

A Project/Dissertation Report
on
ANALYSIS AND DETECTION OF FAKE NEWS

*Submitted in partial fulfillment of the
requirement for the award of the degree
of*

B.TECH(CSE)



(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

**Under The Supervision of
Name of Supervisor :
DR. NARESH KUMAR**

Submitted By
MOHIT MEHTA
19021011684/19SCSE1010513

**SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA
INDIA
MONTH, YEAR**

Table of Contents

Title	Page No.
Abstract	I
List of Table	II
List of Figures	III
Chapter 1 Introduction	1
1.1 Introduction	2
1.2 Formulation of Problem	3
1.2.1 Tool and Technology Used	
Chapter 2 Literature Survey/Project Design	5

Abstract

These days a lot of information is being shared over social media and we are not able to differentiate between which information is fake and which is real. People immediately start expressing their concern or sharing their opinion as soon as they come across a post, without verifying its authenticity. This further results in spreading of it. Fake news and rumors are the most popular forms of false and unauthenticated information and should be detected as soon as possible for avoiding their dramatic consequences. This paper is a review and comprehensive analysis of the articles in recent literature's which were about detecting fake news over social media.

In recent years, due to the booming development of online social networks, fake news for various commercial and political purposes has been appearing in large numbers and widespread in the online world. With deceptive words, online social network users can get infected by these online fake news easily, which has brought about tremendous effects on the offline society already. An important goal in improving the trustworthiness of information in online social networks is to identify the fake news timely. This paper aims at investigating the principles, methodologies and algorithms for detecting fake news articles, creators and subjects from online social networks and evaluating the corresponding performance. This paper addresses the challenges introduced by the unknown characteristics of fake news and diverse connections among news articles, creators and subjects.

These characteristics aforementioned of fake news pose new challenges on the detection task. Besides detecting fake news articles, identifying the fake news creators and subjects will actually be more important, which will help completely eradicate a large number of fake news from the origins in online social networks. Generally, for the news creators, besides the articles written by them, we are also able to retrieve his/her profile information from either the social network website or external knowledge libraries, e.g., Wikipedia or government-internal database, which will provide fundamental complementary information for his/her background check. Meanwhile, for the news subjects, we can also obtain its textual descriptions or other related information, which can be used as the foundations for news subject credibility inference. From a higher-level perspective, the tasks of fake news article, creator and subject detection are highly correlated, since the articles written from a trustworthy person should have a higher credibility, while the person who frequently posting unauthentic information will have a lower credibility on the other hand.

CHAPTER-1

Introduction

Internet is the most life changing inventions of humans and approximately everyone using it on daily basis. There are enormous platforms available to the users each of them caters different services to the users. Any of the users can use the platforms to post or upload anything with or without concern of anyone. As a result, a large number of people uses internet to spread fake news through the available platforms. These fake news serves as a propaganda against an individual, society, organization or any other. A normal person is unable to detect the fakeness of a certain news which he encounters on his electronic device which may lead to misconception and can harm his intellect, emotions and or in other adverse way. So, there is need of machine learning algorithms or classifier to detect the fake news automatically.

Fake News contains misleading information that could be checked. This maintains lie about a certain statistic in a country. World is changing rapidly. No doubt we have a number of advantages of this digital world but it also has its disadvantages as well. There are different issues in this digital world. One of them is fake news. Someone can easily spread a fake news. Fake news is spread to harm the reputation of a person or an organization. It can be a propaganda against someone that can be a political party or an organization. There are different online platforms where the person can spread the fake news. This includes the Facebook, Twitter etc. Machine learning is the part of artificial intelligence that helps in making the systems that can learn and perform different actions. A variety of machine learning algorithms are available that include the supervised, unsupervised, reinforcement machine learning algorithms. The algorithms first have to be trained with a data set called train data set. After the training, these algorithms can be used to perform different tasks. Machine learning is using in different sectors to perform different tasks. Most of the time machine learning algorithms are used for prediction purpose or to detect something that is hidden. Detecting the fake news is a big challenge because it is not an easy task. If the fake news is not detected early then the people can spread it to others and all the people will start believing it. Individuals, organizations or political parties can be affected through the fake news. People opinions and their decisions are affected by the fake news. Different researchers are working for the detection of fake news. The use of Machine learning is proving helpful in this regard. Researchers are using different algorithms to detect the false news. That is why there is a need to detect fake news. The algorithms of machine learning are trained to fulfill this purpose. Machine learning algorithms will detect the fake news automatically once they have trained.

