

FAKE NEWS PREDICTION USING MACHINE LEARNING FOR SOCIAL MEDIA DATASET

Puja Sunil Erande¹, Monika Dhananjay Rokade²

PG Student and Asst. Prof., Department of Computer Engg., SPCOE, Pune, India^{1 2}

pujaerande27@gmail.com¹, monikarokade04@gmail.com²

Abstract: - Fake news are described with an intention to misdirect or to delude the reader. We have presented a response for the task for fake news, individuals are clashing if not by large poor locators of fake news. For this reason new system is generated for fake news identification. The most approaches are used such as supervised machine learning. The result of this project determines the actual fake news detection for social networks using machine learning. Number of peoples having social media accounts such as Facebook, WhatsApp, twitter, etc. This social network is main source of news.

Keywords: - Content modelling, Fake news detector, Fake news categorization, Stance detection, Machine learning, Social media, online fake news, twitter.

I INTRODUCTION

There are a number of people having profiles on social media platforms (SMPs) is growing, thus hiding their identity for malicious purposes. Over the last few years, ONLINE social networks have seen both the number of users and the amount of information shared explosively rise. Users may use these sources of messages to connect, share, discover and disseminate information. Some of those services provide social connections (Facebook and Twitter, for example). Others (Youtube and Flickr, for starters) are used for sharing content. Knowing the actions of consumers at such pages is one of the major research challenges. System Uses Twitter's Social Network as our case study.

II HISTORY AND BACKGROUND

According to [1] the event-based approach based on consumer curiosity used by LeMeNo for News Recommendation. The network of recommendations is focused on both current events and customer expectations. News articles are recommended using machine learning techniques such as grouping related articles, predicting their content, subject similarity, and keyword extraction. The system learns user preferences based on the amount of time spent reading a post, as well as the user-specified rates of interest in different subjects. In this day and age, where there are so many news sources to choose from, it's critical to develop a solution that can guide customers to relevant articles based on their preferences. To increase the likelihood of users recommending a related post, our architecture integrates several approaches to news recommendations.

According to [2] evaluates some of the most Machine learning techniques are commonly used to automatically identify Nepali data, particularly Naive Bayes, SVM and Neural

Networks. The method is being experimented with a self-created Nepali News Corpus with 20 different categories and a total of 4964 posts, gathered online by crawling various national news portals. Functionality dependent on TF-IDF is derived to train and examine the models from the preprocessed documents. The classification pip. According to [3] Social Poisson factorization (SPF), a Probabilistic model incorporating social network information into a standard factorization method; SPF applies to the algorithmic suggestion a social aspect. It provides a robust method to test SPF data and shows that it outperforms rival methods on six datasets in the real world; data sources include a social reader and Etsy.

According to [4] Privacy risks Similar to numerous emerging and influential automation patterns, including internet customization, behavioral profiling and location-based customization. Program analyzes user behaviors about privacy and personalization, as well as technologies that can help reduce the risks to privacy. Program ends with a review that describes risks and technical solutions as well as places at the nexus of personalization and privacy for further study. Such structures will help programmers and analysts place the data protection issues in perspective of solutions when designing customization systems.

According to [12] a new collection of features for automated identification of false news, as well as evaluating the efficiency of existing methods and features in terms of prediction. Our findings show some intriguing details about the utility and significance of features in detecting fake news. Finally, we explore how to apply fake news identification methods in reality, addressing problems and opportunities.

According to [13] the application of hierarchical structure to the classification of a massive, heterogeneous set of

Amharic News Text The method takes advantage of the hierarchical topic structure to break down the classification challenge into a series of smaller problems, one for each classification tree node. An experiment was carried out using categorical data obtained from Ethiopian News Agency (ENA) and SVM to see how the hierarchical classifiers performed on Amharic news text. The results of the experiment show that as the number of classes and documents (features) grows, the accuracy of flat classification decreases. Furthermore, as the number of top features in the feature set grows, the flat classifier's accuracy decreases. When the top three features were used, the flat classifier's accuracy peaked at 68.84 percent. The results of a hierarchical classification experiment show that as we step down the ladder, the classifiers' output improves.

According to [14] a systematic study of identifying false news on social media, including characterizations of fake news based on psychology and social theories, emerging data mining algorithms, measurement metrics, and representative datasets We also talk about relevant research areas, open issues, and potential research directions for social media fake news identification.

