

THE FAKE NEWS DETECTION USING MACHINE LEARNING ALGORITHMS

Dr. C K GOMATHY¹, Ms. C.V.S.VASAVI², Mr. D.Y.V.RAJESH³, Ms. A.SRIJA⁴

ABSTRACT: This project comes up with the applications of Natural Processing Language (NLP) to detect the true and fake news .These techniques are used for detecting the fake news happening from some trusted and untrusted platforms/non-trustable sources these days. These models are built using Various Machine Learning techniques to calculate the frequency and count of news from a set. We will use several news platforms like twitter and face book for analyzing the news to detect them whether true or fake.

Keywords: NLP, Fake News, True News, Logistic regression, Decision Tree.

I. INTRODUCTION

Now a day's fake news is creating different issues from sarcastic articles to fabricated news and government propaganda in some outlets. Even this fake news was destroying countries and one's individual peace and harmony and even some families are being destroyed through this.

The data which was collected might contain missing values that may lead to inconsistency .To gain better values the data must be in the media are growing problems with huge ramifications in our society .Obviously , a purposely misleading story is a "fake news" but lately blathering social media 's discourse is changing its definition. Some of them now use the term to dismiss the facts counter to their preferred view points.

II. LITERATURE SURVEY

There are several algorithms for detecting fake news. For that we analyze through different classifiers in different way. Some of the classifiers used here are Random Forest Classifier, Logic Regression; Decision Tree Classifiers are the classifiers which we used for obtaining accuracy.

Random Forest Classifier:

A random forest classifier is a machine learning technique that is used to solve regression and classification problems .It utilizes ensemble learning, which is a technique that combines many classifiers to provide solutions to complex problems. A random forest algorithm consists of many decision trees.

Logistic Regression:

It is a classification algorithm, used when the value of the target value is categorical in nature .Logistic Regression is most commonly used when the data in question has binary output .So when it belongs to one class or another is either 0 or 1.

Decision Tree Classifier:

A Decision Tree is a graphical representation of all possible solutions to a decision based on certain conditions .On each step or node of a decision tree, used for classification .We try to form a condition on the features to seperate all the labels or classes contained in a dataset to the fullest purity.

All these classifiers are used to obtain the best accuracy in the classification of results.

III. EXISTING SYSTEM

There exists a large body of research on the topic of machine learning methods for deception detection, most of it has been focusing on classifying online reviews and publicly available social media posts, Particularly now a days we can't even trust the news papers. So We used old classifiers with new modifications. The news taken can be listed as a small table below.

Table 1: Representing News Sources

Top Five Unreliable News Sources		Top Five Reliable News Sources	
Before It's News	2066	Reuters	3898
Zero Hedge	149	BBC	830
Raw Story	90	USA Today	824
Washington Examiner	79	Washington Post	820
Infowars	67	CNN	595

IV. PROPOSED SYSTEM

In the proposed model The existing data models have been used for training data with few modifications, Later the training models are being tested as multiple processes like passive aggression model, multinomial naive bayes and tested on holdout assets later the trained data is spitted and then generates count victimizer to test and train the data later the results are picked for models.

Here the diagram representing the training data models. In this way by using these models we can obtain accuracy in results.

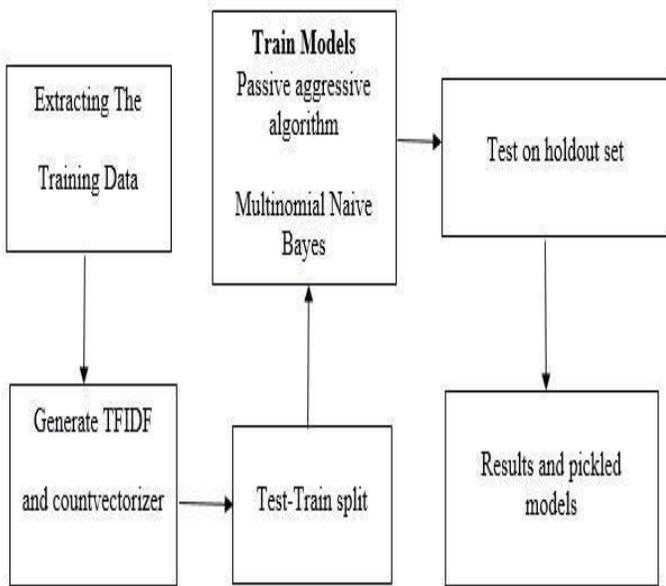


Fig 1: Representing Training Models

IV. METHODOLOGY

The methodology for it was how we are using machine learning algorithms in big data for extraction or classification and evaluation of data through various training data models and some random classifiers which are all listed in literature surveys.

