	Models		
		GBM	MLP	RF
Facebook	R^2	0.98	**0.983**	0.957
	MAE	0	**0.001**	0
	RMSE	0.003	**0.003**	0.004
Google+	R^2	0.931	0.959	**0.992**
	MAE	0	0.001	**0**
	RMSE	0.006	0.005	**0.002**
Linkedln	R^2	0.927	**0.975**	0.664
	MAE	0	**0**	0
	RMSE	0.001	**0.001**	0.003

Tab. 8 shows the performances of the models on social media sources for Palestine dataset. As shown in this table;

a) All algorithms have satisfactory performance on Facebook source for Palestine dataset. Further, MAE measures are same for all models. The maximum R^2 and

minimum RMSE measures are obtained with MLP based model on this source.

b) All algorithms have satisfactory performance on Google+ source for Palestine dataset. Further, MAE measures are same for all models. The maximum R^2 and minimum RMSE measures are obtained with RF based model on this source.

c) All algorithms have satisfactory performance on Linkedln source for Palestine dataset. Further, MAE measure is same for all models. The maximum R^2 and minimum RMSE measures are obtained with MLP based model on this source.

Since the datasets used in this study were newly released in February 2018, there is no published study that uses these datasets. But the studies were performed on other datasets based on machine learning because this subject is popular. For this reason, sample studies on the use of machine learning for different datasets are presented in Tab. 9.

Table 9: Sample of studies performed on different datasets

Studies	Method	Data size	Accuracy (%)	R^2	MAE	RMSE
[Alswiti and Rodan (2017)]	Random Forest (100 trees)	39000 records	66.34 %	---	---	---
[Bandari, Asur and Huberman, 2012]	Bagging	44000 records	83.96%	---	---	---
Keneshloo, Wang, Han et al. 2016]	Multi Linear Regression	Twitter data	---	78.2	---	---
Fernandes, Vinagre and Cortez (2015)]	Random Forest	39000 records	67%	---	---	---
Proposed study	Economy dataset (Linkedln)	93239 records	---	0.981	0	0.003
	Microsoft dataset (Google+)		---	0.994	0.001	0.002
	Obama dataset (Google+)		---	0.989	0	0.003
	Palestine dataset (Facebook)		---	0.992	0	0.002

4 Conclusion

News conveys newsworthy events which occur during day to people. News popularity is measured through people's interaction with news website or social media platforms. They cast in their opinions or news links. The scientists use the social media data since it is the reflection of user behaviors in the real world. This study uses a set of the data consisting of news items and their popularity in the social media sources: Facebook, Google+ and LinkedIn. It is composed of 12 data files, for all combinations of the Economy, Microsoft,

Obama and Palestine categories, and the social media sources. The study consists of two phrases which are the preparation of the data and the design of prediction models. The pre-processed datasets are described by 147 attributes (2 sentiment values, title and headline, 144 measurements of popularity in 20-minute intervals for a total of 2 days and outcome variable, the new items' popularity). The prediction models designed by utilizing GBM, MLP and RF learning algorithms are introduced for twelve datasets and empirical tests are performed. The success of most models for each dataset is approximately same. Further, this study will provide a beneficial reference for news popularity prediction.

Acknowledgement: The authors would like to thank the Fernandes et al. [Fernandes, Vinagre and Cortez (2015)] for providing the datasets.

References

Alexander, D. L. J.; Tropsha, A.; Winkler, D. A. (2015): Beware of R^2: simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models. *Journal of Chemical Information and Modeling*, vol. 55, no. 7, pp. 1316-1322.

Alpaydin, E. (2010): *Introduction to Machine Learning*, 2nd Ed. Cambridge, MA, USA: The MIT Press.

Alswiti, W.; Rodan, A. (2017): Features selection effect on predicting the popularity of online news. *Proceedings of the New Trends in Information Technology*.

Anil, R.; Indiramma, M. A. (2015): Survey of personalized recommendation system with user interest in social network. *International Journal of Computer Science and Information Technologies*, vol. 6, no. 1, pp. 413-415.