The importance of machine learning to detect fake news will be proved in this literature review. It will also be discussed how machine learning can be used for detecting the false news. Increasing use of internet has made it easy to spread the false news. Different social media platforms can be used to spread fake news to a number of persons. With the share option of these platforms, the news spread in a fast way. Fake news just not only affects an individual but it can also affect an organization or business. So, controlling the fake news is mandatory. A person can know the news is fake only when he knows the complete story of that topic. It is a difficult task because most of the people do not know about the complete story and they just start believing in the fake news without any verification.

Detecting the fake news is one of the most difficult tasks for a human being. The fake news can easily be detected through the use of machine learning. There are different machine learning classifiers that can help in detecting the news is true or false. Nowadays, the dataset can easily be collected to train these classifiers. Different researchers used machine learning classifiers for checking the authenticity of news.

Fake news definition is made of two parts: authenticity and intent. Authenticity means that fake news content false information that can be verified as such, which means that conspiracy theory is not included in fake news as there are difficult to be proven true or false in most cases. The second part, intent, means that the false information has been written with the goal of misleading the reader.

They often have grammatical mistakes. They are often emotionally colored. They often try to affect readers opinion on some topics. Their content is not always true. They often use attention seeking words and news format and click baits. They are too good to be true. Their sources are not genuine most of the times

The spread of fake news has far-reaching consequences like the creation of biased opinions to swaying election outcomes for the benefit of certain candidates. Moreover, spammers use appealing news headlines to generate revenue using advertisements via click- baits.

We aim to provide the user with the ability to classify the news as fake or real and also check the authenticity of the website publishing the news.

The rise of fake news during the 2016 U.S. Presidential Election highlighted not only the dangers of the effects of fake news but also the challenges presented when attempting to separate fake news from real news. Fake news may be a relatively new term but it is not necessarily a new phenomenon. Fake news has technically been around at least since the appearance and popularity of one-sided, partisan newspapers in the 19th century. However, advances in technology and the spread of news through different types of media have increased the spread of fake news today. As such, the effects of fake news have increased exponentially in the recent past and something must be done to prevent this from continuing in the future.

In our modern era where the internet is ubiquitous, everyone relies on various online resources for news. Along with the increase in the use of social media platforms like Facebook, Twitter, etc. news spread rapidly among millions of users within a very short span of time. The spread of fake news has far-reaching consequences like the creation of biased opinions to swaying election outcomes for the benefit of certain candidates. Moreover, spammers use appealing news headlines to generate revenue using advertisements via click-baits. In this paper, we aim to perform binary classification of various news articles available online with the help of concepts pertaining to Artificial Intelligence, Natural Language Processing and Machine Learning. We aim to provide the user with the ability to classify the news as fake or real and also check the authenticity of the website publishing the news.

Fake news is a phenomenon which is having a significant impact on our social life, in particular in the political world. Fake news detection is an emerging research area which is gaining interest but involved some challenges due to the limited number of resources (i.e., datasets, published literature) available.

The social context during news dissemination process on social media forms the inherent tri-relationship, the relationship among publishers, news pieces, and users, which has the potential to improve fake news detection. For example, partisan-biased publishers are more likely to publish fake news, and low-credible users are more likely to share fake news.

The main objective is to detect which is a classic text classification problem with a straight forward proposition (NPL).

It is needed to build a model that can differentiate ‘Real’ news and ‘Fake ‘news.

TOOLS AND TECHNOLOGY USED

Various python libraries:

numpy

pandas

itertools

matplotlib

sklearn

nlTK

The Dataset: The dataset we’ll use is news.csv (from kaggle.com). This dataset has a shape of (7796×4). The first column identifies the news, the second and third are the title and text, and the fourth column has labels denoting whether the news is REAL or FAKE.