All these classifiers are used in obtaining the exact and nearly values to the problems. It particularly extracts the true and fake news lists in a graphical form and how they are impacting all as shown in fig5 in results. The methodology is very simple and quite interesting. All the classifiers implementation can be done by using Jupiter notebook tool to gain Results in a Way.

Datasets:

The datasets have been collected from various news journals and formed as datasets for two types as true and fake as shown in fig 3&4. All the collected is tested by using all the classifiers to obtain the accurate results from them.

```

1 title,text,subject,date
2 "As U.S. budget fight looms, Republicans flip their fiscal script,"WASHINGTON (Reuters) - The head of a conservative Republican faction in the U.S. Congress, who voted this month for a huge expansion of the national debt to pay for tax cuts, called himself a "fiscal conservative" on Sunday and urged budget restraint in 2018. In keeping with a sharp pivot under way among Republicans, U.S. Representative Mark Meadows, speaking on CBS' "Face the Nation," drew a hard line on federal spending, which lawmakers are bracing to do battle over in January. When they return from the holidays on Wednesday, lawmakers will begin trying to pass a federal budget in a fight likely to be linked to other issues, such as immigration policy, even as the November congressional election campaigns approach in which Republicans will seek to keep control of Congress. President Donald Trump and his Republicans want a big budget increase in military spending, while Democrats also want proportional increases for non-defense "discretionary" spending on programs that support education, scientific research, infrastructure, public health and environmental protection. "The (Trump) administration has already been willing to say: 'We're going to increase non-defense discretionary spending ... by about 7 percent,'" Meadows, chairman of the small but influential House Freedom Caucus, said on the program. "Now, Democrats are saying that's not enough, we need to give the government a pay raise of 10 to 11 percent. For a fiscal conservative, I don't see where the rationale is. ... Eventually you run out of other people's money," he said. Meadows was among Republicans who voted in late December for their party's debt-financed tax overhaul, which is expected to balloon the federal budget deficit and add about $1.5 trillion over 10 years to the $20 trillion national debt. "It's interesting to hear Mark talk about fiscal responsibility," Democratic U.S. Representative Joseph Crowley said on CBS. Crowley said the Republican tax bill would require the United States to borrow $1.5 trillion, to be paid off by future generations, to finance tax cuts for corporations and the rich. "This is one of the least ... fiscally responsible bills we've ever seen passed in the history of the House of Representatives. I think we're going to be paying for this for many, many years to come," Crowley said. Republicans insist the tax package, the biggest U.S. tax overhaul in more than 30 years, will boost the economy and job growth. House Speaker Paul Ryan, who also supported the tax bill, recently went further than Meadows, making clear in a radio interview that welfare or "entitlement reform," as the party often calls it, would be a top Republican priority in 2018. In Republican parlance, "entitlement" programs mean food stamps, housing assistance, Medicare and Medicaid health insurance for the elderly, poor and disabled, as well as other programs created by Washington to assist the needy. Democrats seized on Ryan's early December remarks, saying they showed Republicans would try to pay for their tax overhaul by seeking spending cuts for social programs. But the goals of House Republicans may have to take a back seat to the Senate, where the votes of some Democrats will be needed to approve a budget and prevent a government shutdown. Democrats will use their leverage in the Senate, which Republicans narrowly control, to defend both discretionary non-defense programs and social spending, while tackling the issue of the "dreamers," people brought illegally to the country as children. Trump in September put a March 2018 expiration date on the Deferred Action for Childhood Arrivals, or DACA, program, which protects the young immigrants from deportation and provides them with work permits. The president has said in recent Twitter messages he wants funding for his proposed Mexican border wall and other immigration law changes in exchange for agreeing to help the Dreamers. Representative Debbie Dingell told CBS she did not favor linking that issue to
  