Bandari, R.; Asur, S.; Huberman, B. A. (2012): The pulse of news in social media: forecasting popularity. *Proceedings of the Sixth International Association for the Advancement of Artificial Intelligence Conference on Weblogs and Social Media*, pp. 26-33.

Breiman, L. (2001): Random forests. *Machine Learn*, vol. 45, no. 1, pp. 5-32.

Canneyt, S. V.; Leroux, P.; Dhoedt, B.; Demeester, T. (2018): Modeling and predicting the popularity of online news based on temporal and content-related features. *Multimedia Tools & Applications*, vol. 77, no. 1, pp. 1409-1436.

Fernandes, K.; Vinagre, P.; Cortez, P. (2015): A proactive intelligent decision support system for predicting the popularity of online news. In: Pereira F, Machado P, Costa E, Cardoso A (eds.), *Progress in Artificial Intelligence*, vol. 9273, pp. 1-12. EPIA Lecture Notes in Computer Science, Springer, Cham.

Fletcher, R.; Park, S. (2017): The impact of trust in the news media on online news consumption and participation. *Digital Journalism*, vol. 5, no. 10, pp. 1281-1299.

Friedman, J. H. (2001): Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*, vol. 29, no. 5, pp. 1189-1232.

Keneshloo, Y.; Wang, S.; Han, E. H.; Ramakrishnan, N. (2016): Predicting the popularity of news articles. *Proceedings of the 2016 Society for Industrial and Applied Mathematics International Conference on Data Mining*, pp. 441-449.

Kümpel, A. S.; Karnowski, V.; Keyling, T. (2015): News sharing in social media: a review of current research on news sharing users, content and networks. *Social Media+Society*, vol. 1, no. 2, pp. 1-14.

Kywe, S. M.; Lim, E. P.; Zhu, F. (2012): A survey of recommender systems in twitter. *SocInfo'12 Proceedings of the 4th International Conference on Social Informatics*, pp. 420-433.

Lawrence, P.; Chase, D.; Kyle, S.; Nathan, H.; Svitlana, V. (2017): Using social media to predict the future: a systematic literature review. arXiv:1706.06134.

Lee, J. G.; Moon, S.; Salamatian, K. (2012): Modeling and predicting the popularity of online contents with cox proportional hazard regression model. *Neurocomputing*, vol. 76, no. 1, pp. 134-145.

Lerman, K.; Ghosh, R. (2010): Information contagion: an empirical study of the spread of news on digg and twitter social networks. *4th International Conference on Weblogs and Social Media*, pp. 90-97.

Liu, B.; Zhang, L. (2017): *Sentiment Analysis and Opinion Mining*. In: Sammut C., Webb G.I. (eds.), *Encyclopedia of Machine Learning and Data Mining*. Springer, Boston, MA.

Liu, C.; Wang, W.; Zhang, Y.; Dong, Y.; He, F. et al. (2017): Predicting the popularity of online news based on multivariate analysis. *IEEE International Conference on Computer and Information Technology*.

Moniz, N.; Torgo, L. (2018): Multi-source social feedback of online news feeds. arXiv:1801.07055.

Rinker, T. W. (2013): *QDAP: Quantitative Discourse Analysis Package*. University at Buffalo/SUNY, Buffalo, New York.

Tatar, A.; Antoniadis, P.; Amorim, M.; Fdida, S. (2014): From popularity prediction to ranking online news. *Social Network Analysis and Mining*, vol. 4, no. 174, pp. 1-12.

Uddin, T.; Patwary, M. J. A.; Ahsan, T.; Alam, M. S. (2016): Predicting the popularity of online news from content metadata. *International Conference on Innovations in Science, Engineering and Technology*.

Willmott, C. J.; Matsuura, K. (2005): Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, vol. 30, no. 1, pp. 79-82.

Wu, B.; Shen, H. (2015): Analyzing and predicting news popularity on Twitter. *International Journal of Information Management*, vol. 35, no. 6, pp. 702-711.