AI4TRUTH: AN IN-DEPTH ANALYSIS ON MISINFORMATION USING MACHINE LEARNING AND DATA SCIENCE

Kevin Qu and Yu Sun

California State Polytechnic University, Pomona, CA, 91768

ABSTRACT

A number of social issues have been grown due to the increasing amount of "fake news". With the inevitable exposure to this misinformation, it has become a real challenge for the public to process the correct truth and knowledge with accuracy. In this paper, we have applied machine learning to investigate the correlations between the information and the way people treat it. With enough data, we are able to safely and accurately predict which groups are most vulnerable to misinformation. In addition, we realized that the structure of the survey itself could help with future studies, and the method by which the news articles are presented, and the news articles itself also contributes to the result.

KEYWORDS

Machine Learning, Cross Validation, Training and Prediction, Misinformation

1. INTRODUCTION

With the advent of the information age, the internet has given us access to previously unimaginable wealth of information [7]. With tools such as google, we can access all the collective knowledge of humanity at the press of a button [8]. Yet, with all this power and knowledge, misinformation is somehow more prevalent than ever before [9]. Social media platforms such as Facebook allow misleading headlines and sometimes outright lies to spread to millions of users before anyone can do anything about it. There is a quote - commonly attributed to Mark Twain - that states that "a lie can travel halfway around the world while the truth is still putting on its shoes" [10]. This is made all the more ironic by the fact that Mark Twain most likely never said those words. That does not, however, take away from the truth in the statement, especially in this day and age.

According to the Washington Post, 59% of people comment on fake news headlines before they read the actual article. This can be especially devastating as headlines are often specifically crafted to grab a reader's attention. They often leave out information or straight up lie for views. This means that the majority of people will not get the full story.

According to statists, there are around 4.2 billion internet users across the globe. That is over half of the almost 7.87 billion people in the world, according to world meters. This means that it has become trivially easy to post practically anything and have it be seen. This means that it has become trivially easy to post practically anything and have it be seen. While this does mean that it is easier to spread information, it is also easier to spread falsehoods and rumors.

David C. Wyld et al. (Eds): NLP, MLTEC, CLBD, SEAPP, NeTIoT, VLSIE, ITCS, ARIA, DMS - 2021
pp. 355-362, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.112327

This study focuses on how many people actually read an article after they have seen the headline [11]. Instead of a survey, they simply divided the amount of people who actually clicked on the URL by the people that saw the post. This approach is different from ours mainly because while our study relies on direct user interaction, this one uses a more indirect method [12]. One potential shortfall of this method is that people who have clicked on the links may not necessarily have fully read the article. This study functions in an incredibly similar fashion to this one. They were presented with misinformation and asked whether or not they believed it.

This method, while incredibly similar, is not the same. Instead of directly presenting the participants with misinformation, this survey asks them to identify which one they think is misinformation. This may not sound like a large difference but it is. Their method can introduce unconscious biases that may affect the results. They may also be hesitant to directly admit they believe in misinformation. These two factors could lead to skewed and biased results that cannot be accounted for. This survey, with the randomized questions and intentionally ridiculous headlines, attempts to address this issue by making it so that the real news story cannot be distinguished with ease.

This study uses a very similar approach as this one [14]. It also uses a survey of sorts and focuses on WhatsApp. The information of the participants (Age and Occupation) are taken and their responses filed under those two categories. The questions themselves ask the participant to identify which messages contain real information and which ones don’t. One of the messages will also have a link of some kind to source materials while the other won’t. A “score” is then calculated using what the participant thought were true or false.

This study focuses on whether or not a user will believe the information at first glance. It does not take into account the link that was provided or the website it leads to. It also only includes two factors (age and occupation) while this study has seven. This study also focuses on multiple areas of misinformation, not just health misinformation.

This study aims to send out a survey for the general populous to take. This survey would ask them to identify whether or not a news headline is real or fake based on a screenshot of the website. There are ten such questions and they address multiple areas of interest, including healthcare. The results, along with the attributes of the participants are then fed into a python script where multiple methods of classification are tested. The purpose of the algorithm is to predict the answers of the participants using their attributes. It does not necessarily predict whether or not they will believe in misinformation but rather what factors influence their decision. Unfortunately, we were unable to send this study out so we have opted to use dummy data for the purposes of refining the algorithm [15].

