Journal of Engineering Design and Analysis
Volume 3, Issue 2 - 2020, Pg. No. 29-32
Peer Reviewed Journal

Research Article

Fake News Detection on Social Media using Machine Learning Techniques

Bimal Prasad Kar[1], Lakavath Kavitha[2], Sarat Chandra Nayak[3]

[1,2]Department of Computer Science and Engineering, Gandhi Institute for Technological Advancement, Bhubaneswar, Odisha, India.
[3]Department of Computer Science and Engineering, CMR College of Engineering and Technology, Hyderabad, Telangana, India.

I N F O

Corresponding Author:
Bimal Prasad Kar, Gandhi Institute for Technological Advancement, Bhubaneswar, Odisha, India.
E-mail Id:
bimalpk@gmail.com
Orcid Id:
https://orcid.org/0000-0002-2333-9758
How to cite this article:
Kar BK, Kavitha L, Nayak SC. Fake News Detection on Social Media using Machine Learning Techniques. *J Engr Desg Anal* 2020; 3(2): 29-32.

Date of Submission: 2020-11-27
Date of Acceptance: 2020-12-11

A B S T R A C T

Social media has been affecting life style of human being since last two decades by means of providing technologies, high end equipment and free environment for everyone to share their thoughts and ideas and also news. It also provide easy ways to create and propagate news rapidly across the globe. There are possibilities of spreading false news which might misguide the community, thesenews are called as Fake News. With the exponential growth in social media users, framing a central control is quite complex and challenging. Recongnizinga Fake News in Social Media is now a big issue for researchers in this field. Quite a few attempts are found in the literature for systematic recognization of such fake news. Fake news in social media has exclusive features which cannot be found in traditional methods and approaches for reality detection. This work explores the suitability of few popular machine learning techniques such as Naïve Bayes Classifier (NBC), Support Vector Machine (SVM), Linear Regression (LR), Random Forest (RF) and Decision Tree (DT) for detection of fake news in social media data.

Keywords: Machine Learning, Support Vector Machine

Introduction

Fake news is a low quality news with intensionally false information. Fake news is news storiesor hoaxes created to deliberately misinform or deceive readers.Fake news is a misleading news stories that comes from non-reputable sources. The background work in the research domains of rumour detection and virality prediction on social media. We present a state-of-art review of rumour detection on online social media. The research gaps have been identified as issues and challenges within the domain which make it an active and dynamic area of research.

Fake news typically generated for commercial interests to attract viewers and collect the advertising revenue. The term 'fake news' became common parlance for the issue, particularly to describe factually incorrect and misleading articles published mostly for the purpose of making money through page view.Fake news detection is an important and technically challenging problem. In an attempt to tackle the growing misinformation, several fact-checking websites have been deployed to expose the fake news. These websites play a crucial role in clarifying fake news, but they require expert analysis which is time-consuming. Fake news shows negative impact on individual and society. Fake news detection is difficult mainly because there is no governance in place to control over what citizens can read and what carrier they are using to get that particular news nor who is behind that particular news story.

The data science community has responded by taking action

against the problem.There is competition called as the "Fake news challenge" Facebook is employing AI to filter fake news stories out users and feed. Describing factually incorrect & misleading articles published for making money through page views. Fake news is an information which mimic news media content in form but not in original process. With the increasing popularity of social media, more and more people consume news from social media instead of traditional news media.

The people who did not know side story of content they not check for fact to find truth. Solving this problem is not easy. Social media on the rise, news story are reachable & high impact detection is difficult because no governance in place to control over what citizens can read and what carries they are used to get news.Fackbook implementing news system for checking Fake News.

Related Work

In[1] Shlok Gilda (2017) presented concept approximately how NLP is relevant to stumble on fake information. They have used time period frequency-inverse record frequency (TFIDF) of bi-grams and Probabilistic Context Free Grammar (PCFG) detection. They have examined their dataset over more than one class algorithms to find out the great model. They locate that TF-IDF of bi-grams fed right into a Stochastic Gradient Descent model identifies non-credible resources with an accuracy of seventy seven.

In[2] Shivam B, Parikh and Pradeep k, Atrey. 2018, on Media-Rich Fake News Detection: A survey proposed Linguistic feature based methods, Clustering based methods. They have explained about types of data in news, fake news types and various news carrier platforms such as Social Media, Emails and radio services etc.

In[3] Subhadra Gurav, Swati Sase, Supriya Shinde, Prachi Wabale, Sumit Hirve here they followed twitter data for detecting post which are false by using Natural Language Processing (NLP). In this paper, an innovative model for fake news detection using machine learning algorithms has been presented. This model takes news events as an input and based on twitter reviews and classification algorithms it predicts the percentage of news being fake or real.

In[4] Kai Shu,Deepak Mahudeswaran, Huan Liu (Oct 2018) explained about a tool for fake news collection ,detection and visualization using Twitter dataset applied neural network methods to detect news article whether true or false. Fake News Tracker is a tool can automatically collect data for news piece and predicting fake news with visualizations techniques.