CHAPTER-2

Literature Survey/Project Design

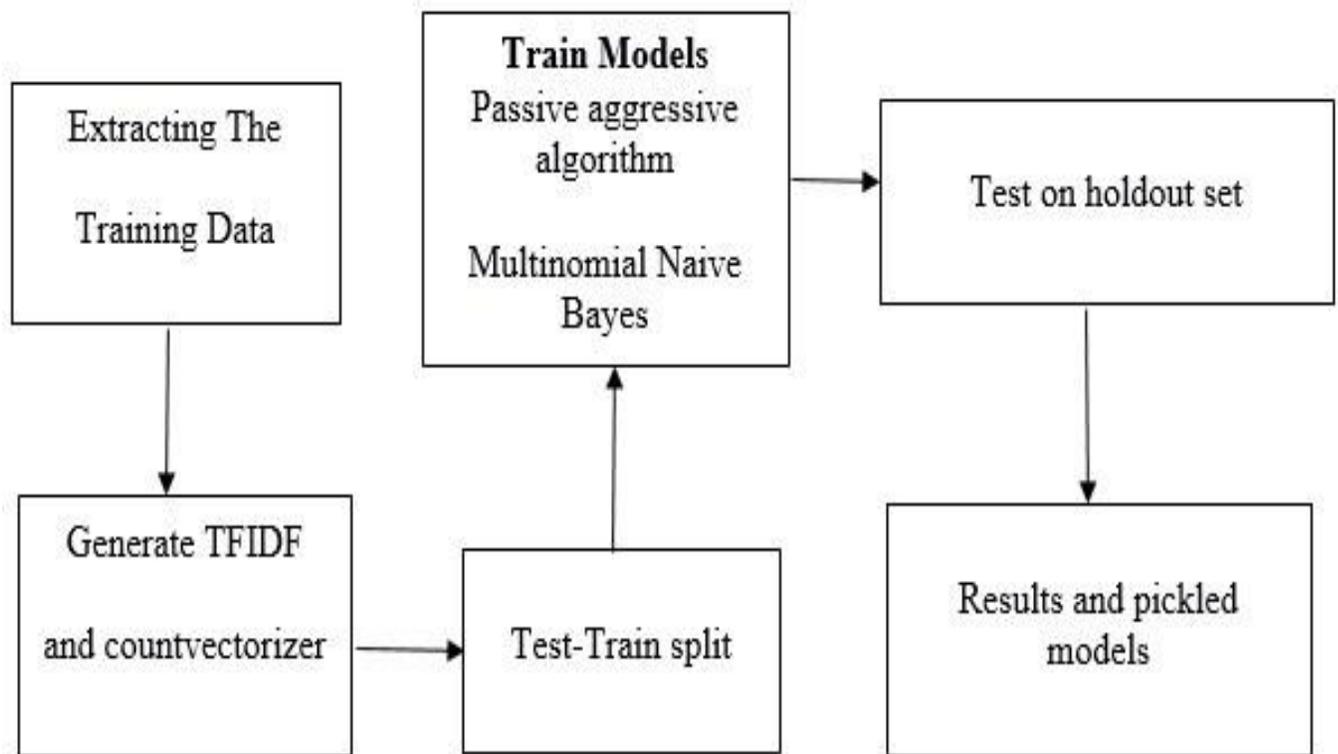
PROPOSED SYSTEM

The actual goal is in developing a model which was the text transformation (count vectorizer vs. tfidf vectorizer) and choosing which type of text to use (headlines vs full text).

Now the next step is to extract the most optimal features for countvectorizer or tfidf-vectorizer, this is done by using a n-number of the most used words, and/or phrases, lower casing or not.

Mainly removing the stop words which are common words such as “the”, “when”, and “there” and only using those words that appear at least a given number of times in a given text dataset. Hence classifying through passive aggressive classifier and finally detection takes place.