The factors being looked at are: gender, education, age, main source of news, social media use, income and political standing. All of those things are collected in the survey itself, with each one split into multiple categories. Age, for example, is split into the 0-13, 14-17, 18-21, 22-27, 28-35, 35-50, 50-60, and 60+ groups while political standing is split into the far left, left leaning, moderate, right leaning and far right groups. This is to make data collection easier as, even though it somewhat limits the algorithm's range, it does not allow for answers not easily parsed by the algorithm.

The way we will be determining how accurate the results are will be by comparing the algorithms output to the original data. 30% of the original data will be set aside for the algorithm to test on, with the remaining 70% to go towards testing. Unfortunately, as we have mentioned before, we do not have actual data. We only have dummy data for the sole purpose of testing the algorithm so we unfortunately cannot compare our results to that of other studies.

The rest of the paper is organized as follows: Section 2 gives the details on the challenges that we met during the experiment and designing the sample; Section 3 focuses on the details of our solutions corresponding to the challenges that we mentioned in Section 2; Section 4 presents the relevant details about the experiment we did, following by presenting the related work in Section 5. Finally, Section 6 gives the conclusion remarks, as well as pointing out the future work of this project.

2. CHALLENGES

In order to build the tracking system, a few challenges have been identified as follows.

2.1. Designing the Algorithm

One of the challenges we faced was how we wanted to design the algorithm itself [13]. We needed a way for the algorithm to determine which attribute was a contributor to a person's decisions. There were many ways we could have approached this; an example was to feed the algorithm a simple percentage of how many survey questions were correct and have it predict the person's attributes but this would not have been ideal. First off, there were simply too many attributes for an algorithm such as this to have reliably pulled off. Secondly this would not tell us much about which specific attribute contributed the most; it would only tell us which combination would make a person get a certain percentage. The system we ended up going with allowed us to accurately see exactly which attributes would lead to which decisions on all the questions. This allowed for much
more information to be collected from the same amount of data.

2.2. Picking Out the News Articles

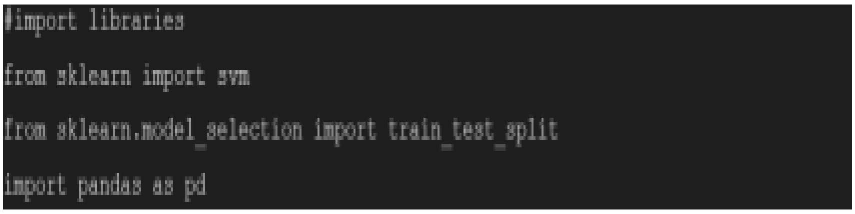
Once the general structure of the algorithms had been decided, the next challenge was to actually pick out the news articles. This was surprisingly difficult as - to provide the most unbiased and accurate set of data - the real news articles have to sound as ridiculous as possible and the fake ones have to sound as real as possible. An example of this would be in the sports section where the real news article was titled "Olympic athlete stuck in quarantine calls lack of fresh air 'inhumane'" (from CNN) while the fake one was titled "Olympics under fire for human rights violations after forcing athletes to exert themselves" (from the Onion). Both of these sound rather far fetched and while the one from the Onions sounds a little more so, both seem to be within the realm of reality. This means that the test comes down to the participants' knowledge of satirical sites (of which the Onion is one) and other factors. It removes the potential for people to easily discern which is which based simply on the ridiculousness of the headline.

2.3. Getting Necessary Data

Another major challenge in conducting this study has been actually getting the necessary data from the survey we constructed. While we did manage to get a handful of responses, it was not nearly enough to both train the algorithm and test it. Beyond that, any survey responses we did manage to get would be heavily skewed and biased seeing as our own friend groups would most likely share the same or at least similar views with us. This means, short of sending out the survey en masse, that any data collected would be more or less useless. As a result, we decided to use dummy data to train the algorithm and make sure it works just so that the experiment can go on.