III METHODOLOGY

Sentiment analysis is a huge term to classify user's opinion using Natural Language Processing (NLP) and Machine Learning (ML) Approach. Various researchers has done different methods for positive negative classification, aspect base classification, polarity based classification etc. Product review based sentiment analysis is similar to proposed sentiment analysis approach. Most existing approaches methods utilize a preparation set and a test set for arrangement or classification. Training set is made of input feature courses and their corresponding class labels. Utilizing this preparation or training set, an arrangement (classification) model is created which tries to order the input courses into corresponding class names or labels. Then a test set is utilized to confirm the model by deriving the class labels of unknown feature courses. Most existing approaches methods utilize a preparation set and a test set for arrangement or classification. Training set is made of input feature courses and their corresponding class labels. Utilizing this preparation or training set, an arrangement (classification) model is created which tries to order the input courses into corresponding class names or labels. Then a test set is utilized to confirm the model by deriving the class labels of unknown feature courses. A variety of machine learning techniques like Simple Bayes (NB), NLP [4, 5], and Support Course Machines (SVM) [6] are utilized to classification of reviews. Some of the components that can be utilized for semantic classification are Term Absence or Presence, Term Repetition, invalidation, n-grams and Parts of Speech.

IV PROPOSED SYSTEM

Proposed work of project topic:-

A. Data Acquisition:

First of all the information for different Social Media accounts based on certain parameters is extracted from API.

B. Pre- Processing

Then we will apply various pre-processing steps such as lexical analysis, stop word removal, stemming (Porters algorithm), index term selection and data cleaning in order to make our dataset proper.

C. Lexical Analysis

Lexical analysis separates the input alphabet into 1, Word characters (e.g. the letters a-z and 2) Word separators (e.g. space, newline, tab).

D. Stop Word Removal

Stop word removal refers to the removal of words that occur most frequently in documents.

E. Stemming

Stemming replaces all the variants of a word with a single stem word. Variants include plurals, gerund forms (ing forms), third person suffixes, past tense suffixes, etc.).

F. Data Training

We compile artificial as well as real time using online news data and provide training with any machine learning classifier.

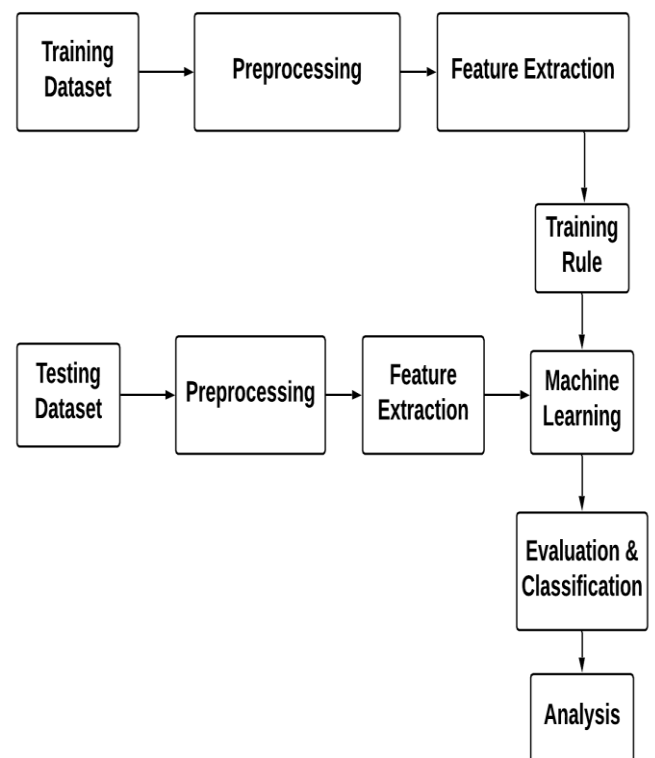


Figure 1: System Architecture

G. Data Testing with Machine Learning

We predict online news using any machine learning classifier, weight calculator for real time or synthetic input data accordingly.

H. Analysis

We demonstrate the accuracy of proposed system and evaluate with other existing systems.

V. RESULT AND DISCUSSION

The results and discussion may be combined into Expected result is as follows:

- System service providers predict news data in social media.
- Classify news types using trained classification, which can automatically predict some predefined class with a news type.
- Performance has been increased with the reduced cost.

VI. CONCLUSION

Fake News detection is done using satisfiability analysis and NP Hard, NP-Complete or P type using modern algebra and relevant mathematical models. The method proposed performed better than those accepted for the three approaches. Using that proposed approach, the accuracy, retrieval, and error of recognition were enhanced. The reason for the move was that it scrapped some redundant functions which did not provide gender separability. The proposed method abused characteristics that were not recognized by the three chosen approaches. The proposed system describes a personalized based news recommendation from social media. The online news population dataset also available on machine learning UCI repository. During the initial research process, the system's output is assessed using this dataset, and accuracy is calculated. However, there is still room for development by introducing a hybrid model that uses a range of feature selection approaches.

