Determining Fake Statement Made Public Figures by Means Artificial Intelligences

[1]Mr. Vishal Shinde, [2]Mr.Roshan Dandkar, [3]Miss.Sonal Vanave, [4]Miss.Dipti Vide

[1]Asst.Professor,[2,3,4]UG Student,[1,2,3,4]Computer Engg. Dept. Shivajirao S. Jondhle College of Engineering & Technology, Asangaon, Maharashtra, India.

[1]mailme.vishalshinde@gmail.com,[2]roshandandkar6@gmail.com,[3]sonal29vanve@gmail.com,[4]diptivide@gmail.com

Abstract-The paper shows an approach for detecting fake statements made by public figures by means of Artificial Intelligence. Fake news problem may be solved automatically, without human interference, by means of artificial intelligence. This project describes a way for classification of short political statements by means of artificial intelligence. Several artificial intelligence algorithms were used for statement classification. i.e. Logistic regression, Naive Bayes classifier, Random forest classifier, Support vector machines, Deep neural networks, etc. Before actually applying the artificial intelligence algorithms to the data it should be pre-processed.

Keywords: Artificial intelligence, cosmonaut training, automatic information system, information support

I. INTRODUCTION

Every social platform has its own online portal, Facebook account, Twitter account etc., people can get news information very quickly. Unluckily that news information is not always true. Internet check the available information, because there are too many sources that often even contradict each other. All of this lead to the appearance of fake news. Social media and mass media have an great impact on a public. Using social media platform politician spared fake news to achieve their political goal.

This paper 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 Mass media and social media have a great influence on a public. They provide false information in form of news to handle people in different ways. [1]There exist lots of websites with a single purpose of grow of false information. They publish fake news, propaganda materials, hoaxes, conspiracy theories in disguise of a real news information. The main purpose of fake news websites is to affect the public opinion on certain matters (mostly political). 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.It contains of short statements made popular by public figures. Six possible labels were available for the statement.

They are: Pants on Fire (completely false)
1. False
2. Mostly False
3. Half-True
4. Mostly True
5. True

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 Political website. The Political is a project operated by Tampa Bay Times in which reporters from the Times and affiliated media fact check statements by members of the United States Congress, the White House, lobbyists and interests groups.

II. AIMS AND OBJECTIVE

a) Aim

The aim of making this system is for detecting fake statements made by public figure. Several approaches were implemented as a software system and tested against a data set of statement identify intentionally deceptive

news. Facebook, Twitter and other social media sites have been stepping up their efforts to combat the spread of fake news on their platforms this project classify that fake news.

b) Objective

Input style is that the method of changing a user-position description of the input into a computer-based system.

This method is vital to avoid error within the knowledge input method and show the proper way to the management for secure correct info from the processed system.

III. LITERATURE SURVEY

Paper 1: A fake news detection using naive Bayes classifier M. Granik, V. Mesyura 2017 IEEE First Ukraine Conference on …, 2017 - ieeexplore.ieee.org

This paper shows a simple approach for fake news detection using naive Bayes classifier. This approach was implemented as a software system and tested against a data set of Facebook news posts. We achieved classification accuracy of approximately 74% on the test set which is a decent result considering the relative simplicity of the model.

Paper 2: The principles of the truth-o-meter: politifact's methodology for independent fact-checking

Fact-checking journalism is the heart of PolitiFact. Our core principles are independence, transparency, fairness, thorough reporting and clear writing. The reason we publish is to give citizens the information they need to govern themselves in a democracy. Since our launch in 2007, we've received many questions about how we choose facts to check, how we stay nonpartisan, how we go about fact-checking and other topics. This document attempts to answer those questions and many more.

Paper 3: A statistical interpretation of term specificity and its application in retrieval

The exhaustively of document descriptions and the specificity of index terms are usually regarded as independent. It is suggested that specificity should be interpreted statistically, as a function of term use rather than of term meaning. The effects on retrieval of variations in term specificity are examined, experiments with three test collections showing, in particular, that frequently-occurring terms are required for good overall performance.