In[5] Kai Shu, Amy Sliva , Suhang Wang, Jiliang Tang, and Huan Liu proposed work on various methods used for detecting fake news on social media .In this survey, present a comprehensive review of detecting fake news on social media, including fake news characterizations on psychology and social theories, existing algorithms from a data mining perspective. In these article discussed about characterization and detection techniques for fake news.

In[6] Reema Aswani & Arpan Kumar Kar & P Vigneswar Ilavarasan (2017). Detection of Spammers in Twitter marketing: A Hybrid Approach Using Social Media Analytics and Bio Inspired Computing, detection on spammers in twitter marketing using machine learning such as K-Means levy fight firefly algorithms with chaos for tuning the absorption coefficient. The study thus effectively combines relevant factors from user, descriptive and semantic statistics to model the Twitter profiles for detecting social media spam.

Methodology

Naive Bayes Classifier

In machine learning, Naive Bayes Classifier or family of simple "probabilistic classifiers" based on applying Bayes theorem. The dataset was randomly shuffled and then it was divided into two subsets, Training dataset and testing dataset. Training dataset was used to train the naive Bayes classifier. Test dataset was used to get the results. It is used to predict the likelihood that an event will occur given evidence that's present in ort dataset.

$$\mathrm{P}\left(\frac{A}{B}\right) = \frac{P\left(\frac{B}{A}\right).P(A)}{P(B)}$$

Support Vector Machine

It is a supervised machine learning algorithm which can be used for both classification and regression challenges. In this algorithm we plot each data as a point in n dimensional space. We perform classification by finding the hyper-plane that differentiates the two classes. Measuring accuracy by applying Mean Squared Error (MSE), Mean Absolute Error(MAE), Root Mean Squared Error(RMSE). For support Vector Machine got Accuracy 97%.

$Y = b+\sum \propto (1)\ y(i)*x(i)\ .x$

Where y(i) is the class value of training example x(i), · represents the dot product. The vector x represents a test example and the vectors x(i) are the support vectors, b and $\propto$ (1) are parameters that determine the hyperplane.

Linear Regression

Linear Regression is used to determine linear relationship between independent and dependent variable or in other words, It is used in estimating exactly how much of y will linearly change, when x changes by a certain amount.

$y = mx+c$

Decision Tree

Decision tree can be used for various machine learning

applications. But trees that are grown really deep to learn highly irregular patterns tend to overfit the training sets. Noise in the data may cause the tree to grow in a completely unexpected manner. Random Forests overcome this problem by training multiple decision trees on different subspaces of the feature space at the cost of slightly increased bias. This means that none of the trees in the forest sees the entire training data. The data is recursively split into partitions. At a particular node, the split is done by asking a question on an attribute.

Experimental Result

Dataset is collected from kaggle website and also get from UCI repository. The datasets of every index include 4 attributes: Index, Title, Text and Label.

For experimental analysis we are using Python software, Jupyter Notebook platform and Windows 10 operating system.

Data preprocessing is converting data from raw form to structed or desired form by using Rescaling and Label Encoder of Sklearn library. Label Encoder is used to convert text data into numerical form then Split dataset into training and testing and apply machine learning techniques to predict appropriate value such machine learning models are Support Vector Machine, Naïve Bayes Classifier using various performance metrics such as Mean Absolute Percentage error, Mean Squared Error and Root Mean Squared Error.

Mean Absolute Percentage Error (MAPE) is calculated using the absolute error in each period divided by the observed values that are evident for that period. Then, averaging those fixed percentages. This approach is useful when the size or size of a prediction variable is significant in evaluating the accuracy of a prediction. MAPE indicates how much error in predicting compared with the real value.

$$MAPE = =\sum((y_{actual} - y_{predcited} \div n)) * 100\%$$

Mean Squared Error (MSE) is estimator (of a procedure for estimating an unobserved quantity) measures the average of the squares of the errors that is, the average squared difference between the estimated values and actual values.

$$MSE = 1/n \sum(y_{actual} - y_{predicted})2$$

Root Mean Squared Error (RMSE) is a frequently used measure square root of the difference between values predicted by model and observed values. RMSE is a measure of accuracy ,to compare forecasting error of different models for particular dataset and it is scale dependent

$$RMSE = =\sqrt{\sum}1/N(y_{acutal} - y_{predicted})2$$

Below figure shows various steps for forecasting financial time series data.

