International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

2020 IJIEMR.Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 2nd Jan 2021. Link
:http://www.ijiemr.org/downloads.php?vol=Volume-09&issue=ISSUE-12

DOI: 10.48047/IJIEMR/V09/I12/149

Title: **DETERMINING FAKE STATEMENTS MADE BY PUBLIC FIGURES BY MEANS OF ARTIFICIAL INTELLIGENCE**
Volume 09, Issue 12, Pages: 867-872
Paper Authors

THATI RAMYATEJA, KOLGURI ANJALI, A.SRAVANTHI, T E.LAXMAN

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

DETERMINING FAKE STATEMENTS MADE BY PUBLIC FIGURES BY MEANS OF ARTIFICIAL INTELLIGENCE

THATI RAMYATEJA[1], KOLGURI ANJALI[2], A.SRAVANTHI[3], T E.LAXMAN[4]

[1,2,3,] B TECH Students, Department of CSE, Princeton Institute of Engineering & Technology For Women, Hyderabad, Telangana, India.

[4] Assistant Professor, Department of CSE, Princeton Institute of Engineering & Technology For Women, Hyderabad, Telangana, India.

Abstract: This paper shows an approach for detecting fake statements made by public figures by means of artificial intelligence. Several approaches were implemented as a software system and tested against a data set of statements. The best achieved result in binary classification problem (true or false statement) is 86%. The results may be improved in several ways that are described in the article as well.

Keywords: fake news; artificial intelligence; deep learning

I. INTRODUCTION

The progress in modern informational technologies brings us to the era where information is as accessible as ever. It is possible to find the answers to the questions we are interested in a matter of seconds. Availability of mobile devices makes it even more convenient for the users. This factor changed the way of how people get the news information a lot. Every mainstream mass media has its own online portal, Face book account, Twitter account etc., so people can access news information really quickly. Unfortunately, the news information that we get is not always true. Paradoxically, the Internet makes it harder to fact check the available information, because there are too many sources that often even contradict each other. All of this caused the emergence of fake news. Mass media and social media have a great influence on a public. There are sides that are interested in using this to achieve their political goals with the help of fake news. They provide false information in form of news to manipulate people in different ways. There exist lots of websites with a single purpose of spreading of false information. They publish fake news, propaganda materials, hoaxes, conspiracy theories in disguise of real news information. The main purpose of fake news websites is to affect the public opinion on certain matters (mostly political). Examples of this may be found in Ukraine, United States of America, Great Britain, Russia and many other countries. Thus, fake news is a global issue and an important challenge to tackle. There is a belief that fake news problem may be solved automatically, without human interference, by means of artificial intelligence. This cause by the rise of deep learning and other artificial intelligence techniques showed us that they can be very effective in solving complex, sometimes even non-formal classification tasks. This article describes a way for classification of short political statements by means of artificial intelligence. Several approaches were implemented and tested on a data set of a statement made by real-life politicians.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

II. Description Of A Data Set Used For Training And Testing

The data set that was used for training and testing was collected by a RAMP studio team. It contains of short statements made by famous public figures. Six possible labels were available for the statement. They are: x ‘Pants on Fire!' (Completely false) x 'False' x 'Mostly False' x 'Half-True' x 'Mostly True' x 'True' Each entry in the data set, besides the statement itself, also contains a lot of metadata. It contains the date when the statement was made, the job of the public figure who made that statement, the source where the statement was taken from, some keywords that characterize the content of the statement and many more other features. The data set consists of 10460 entries in total (7569 of them were provided for training and 2891 for testing). There are more than 2000 different sources of the statements. The RAMP studio team collected the data set using PolitiFact website. The PolitiFact is a project operated by Tampa Bay Times in which reporters from the Times and affiliated media factcheck statements by members of the United States Congress, the White House, lobbyists and interests groups. They publish original statements and their evaluations on the PolitiFact.com website, and assign eacit h a "Truth-O-Meter" rating. PolitiFact.com was awarded the Pulitzer Prize for National Reporting in 2009 for "its fact-checking initiative during the 2008 presidential campaign that used probing reporters and the power of the World Wide Web to examine more than 750 political claims, separating rhetoric from truth to enlighten voters". At some points PolitiFact was criticized by both liberal and conservative wings of American politics, but nevertheless it is a viable source of fact-checked information. This makes a data set useful for creating a system which will classify statements as true or false.