Paper 4: Fake News Detection on Social Media: A Data Mining Perspective

This Paper introduces a detailed analysis of the crawled PolitiFact network dataset, which can provide necessary motivations and foundations for our proposed model to be introduced in the next section. The data analysis in this section includes 4 main parts: article credibility analysis with textual content, creator credibility analysis, creator-article publishing historical records, as well as subject credibility analysis, and the results are illustrated

IV. EXISTING SYSTEM

This system fake news detection approaches from a data mining perspective, including feature extraction and model construction. This system also further discussed the data sets, evaluation metrics, and promising future directions in fake news detection research and expand the field to other applications when two enemy politicians cite statistics on privation, with one saying poverty is up while the different says poverty is falling, both can be fair when you look at complete versus relative numbers. Automated fact-checking can look over repeated declare, document where they own been cited before and whether there is proof to back up the statements, as signify here allowing key term and look over from speeches and statements to be checked in opposition to the facts. But it not like just classifying whole pages as either 'true' or 'false' but what it can do in specific cases is does not give proper result to the people details they need to make their personal minds up when they need it,"[4]

V. COMPARTIVE STUDY

Sr.no	Paper name	Author name	Technology	Advantage	Disadvantages
1.	A fake news detection using naive bays	M.garnik, V.mesyura	Naive bays algorithm	This system has better scalability & accuracy	Unsuitable for implementing complex algorithm
2.	The principle of truth-o-meter	Angie Drobnic Holan	Politifact's methodology	Low cost easy access	The system can not properly Analyzed
3.	statistical interpretation of term specify and it is application in retrieval	Sparck jones, k	Machine & deep learning approach	The system capable analyzing keyword	Limited presentation capacity
4.	Fake News Detection on Social Media: A Data Mining Perspective	KShu, A Sliva, S Wang, J Tang, H Liu	Data mining Algorithm	Easy to implement	It not give proper result to the people details they need to make their personal minds up when they need it

VI. PROBLEM STATEMENT

Social media for news utilization is a double-edged blade. On the one fist, its little cost,easy access, and fast dissemination of detail lead people to seek out and ingest news from social media. On the different hand, it enables the broad spread of "fake news", i.e., little quality news with on purpose false details. Therefore, fake news examination on social media has currently become an emerging research that is captivate tremendous attention. Fake news detection on social media give unique property and challenges that make existing notice algorithms from accepted news media misfire or not applicable.[2]

VII. PROPOSED SYSTEM

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. Splitting the statements into separate tokens (words). Removing all numbers. Removing all punctuation marks. Remove all other non-alpha characters 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, skew y are used. System will also show the accuracy of the signature and for verification it will show the result whether the signature is genuine or forged. Random forest classifier and neural network is used in proposed system.[1]

VIII. ALGORITHM

The general idea of working of proposed system algorithm is given as follow:

Algorithms:
Algorithm – SVM
Step 1: Data Preprocessing
1. Import Dataset or add used already stored dataset values
2. Extract Independent and dependent Variable from the dataset
3. Split dataset into training and testing set

Step-2: Create a Support vector classifier

#classifier = SVC(kernel='linear', random_state=0)

we have used **kernel='linear'**, as here we are creating SVM for linearly separable data

Step-3: Predicting the test result
1. Model is first fitted to the training set, for predicting the test result from the available dataset.
2. #y_prediction= classifier.predict(test_data)
3. Above prediction vector and test set real vector can be used to determine the incorrect predictions done by the classifier.

Step-4: Repeat Step 1 & 2.

Step-5: Segregate the data elements into minimum identified sub classes with best matching.

Algorithm 2: : Decision Tree Algorithm – Supervised Learning Classification & Regression Algorithm

Step-1: Begin the tree with the root node, says S, which contains the complete dataset.

Step-2: Find the best attribute in the dataset using **Attribute Selection Measure (ASM).**

Step-3: Divide the S into subsets that contains possible values for the best attributes.

Step-4: Generate the decision tree node, which contains the best attribute.