Training of the classifiers of machine learning is an important task. This plays an important role for the accuracy of results of these classifiers. A classifier must have to be trained in a proper way with proper data set. Different researchers have trained the machine learning classifiers to detect the fake news. The main problem that occurs while training these classifiers is that mostly the training data set in an imbalanced form

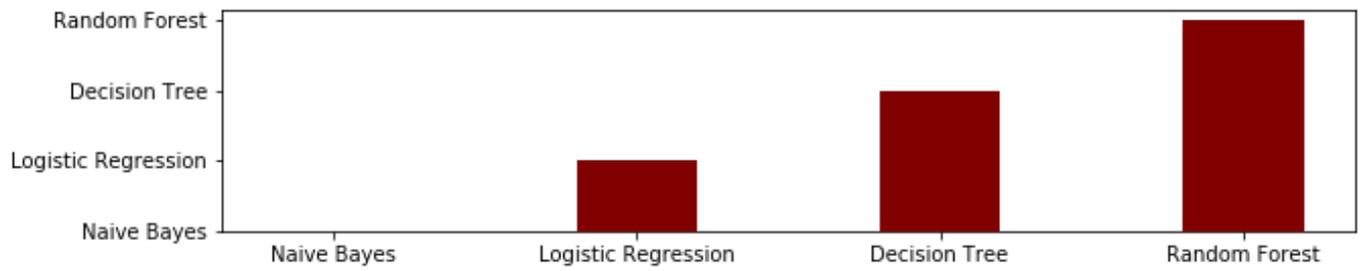
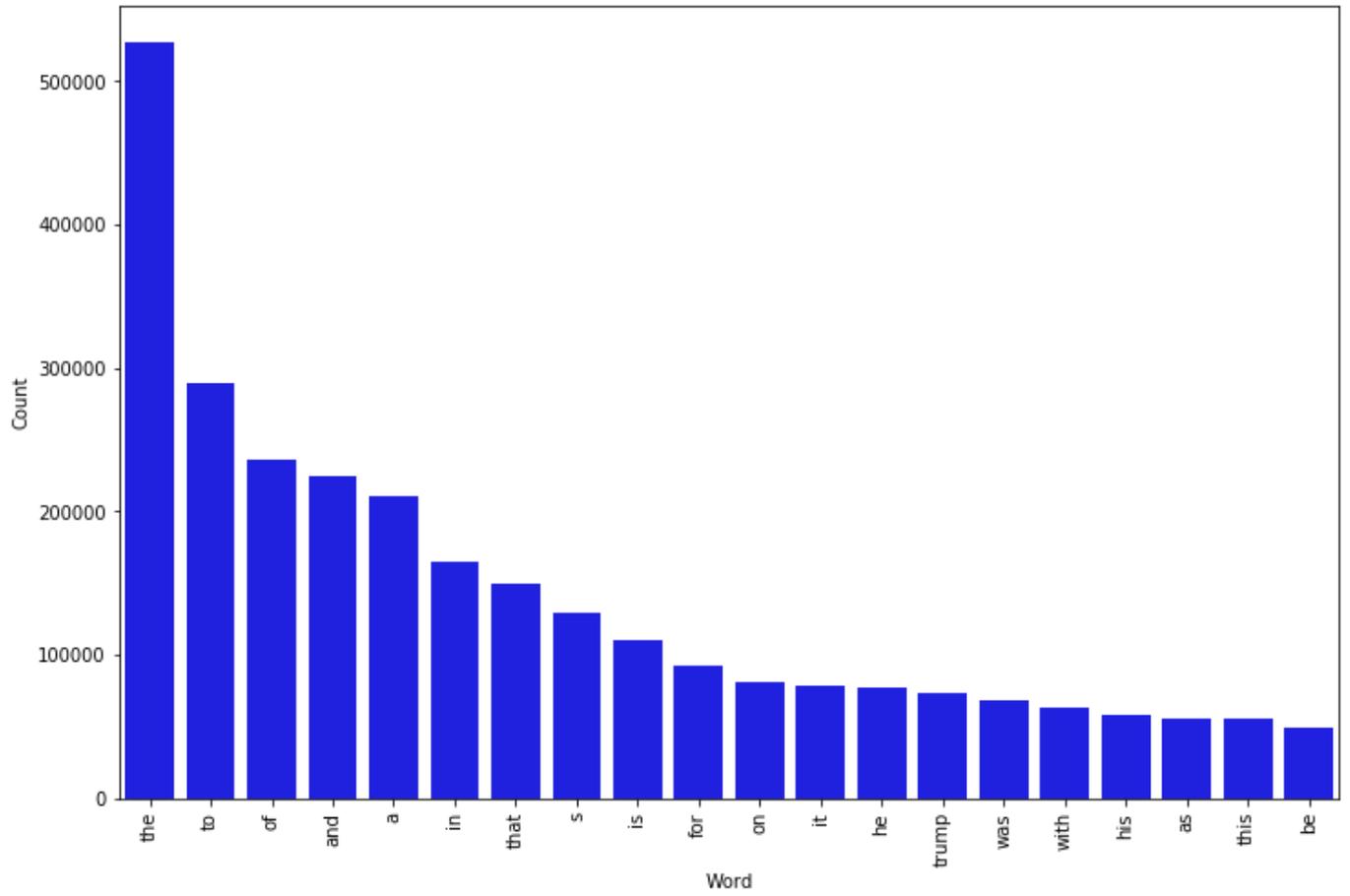