DATA PRE-PROCESSING

Before actually applying the artificial intelligence algorithms to the data, it should be pre-processed. First of all it was decided to use only the statements themselves for classification purposes. This means that none of the metadata provided is used for classification. The classification algorithm might actually be improved in the future by taking into account this metadata. The steps that were used for the pre-processing are the following: x Splitting the statements into separate tokens (words). X Removing all numbers. x Removing all punctuation marks. x Remove all other non-alpha characters x Applying the stemming procedure to the rest of the tokens. In linguistic morphology and information retrieval, stemming (or lemmatization) is the process of reducing inflected or derived words to their word stem, base or root form – generally a written word form. This helps to treat similar words (like “write” and “writing”) as the same words and might be extremely helpful for classification purposes. x Removing stop words. Stop words are the words occur in basically all types of texts. These words are common and they do not really affect the meaning of the textual information, so it might be useful to get rid of them. X Substitution of words with their tf-idf scores. In information retrieval, tf–idf, which is a short for “term frequency–inverse document frequency”, is a numerical statistic measure reflects the importance of a certain word to a document in a collection or corpus. The tf-idf value increases proportionally to the

number of times a word appears in the document and decreases proportionally to the frequency of the word in the corpus, which helps to adjust for the fact that some words appear more frequently in general. According to tf-idf, the weight of a term that occurs in a document is proportional to its frequency, and the specificity of a term can be calculated as an inverse function of the number of documents that contain the specified term.

Implementation of Different Classification Algorithms

Several artificial intelligence algorithms were used for statement classification. All of them are implemented by scikit-learn (a library for Python programming language). For all of the algorithms two different metrics were measured: Classification accuracy based on six categories available x Binary classification accuracy. This metric counts the accuracy as if there were only 2 possible categories for the statement – true (based on the last three categories described above) and false (based on the first three categories described above) For all of the methods the provided data set with known labels was split into training and validation data sets. The training data set was used for the actual process of training of the machine learning models. The validation data set was used for some very basic model tuning. The idea is that having a validation data set we can iteratively tune the machine learning model by repeating the following process: Change a subset of machine learning model meta parameters. Train it on the training data set. Measure its performance on the validation data set. In the end, usually the model, which performed the best on the validation data set, is chosen as a final model. Its performance on the testing data set is considered as an unbalanced estimate of how well the model performs on previously unseen data. A. Classification with logistic regression Logistic regression is a statistical method for analyzing a data set in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes). For the cases when there are more than two labels, the strategy, which is called “One versus all”, is used. In this strategy every category is binary classified against its inverse (a fictional category that states that the example does not belong to the current category). The category with the highest score is picked as a result of a classification. Logistic regression is one of the simplest machine learning techniques. It is easy to implement and easy to interpret. It is usually a good idea to implement logistic regression classifier before proceeding with a more complex approach because it gives you an estimate of how well machine learning algorithms will perform on this specific task. It also helps to eliminate some basic implementation bugs regarding data set treatment. The results that were achieved for logistic regression classifier are the following: x classification accuracy – 72% x binary classification accuracy – 75% B. Classification with naive Bayes classifier In artificial intelligence, naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes theorem with strong (naive) independence assumptions between the features. Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels

are drawn from some finite set. It is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable [1]. Naive Bayes were widely used for e-mail filtering problem. They were invented in the middle of the 90s and they were widely used for classification of e-mails as spam or not spam. Naive Bayes typically use bag of words features to classify texts. Naive Bayes classifiers usually correlate the use of tokens (typically words, or sometimes other constructions, syntactic or not), with the classes that are used for classification, and then apply Bayes theorem to calculate a probability that a text belonging to a certain class. Using naive Bayes classifier is easy to use for both binary and multi-label classification. For the task, described in the paper it is possible to calculate probabilities of the fact that each given statement belongs to the specific group. The results that were achieved for naive Bayes classifier are the following: x classification accuracy – 73% binary classification accuracy – 75% C. Classification with Random Forrest Classifier Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks. Random decision forests consist of number of random decision trees. Each of these random decision trees solves the specified problem independently by their own, and then they "vote" for received results, so the system in general could produce a single result [9]. Random decision forests correct for decision trees' habit of over-fitting to their training set. It is usually beneficial to try a random forest approach for classification tasks. For many tasks, it shows classification accuracy, which is comparable to the accuracy of the most powerful techniques available, while a period of time it takes to train a random forest model is usually much shorter. The results that were achieved for random forest classifier are the following: classification accuracy – 76% binary classification accuracy – 81% D. Classification with support vector machines In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. A support vector machine model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall. In the classification tasks for the cases when there are more than two labels, "One versus all" strategy is used (similarly to logistic regression). Currently support vector machines are not as popular as they used to be (mostly because of the rise of deep learning algorithms), but they are still very useful for some classification problems. The results that were achieved for support vector machines classifier are the following: x classification accuracy – 79% x binary classification accuracy – 83% E. Classification with deep neural networks Artificial neural networks are computing that were inspired by biological neural networks of the animals' brains (though many scientists believe that actual brains are much more complex systems than artificial neural networks – it has much more units, signals are transferred differently etc). The artificial neural networks consist of units (generally grouped by layers) and connections between them. Each of these