Step-5: Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

Random Forest Classifier

Step 1: Data Preprocessing
1. Split the dataset into training and test set (train_data & test_data).
2. Select random N data points from the training set.
3. Extract Independent and dependent Variable from the training dataset
4. Create Feature class based on similar subset variables

Step-2: Build the decision trees associated   with the selected data points (Subsets) using RF Classifier as given below.
#classifier= RandomForestClassifier(n_estimators= x, crite rion="entropy")
classifier.fit(x_train, y_train)
Where,

*   **n_estimators=** The required number of trees in the Random Forest. The default value is 10. We can choose any number but need to take care of the overfitting issue.

*   **criterion=** It is a function to analyze the accuracy of the split. Here we have taken "entropy" for the information gain.

Step-3: Predicting the test result

1. Model is first fitted to the training set, for predicting the test result from the available dataset.

   # x_prediction= classifier.predict(train_data)

   #y_prediction= classifier.predict(test_data)

2. Above prediction vector and test set real vector can be used to determine the incorrect predictions done by the

Step-4: Repeat Step 1 & 2.

Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that fits into the larger subset.

IX. MATHEMATICAL MODEL

The formula for calculating the conditional probability of the fact, that news article is fake given that it contains some specific word looks as following:

$Pr(F|W)=Pr(W|F)·Pr(F)/(Pr(W|F)·Pr(F)+Pr(W|T)·Pr(T))$

$Pr (F|W)$ –that a news article is fake given that word W appears in it;

$Pr (W|F)$ – conditional probability of finding word W in fake news articles;

$Pr (F)$ – overall probability that given news article is fake news article;

$Pr (W|T)$ – conditional probability of finding word W in true news articles;

$Pr (T)$ – overall probability that given news article is true news article;

$p1 = Pr (F|W1) ·...·Pr (Fawns),………. \quad (1)$

$p2 = (1 − Pr (F|W1))(1 − Pr (Fawns))…..(2)$

$p = p1/ (p1 +p2),………………………..(3)$

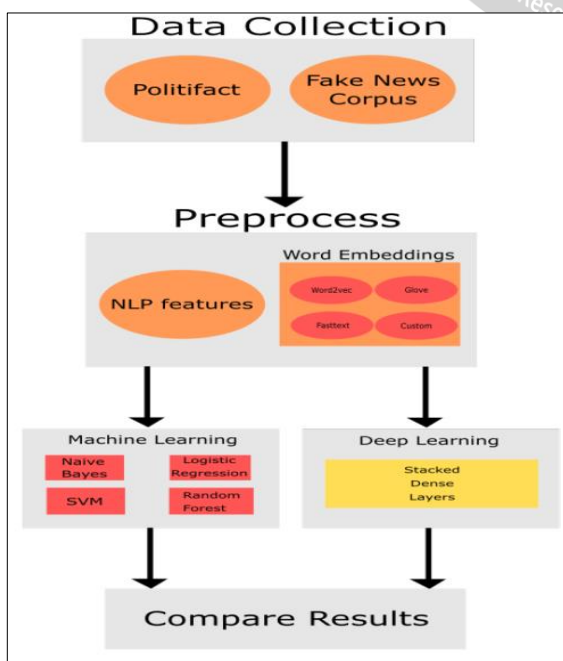## X. SYSTEM ARCHITECTURE

Fig.1: System Architecture

This software is very simple to use as it made in python and hence the GUI is very less complex. The main purpose of this software is to provide services to user and predict the fake statements. The GUI is so simple that can understand easily. The scope of this research