```

Fig 2: True News Data Set

```

1 title,text,subject,date
2 Donald Trump Sends Out Embarrassing New Year's Eve Message; This is Disturbing,"Donald Trump just couldn't wish all Americans a Happy New Year and leave it at that. Instead, he had to give a shout out to his enemies, haters and the very dishonest fake news media. The former reality show star had just one job to do and he couldn't do it. As our Country rapidly grows stronger and smarter, I want to wish all of my friends, supporters, enemies, haters, and even the very dishonest Fake News Media, a Happy and Healthy New Year. President Angry Pants tweeted. 2018 will be a great year for America! As our Country rapidly grows stronger and smarter, I want to wish all of my friends, supporters, enemies, haters, and even the very dishonest Fake News Media, a Happy and Healthy New Year. 2018 will be a great year for America! Donald J. Trump (@realDonaldTrump) December 31, 2017Trump's tweet went down about as well as you'd expect. What kind of president sends a New Year's greeting like this despicable, petty, infantile gibberish? Only Trump! His lack of decency won't even allow him to rise above the gutter long enough to wish the American citizens a happy new year! Bishop Talbert Swan (@TalbertSwan) December 31, 2017One one likes you Calvin (@calvinstowell) December 31, 2017Your impeachment would make 2018 a great year for America, but I'll also accept regaining control of Congress. Miranda Yaver (@mirandayaver) December 31, 2017Do you hear yourself talk? When you have to include that many people that hate you have to wonder? Why do they all hate me? Alan Sandoval (@AlanSandoval11) December 31, 2017Who uses the word haters in a New Year's wish? Harlene (@harlene399) December 31, 2017You can't just say happy new year! Koren politt (@Korencarpenter) December 31, 2017Here's Trump's New Year's Eve tweet from 2016. Happy New Year to all, including to my many enemies and those who have fought me and lost so badly they just don't know what to do. Love! Donald J. Trump (@realDonaldTrump) December 31, 2016This is nothing new for Trump. He's been doing this for years. Trump has directed messages to his enemies and haters for New Year's, Easter, Thanksgiving, and the anniversary of 9/11. pic.twitter.com/4P9Aa2KypA Daniel Dale (@ddale18) December 31, 2017Trump's holiday tweets are clearly not presidential. How long did he work at Hallmark before becoming President? Steven Goodline (@SGoodline) December 31, 2017He's always been like this ... the only difference is that in the last few years, his filter has been breaking down. Roy Schulze (@rbthtttt) December 31, 2017Who, apart from a teenager uses the term haters? Wendy (@WendyWhistles) December 31, 2017He's a Fucking 5 year old who knows (@lilynday88) December 31, 2017So, to all the people who voted for this a hole thinking he would change once he got into power, you were wrong! 70-year-old men don't change and now he's a year older. Photo by Andrew Burton/Getty Images. "News," December 31, 2017
3 Drunk Bragging Trump Staffer Started Russian Collusion Investigation,"House Intelligence Committee Chairman Devin Nunes is going to have a bad day. He's been under the assumption, like many of us, that the Christopher Steele dossier was what prompted the Russia investigation so he's been lashing out at the Department of Justice and the FBI in order to protect Trump. As it happens, the dossier is not what started the investigation, according to documents obtained by the New York Times. Former Trump campaign adviser George Papadopoulos was drunk in a wine bar when he revealed knowledge of Russian opposition research on Hillary Clinton. On top of that, Papadopoulos wasn't just a coffee boy for Trump, as his administration has alleged. He had a much larger role, but none so damning as being a drunken fool in a wine bar. Coffee boys don't help to arrange a New York meeting between Trump and President Abdel Fattah el-
  
```

Fig 3: Fake News Data Set

V. REQUIREMENTS

- Python
- Numpy
- Pandas
- Itertools
- Matplotlib

With all these requirements the project is executed in jupyter notebook tool to obtain the accurate results.

VI. RESULTS

After testing the datasets mentioned above with all the classifiers the accurate and approximate results are obtained as follows. Fig 5 represents the graphical representation of obtained results. Fig 6, 7 & 8 presents a confusion matrix form of results without normalization, and also the count of real and fake news from datasets.