connections has a corresponding weight, which are modified during learning process. There is no formal definition of deep neural network, but usually it is assumed that neural network is deep if it has more than one hidden layer (not input or output layer). Deep neural networks are very popular right now and they show tremendous results in lots of fields. They are also very well suited for classification problems. The results that were achieved for deep neural network classifier the following: classification accuracy – 81% x binary classification accuracy – 86% F. Comparative analysis of the results of all the methods Summarized classification results that were achieved. As one can see, deep neural network shows the best results both in classification accuracy and in binary classification accuracy. It beats the nearest competitor (which turned out to be a support vector machines model) by approximately 3% in classification accuracy and by approximately 2% in binary classification accuracy. This is not surprising, because recent developments in deep neural network area show that it is really well suited for similar classification tasks. The difference in performance between the simplest model (logistic regression model) and the most complex model (deep neural network model) is quite significant. It looks like the general trend for this task is the more complex the model is the better result it shows.

Ways to Improve the Classification Results

There are several ways to improve classification results that we would like to point out: x Include Meta data to the training process. Nothing but the text of the statement were used for making a prediction in the algorithms described above, but it seems promising to use some other available information x Get more data and use this data for more extensive training. The data set that was used for training is quite small which possibly affected the results of classification negatively. Tune the trained model more. For example, for neural network it is possible to change the number of units in each hidden layer, the number of hidden layers them etc. x Investigate the examples that are misclassified in the validation data set. Some of their features might be useful for building a better machine learning model. Use ensembles of different machine learning algorithms. It is possible to join implemented algorithms to a single system, that takes into account verdicts of all of the algorithms and outputs a classification decision based on that. Such systems usually perform better on lots of classification tasks. Try other artificial intelligence approaches. It is important to try out each of the suggested improvements, as they all look quite promising. This should be a subject of future research.

CONCLUSIONS

In this paper, several algorithms for classifying statements made by public figures were implemented. Unsurprisingly, deep neural networks showed the best results both in classification accuracy based on six categories and binary classification. This encourages future research with extensive usage of deep neural networks. Achieved results might be significantly improved. It is possible to both improve the data which is used for training as well as the machine learning models themselves. This might be a subject for future research. Together with the text summarization (the problem that also can be solved by means of artificial intelligence), this approach might be

used for classification of news articles as fake or true. This might also be a subject for future research.

REFERENCES

[1] M. Granik, V. Mesyura, "Fake news detection using naive Bayes classifier," 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kiev, 2017, pp. 900-903.

[2] Cade Metz. (2016, Dec. 16). "The bittersweet sweepstakes to build an AI that destroys fake news". Available: https://www.wired.com/2016/12/bittersweet-sweepstakes-build-aidestroys-fake-news/

[3] Fake news RAMP: classify statements of public figures. (n.d.) [Online]. Available: https://www.ramp.studio/problems/fake_news

[4] The Principles of the Truth-O-Meter: PolitiFact's methodology for independent fact-checking. (2018, Feb. 12) http://www.politifact.com/truth-o-meter/article/2018/feb/12/principlestruth-o-meter-politifacts-methodology-i/. Accessed Mar. 24, 2018.

[5] Rajaraman, A.; Ullman, J. D. "Data Mining". (2011) Available: http://i.stanford.edu/~ullman/mmds/ch1.pdf. Accessed Mar. 24, 2018.

[6] Stemming and lemmatization. (n.d.) Available: https://nlp.stanford.edu/IR-book/html/htmledition/stemming-andlemmatization-1.html. Accessed Mar. 24, 2018

[7] Sparck Jones, K. (1972). "A Statistical Interpretation of Term Specificity and Its Application in Retrieval". Journal of Documentation, vol 28, pp 11–21.

[8] David A. Freedman (2009). "Statistical Models: Theory and Practice". Cambridge University Press. p. 128.

[9] Ho, Tin Kam (1995). "Random Decision Forests". Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995. pp. 278–282.

[10] Cortes, Corinna; Vapnik, Vladimir N. (1995). "Support-vector networks". Machine Learning, vol 20, pp 273–297.