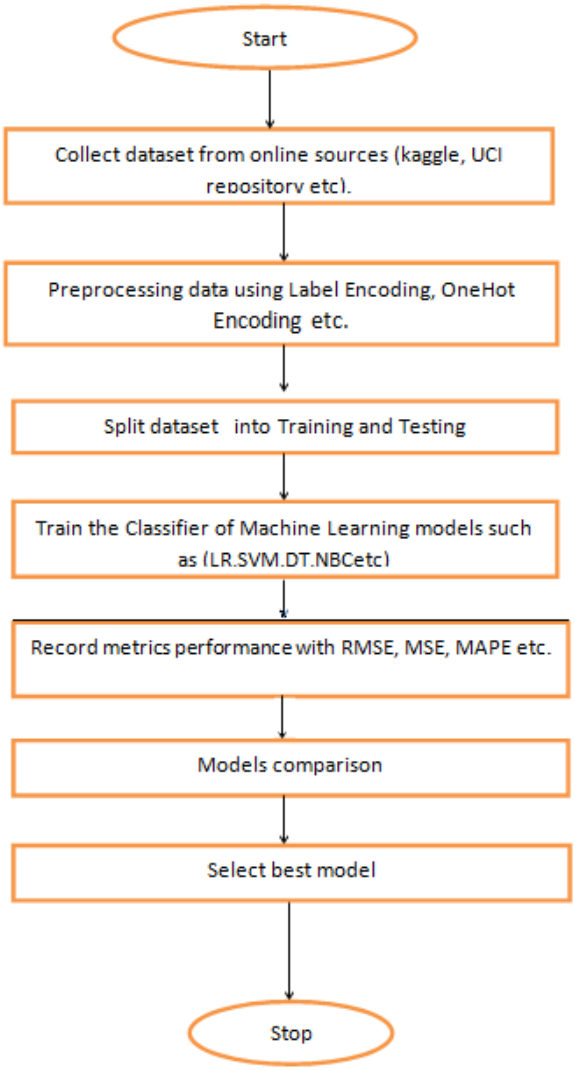

Figure 1.Shows various steps for detecting fake news

Table 1.shows result for various models

Models	MAPE	MSE	RMSE	Accuracy
Linear Regression	4.37	39.32	6.81	83
Naïve Bayes Classifier	4.56	37.86	5.32	82
Support Vector Machine	3.12	24.66	5.11	99
Decision Tree	4.02	25.78	6.02	78

Conclusion

There are possibilities of spreading false news which might misguide the community; these news are called as Fake News. With the exponential growth in social media users,

ISSN: 2582-5607

framing a central control is quite complex and challenging. Recognizing Fake News in Social Media is now a big issue for researchers in this field. This work explores the suitability of few popular machine learning techniques such as Naïve Bayes Classifier (NBC), Support Vector Machine (SVM), Linear Regression (LR), Random Forest (RF) and Decision Tree (DT) for detection of fake news in social media data. The machine learning based classifiers are compared based on metrics such as RMSE, MSE, and MAPE. It has been observed that Support Vector machine has the highest accuracy, i.e. 99.12% and therefore the best classification accuracy.

References

1. Gilda S. Evaluating Machine Learning Algorithms for Fake News Detection. IEEE 15th Student Conference on Research and Development (SCOReD). 2017.
2. Shu K et al. Fake news detection on social media: a data mining perspective. *ACM SIGKDD Explor Newslett* 2017b; 2017b; 19(1): 22-36.
3. Krishnan S, Chen M. Identifying Tweets with Fake News. IEEE International Conference on Information Reuse and Integration for Data Science. 2018.
4. Shu K, Mahudeswaran D, Liu H. FakeNewsTracker: a tool for fake news collection, detection and visualization, Computational and Mathematical Organization Theory. *Springer* 2019; 25(1): 60-71.
5. Shu K, Sliva A, Wang S et al. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter* 2017; 19(1): 22-36.
6. Aswani R, Kumar KA. Vigneswar IP. Exploring content virality in Facebook: A semantic based approach. Forthcoming in Lecture Notes in Computer Science. In Proceedings of 16th IFIP Conference on e-Business, e-Services and e-Society. Springer International Publishing., in press. 2017d.
7. Buzzfeednews: 2017-12-fake-news-top-50. https:// github. com/BuzzFeedNews/2017-12-fake-news-top-50.
8. Maheshwari S. How fake news goes viral: A case study. 2016. [Online]. Available: https://www.nytimes.com/2016/11/20/business/media/how-fake-news-spreads.html.
9. Gupta A, Kaushal. Improving Spam Detection in Online Social Networks. 2015; 978-1-4799-71718/15/$31.00 ©2015 IEEE.
10. Shu K, Wang S, Liu H. Understanding user profiles on social media for fake news detection. In: IEEE conference on multimedia information processing and retrieval (MIPR). 2018b; 430-435.
11. Wang WY. liar, liar pants on fire: a new benchmark dataset for fake news detection. 2017. arXiv:1705.00648.
12. Ruchansky N, Seo S, Liu Y. CSI: a hybrid deep model for fake news detection. In: Proceedings of the 2017 ACM on conference on information and knowledge management. ACM. 2017.
13. Hu X, Tang J, Gao H et al. Social spammer detection with sentiment information. In ICDM'14.
14. Wang S, Tang J, Aggarwal C et al. Linked document embedding for classification. In CIKM'16.
15. https://towardsdatascience.com/full-pipeline-project-python-ai-for-detecting-fake-news-with-nlp-bbb1eec4936d