is to intend to build up a structure reached out to other undertaking, for example. Characterization of an extensive number of comparable articles which require expert information there are numerous such difficulties including newspaper ,articles, Facebook ,twitter ,Instagram. User services without leakage of user live location.
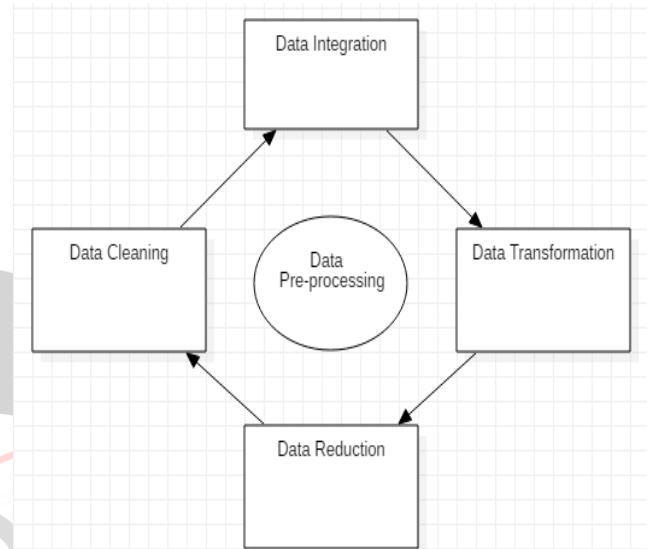

Fig.2: The Architectural Overview of the System

XI. ADVANATGES

1. Detecting number of statement true or false.
2. Detecting number of statement created in given time interval.
3. Easily acquired
4. Widely Accepted
5. Non-intrusive
6. Widespread use
7. Resistance to forgery
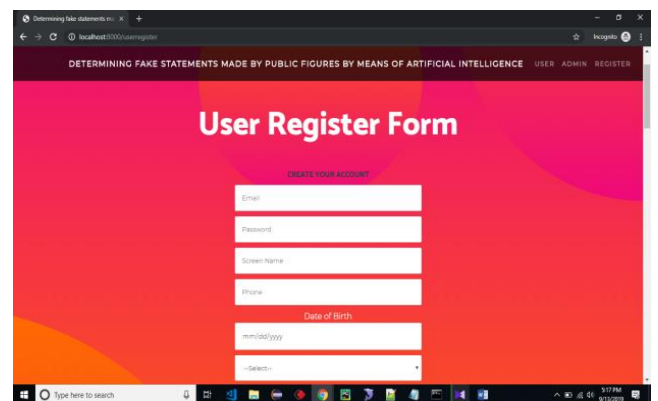
XII. DESIGN DETAILS
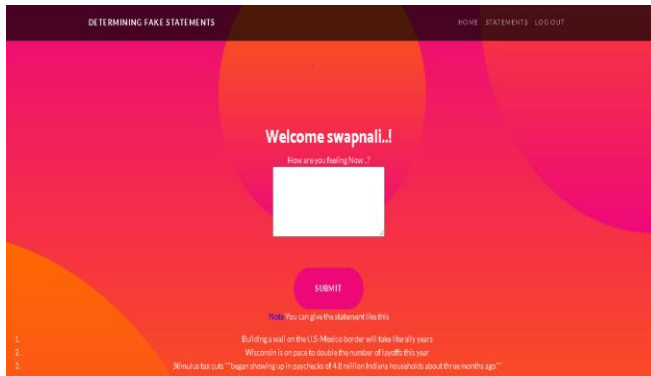

Fig.1: User Registration Form

Fig2: User Checking Statement

XIII. CONCLUSION

Thus, We have tried to implement the paper "Determining fake statements made by public figures by means of artificial intelligence", *"M Granik, V Mesyura, A Yarovyi"* 2018 IEEE and according that implementation the conclusion is as follows: qresults both in classification and 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. Together with the text summarization (the problem that also can be solved by means of artificial intelligence), this approach might be used forclassification of news articles as fake or true.

REFERENCE

[1] M. Granik, V. Mesyura, "Fake news detection using naive Bayes classifier," 2018 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kiev, 2018, pp. 900-903.

[2] Artificial Neural Networks as Models of Neural Information Processing. Frontiers Research Topic (n.d.).https://www.frontiersin.org/researchtopics/ 4817/artificial-neural-networks-as-models-of-neural-informationprocessing. Accessed Mar. 24, 2018.

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

[4]fake news detection on social media: A data mining perspective K Shu, ASliva, S Wang, J Tang, H Liu - ACM SIGKDD Explorations …, 2017 dl.acm.org

[5] 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/

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

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