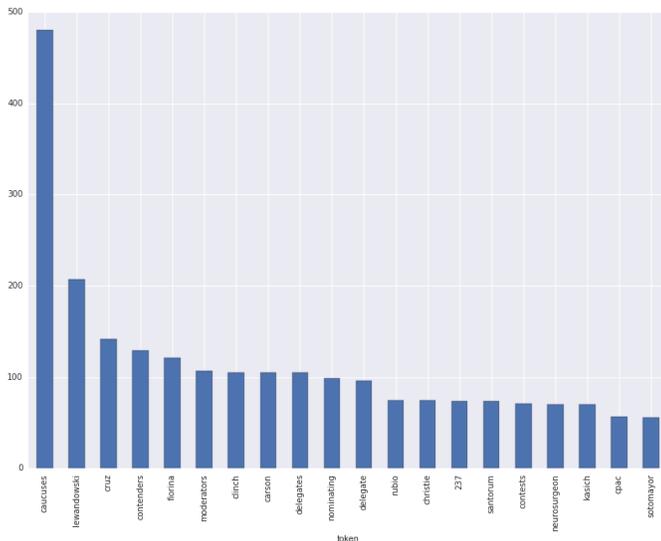


Fig 4: Graph Representing Twitter Fake News 1

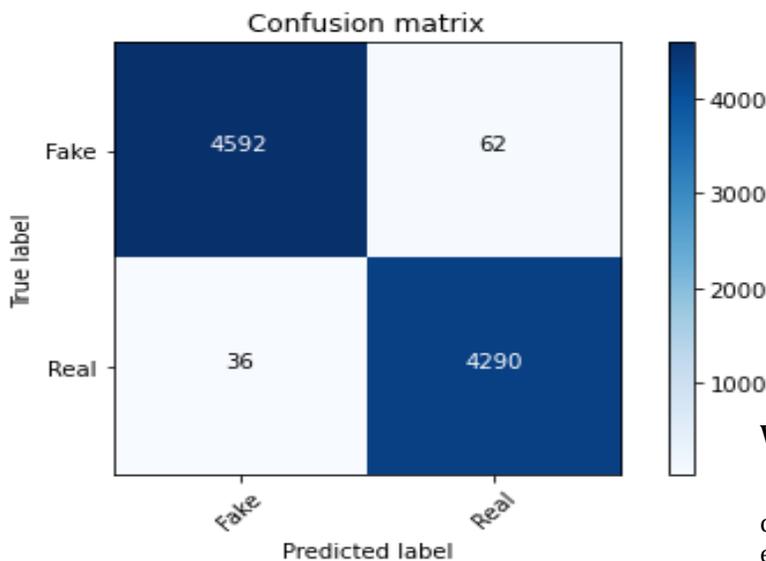


Fig 5: Confusion Matrix, Without Normalization

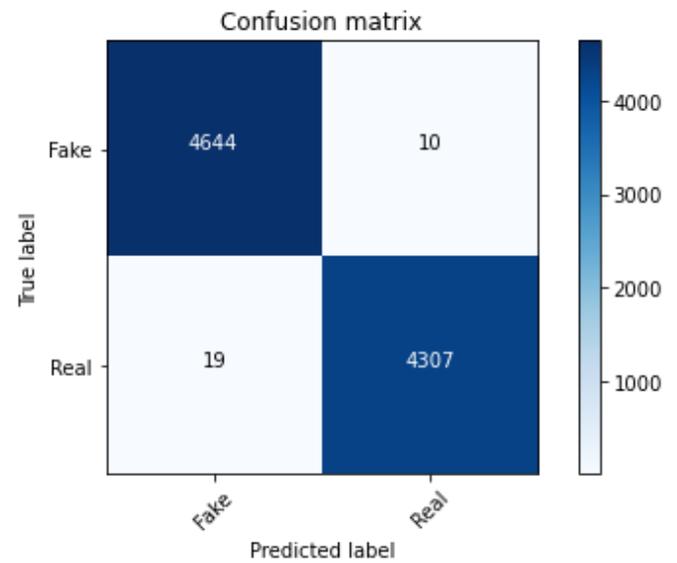


Fig 6: Confusion Matrix, Without Normalization

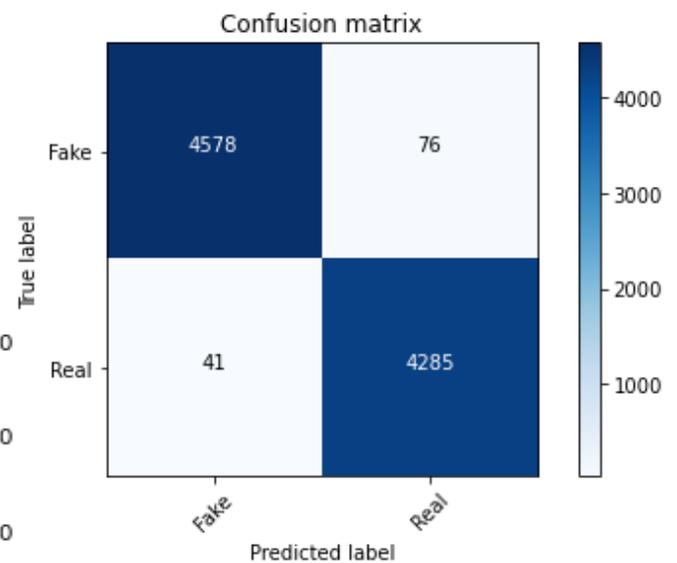


Fig 7: Confusion matrix, without normalization

VII .CONCLUSION

From all these we can conclude that all the results obtained here are accurate of our knowledge and are very easy to understand. As we all know that how people are addicted to social Media now-a-days even for small things they are very dependent on online sources. In that they are forgetting to analyze the thing they have seen all these happens in a large case when it comes to news, So this project is very much helpful in analyzing them to find out the truth.

VIII. REFERENCES

1) N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," Proceedings of the Association for Information Science and Technology, vol. 52, no. 1, pp. 1-4, 2015.

2) S. Feng, R. Banerjee, and Y. Choi, "Syntactic stylometry for deception detection," in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2, Association for Computational Linguistics, 2012, pp. 171-175.

3)Shlok Gilda,Department of Computer Engineering, Evaluating Machine Learning Algorithms for Fake News Detection,2017 IEEE 15th Student Conference on Research and Development (SCORed)

4. Dr.C K Gomathy, Article: A Semantic Quality of Web Service Information Retrieval Techniques Using Bin Rank A Cloud Monitoring Framework Perform in Web Services, International Journal of Scientific Research in Computer Science Engineering and Information Technology IJSRCSEIT | Volume 3 | Issue 5 | ISSN : 2456-3307,May-2018