CONCLUSION

Due to increasing use of internet, it is now easy to spread fake news. A huge number of persons are regularly connected with internet and social media platforms. There is no any restriction while posting any news on these platforms. So, some of the people takes the advantage of these platforms and start spreading fake news against the individuals or organizations. This can destroy the reput of an individual or can affect a business. Through fake news, the opinions of the people can also be changed for a political party. There is a need for a way to detect this fake news. Machine learning classifiers are using for different purposes and these can also be used for detecting the fake news. The classifiers are first trained with a data set called training data set. After that, these classifiers can automatically detect fake

Fake news detected with Python by taking political dataset, implemented a TfidfVectorizer, initialized a PassiveAggressiveClassifier, and fit our model.

We ended up obtaining an accuracy of 92.82% in magnitude



References

- Abdullah-All-Tanvir, Mahir, E. M., Akhter S., & Huq, M. R. (2019). Detecting Fake News using Machine Learning and Deep Learning Algorithms. 7th International Conference on Smart Computing & Communications (ICSCC), Sarawak, Malaysia, Malaysia, 2019, pp.1-5, <https://doi.org/10.1109/ICSCC.2019.8843612>
- Ahmed, H., Traore, I., & Saad, S. (2017). Detection of online fake news using n-gram analysis and machine learning techniques. *Proceedings of the International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*, 127–138, Springer, Vancouver, Canada, 2017. https://doi.org/10.1007/978-3-319-69155-8_9
- Ahmed, H., Traoré, I., & Saad, S. (2018). Detecting opinion spams and fake news using text classification. *Secur. Priv.*, 1(1), 1-15. <https://doi.org/10.1002/spy2.9>
- Al Asaad, B., & Erascu, M. (2018). A Tool for Fake News Detection. *2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, Timisoara, Romania, 2018, pp.379-386. <https://doi.org/10.1109/SYNASC.2018.00064>
- Aphiwongsophon, S., & Chongstitvatana, P. (2018). Detecting Fake News with Machine Learning Method. *2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, 528-531. <https://doi.org/10.1109/ECTICon.2018.8620051>
- Della Vedova, M. L., Tacchini, E., Moret, S., Ballarin, G., DiPierro, M., & de Alfaro, L. (2018). *Automatic online fake news detection combining content and social signals*. FRUCT'22: Proceedings of the 22st Conference of Open Innovations Association FRUCT. Pages 272–279. <https://dl.acm.org/doi/10.5555/3266365.3266403>
- Dewey, C. (2016). Facebook has repeatedly trended fake news since firing its human editors. The Washington Post, Oct. 12, 2016.
- Donepudi, P. K. (2019). Automation and Machine Learning in Transforming the Financial Industry. *Asian Business Review*, 9(3), 129-138. <https://doi.org/10.18034/abr.v9i3.494>
- Donepudi, P. K. (2020). Crowdsourced Software Testing: A Timely Opportunity. *Engineering International*, 8(1), 25-30. <https://doi.org/10.18034/ei.v8i1.491>
- Donepudi, P. K., Ahmed, A. A. A., Saha, S. (2020a). Emerging Market Economy (EME) and Artificial Intelligence (AI): Consequences for the Future of Jobs. *Palarch's Journal of Archaeology of Egypt/Egyptology*, 17(6), 5562-5574. <https://archives.palarch.nl/index.php/jae/article/view/1829>
- Donepudi, P. K., Banu, M. H., Khan, W., Neogy, T. K., Asadullah, ABM., Ahmed, A. A. A. (2020b). Artificial Intelligence and Machine Learning in Treasury Management: A Systematic Literature Review. *International Journal of Management*, 11(11), 13-22.
- Granik, M., & Mesyura, V. (2017). Fake news detection using naive Bayes classifier. *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, Kiev, pp.900-903, <https://doi.org/10.1109/UKRCON.2017.8100379>