ACKNOWLEDGMENT

“M.D. Rokade and S.S. Khatal thanks” for giving their valuable guidance and for providing all the necessary facilities, which where indispensable in the completion of this project.

REFERENCES

- [1] Khandelwal, Dhruv, et al. "LeMeNo: Personalised News Using Machine Learning." 2018 Fourth International Conference on Computing Communication Control and Automation (ICCCAA). IEEE, 2018.
- [2] Manzoor, Syed Ishfaq, and Jimmy Singla. "Fake new detection using machine learning approaches: A systematic review." 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI). IEEE, 2019.

- [3] Shahi, Tej Bahadur, and Ashok Kumar Pant. "Nepali news classification using Naïve Bayes, Support Vector Machines and Neural Networks." 2018 International Conference on Communication information and Computing Technology (ICCICT). IEEE, 2018.

- [4] Chaney, Allison JB, David M. Blei, and Tina Eliassi-Rad. "A probabilistic model for using social networks in personalized item recommendation." Proceedings of the 9th ACM Conference on Recommender Systems. ACM, 2015.

- [5] Toch, Eran, Yang Wang, and Lorrie Faith Cranor. "Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems." User Modeling and User-Adapted Interaction 22.1-2 (2012): 203-220.

- [6] Krishnan, Gokul S., and S. Sowmya Kamath. "Dynamic and temporal user profiling for personalized recommenders using heterogeneous data sources." 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, 2017.

- [7] Vaidya, Nayana, and A. R. Khachane. "Recommender systems-the need of the ecommerce ERA." 2017 International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2017.

- [8] Zhu, Zhiliang, et al. "A Dynamic Personalized News Recommendation System Based on BAP User Profiling Method." IEEE Access 6 (2018): 41068-41078.

- [9] Khater, Shaymaa, Denis Gračanin, and Hicham G. Elmongui. "Personalized recommendation for online social networks information: Personal preferences and location-based community trends." IEEE Transactions on Computational Social Systems 4.3 (2017): 104-120.

- [10]. Liu, Jiahui, Peter Dolan, and Elin Rønby Pedersen. "Personalized news recommendation based on click behavior." Proceedings of the 15th international conference on intelligent user interfaces. ACM, 2010.

- [11]. Liu, Shan, Yao Dong, and Jianping Chai. "Research of personalized news recommendation system based on hybrid collaborative filtering algorithm." 2016 2nd IEEE International Conference on Computer and Communications (ICCC). IEEE, 2016.

- [12]. Monika D. Rokade, Dr. Yogesh Kumar Sharma, "Identification of Malicious Activity for Network Packet using Deep Learning", International Journal of Advanced Science and Technology, 29(9s), 2324 -2331. Retrieved from <http://sersc.org/journals/index.php/IJAST/article/view/14826>.

- [13]. Reis, Julio CS, et al. "Supervised learning for fake news detection." IEEE Intelligent Systems 34.2 (2019): 76-81.

- [14]. A. K. Tegegnie, A. N. Tarekegn, and T. A. Alemu, "A comparative study of flat and hierarchical classification for amharic news text using svm," Culture, vol. 2007, p. 1, 2010.

- [15].K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake news detection on social media: A data mining perspective,” ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, 2017.
- [16]. S. Helmstetter and H. Paulheim, “Weakly supervised learning for fake news detection on Twitter,” Proc. 2018 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Mining, ASONAM 2018, pp. 274–277, 2018.
- [17]. N. Kim, D. Seo, and C. S. Jeong, “FAMOUS: Fake News Detection Model Based on Unified Key Sentence Information,” Proc. IEEE Int. Conf. Softw. Eng. Serv. Sci. ICSESS, vol. 2018–November, pp. 617–620, 2019.
- [18].M. Gahirwal, “Fake News Detection,” International Journal of Advance Research, Ideas and Innovations in Technology, vol. 4, no. 1, pp. 817–819.
- [19]. J. Thorne, M. J. Chen, G. Myrianthous, J. S. Pu, X. X. Wang, and A. Vlachos, Fake news detection using stacked ensemble of classifiers, in Proc. EMNLP Workshop on Natural Language Processing Meets Journalism, Copenhagen, Denmark, 2017.
- [20]. Reis, Julio CS, et al. ”Supervised learning for fake news detection.” IEEE Intelligent Systems 34.2 (2019): 76-81.