5.Dr.C K Gomathy, Article: Supply chain-Impact of importance and Technology in Software Release Management, International Journal of Scientific Research in Computer Science Engineering and Information Technology (IJSRCSEIT) Volume 3 | Issue 6 | ISSN : 2456-3307, P.No:1-4, July-2018

6.Dr.C K Gomathy, Article: A Study on the recent Advancements in Online Surveying , International Journal of Emerging technologies and Innovative Research (JETIR) Volume 5 | Issue 11 | ISSN : 2349-5162, P.No:327-331, Nov-2018

Author's Profile:-



1. **Ms.C.V.S.Vasavi**, Student, B.E. Computer Science and Engineering, Sri Chandrasekharendra SaraswathiViswa Mahavidyalaya deemed to be university, Enathur, Kanchipuram, India. Her Area of Big data analytics



2. **Mr. D.Y.V.Rajesh**, Student, B.E. Computer Science and Engineering, Sri Chandrasekharendra SaraswathiViswa Mahavidyalaya deemed to be university, Enathur, Kanchipuram, India. His Area of Big data analytics



3. **Ms.A.Srija** Student, B.E. Computer Science and Engineering, Sri Chandrasekharendra SaraswathiViswa Mahavidyalaya deemed to be university, Enathur, Kanchipuram, India. Her Area of Big data analytics



4. **Dr.C.K.Gomathy** is Assistant Professor in Computer Science and Engineering at Sri Chandrasekharendra SaraswathiViswa Mahavidyalaya deemed to be university, Enathur, Kanchipuram, India. Her area of interest is Software Engineering, Web Services, Knowledge Management and IOT.