Jadhav, S. S., & Thepade, S. D. (2019). Fake News Identification and Classification Using DSSM and Improved Recurrent Neural Network Classifier, *Applied Artificial Intelligence*, 33(12), 1058-1068, <https://doi.org/10.1080/08839514.2019.1661579>

Kaliyar, R. K., Goswami, A., Narang, P., & Sinha, S. (2020). FNDNet–A deep convolutional neural network for fake news detection. *Cognitive Systems Research*, 61, 32-44. <https://doi.org/10.1016/j.cogsys.2019.12.005>

Kaur, S., Kumar, P. & Kumaraguru, P. (2020). Automating fake news detection system using multi-level voting model. *Soft Computing*, 24(12), 9049–9069. <https://doi.org/10.1007/s00500-019-04436-y>

Kesarwani, A., Chauhan, S. S., & Nair, A. R. (2020). Fake News Detection on Social Media using K-Nearest Neighbor Classifier. 2020 International Conference on Advances in Computing and Communication Engineering (ICACCE), Las Vegas, NV, USA, pp.1-4, <https://doi.org/10.1109/ICACCE49060.2020.9154997>

Khan, J. Y., Khondaker, M., Islam, T., Iqbal, A., & Afroz, S. (2019). A benchmark study on machine learning methods for fake news detection. *Computation and Language*. <https://arxiv.org/abs/1905.04749>

Kotteti, C. M. M., Dong, X., Li, N., & Qian, L. (2018). Fake news detection enhancement with data imputation. 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech). Athens, 2018, pp.187-192. <https://doi.org/10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00042>

Kudarvalli, H. & Fiaidhi, J.(2020). Experiments on Detecting Fake News using Machine Learning Algorithms. *International Journal of Reliable Information and Assurance*, 8(1), 15-26. <https://dx.doi.org/10.21742/IJRIA.2020.8.1.03>

Kurasinski, L. (2020). Machine Learning explainability in text classification for Fake News detection. Malmö University Publications. <https://urn.kb.se/resolve?urn=urn:nbn:se:mau:diva-20058>

Levin, S. (2017). Pay to sway: report reveals how easy it is to manipulate elections with fake news. *The Guardian*. 13 Jun 2017

Ni, B., Guo, Z., Li, J., & Jiang, M. (2020). Improving Generalizability of Fake News Detection Methods using Propensity Score Matching. *Social and Information Networks*. <https://arxiv.org/abs/2002.00838>

Pratiwi, I. Y. R., Asmara, R. A., & Rahutomo, F. (2017). Study of hoax news detection using naïve bayes classifier in Indonesian language. 2017 11th International Conference on Information & Communication Technology and System (ICTS), Surabaya, pp.73-78. <https://doi.org/10.1109/ICTS.2017.8265649>

Rahman, M. M., Chowdhury, M. R. H. K., Islam, M. A., Tohfa, M. U., Kader, M. A. L., Ahmed, A. A. A., & Donepudi, P. K. (2020). Relationship between Socio-Demographic Characteristics and Job Satisfaction: Evidence from Private Bank Employees. *American*

Journal of Trade and Policy, 7(2), 65-72. <https://doi.org/10.18034/ajtp.v7i2.492>