3. Solution

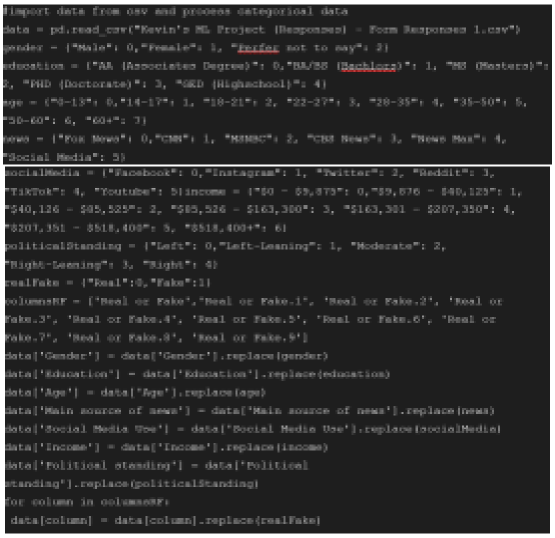
The purpose of the code is to predict, given ample training data, which attributes contribute the most to believing in misinformation. The machine first takes in roughly 70% of the survey responses to train the model. It then uses the remaining 30% to test the accuracy. If the model is able to accurately determine the attributes given the person's responses to the survey, then we know that this would have been a determining factor in whether or not they may believe in misinformation. This test can be repeated for each attribute to determine which one is most likely a determining factor. The first step of this process is to collect data. After the raw data is collected using the survey, it is imported using the pandas library and all the words are swapped with numbers for the machine learning library to understand. The Scikit Learn library is then used for the actual machine learning aspect of the code. Finally, the scores for each of the attributes is printed out at the end to determine whether or not a person with that attribute is likely to believe in misinformation. The first segment of code is the importing of all the libraries.



```
#import libraries

from sklearn import svm

from sklearn.model_selection import train_test_split

import pandas as pd
```

Figure 1. Code of importing libraries

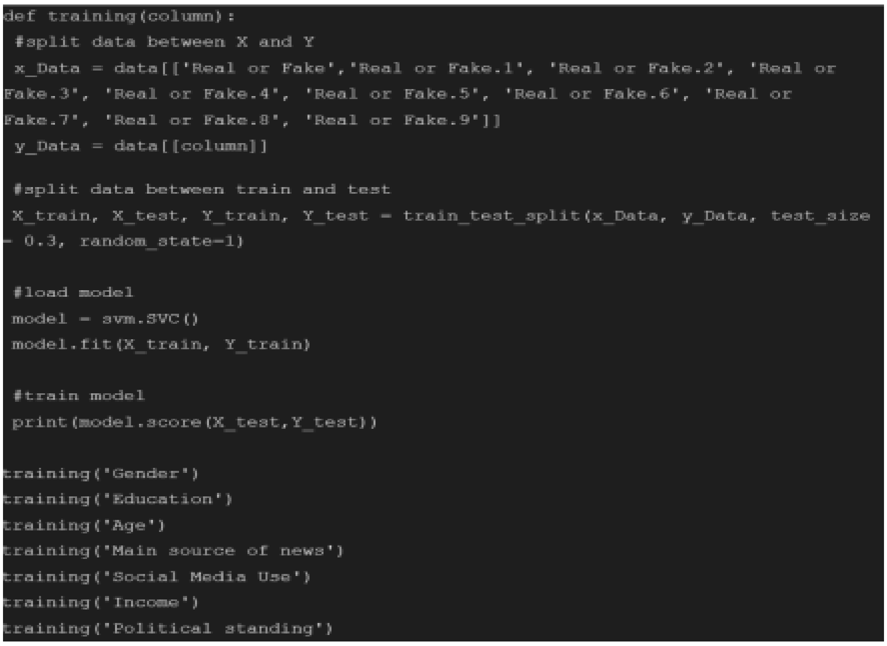
Then comes the data preparation. This includes importing the data with the pandas library and swapping all the words with numbers so that the machine learning library can understand it. An exanple of this process would be that “Male” is replaced with 0 and “Female” is replaced by 1 in the Gender column.



```
#import data from csv and process categorical data
data = pd.read_csv("Kevin's ML Project (Responses) - Form Responses 1.csv")
gender = {"Male": 0,"Female": 1, "Perfer not to say": 2}
education = {"AA (Associates Degree)": 0,"BA/BS (Bachlors)": 1, "MS (Masters)":
2, "PHD (Doctorate)": 3, "GED (Highschool)": 4}
age = {"0-13": 0,"14-17": 1, "18-21": 2, "22-27": 3, "28-35": 4, "35-50": 5,
"50-60": 6, "60+": 7}
news = {"Fox News": 0,"CNN": 1, "MSNBC": 2, "CBS News": 3, "News Max": 4,
"Social Media": 5}
socialMedia = {"Facebook": 0,"Instagram": 1, "Twitter": 2, "Reddit": 3,
"TikTok": 4, "Youtube": 5}income = {"$0 - $9,875": 0,"$9,876 - $40,125": 1,
"$40,126 - $85,525": 2, "$85,526 - $163,300": 3, "$163,301 - $207,350": 4,
"$207,351 - $518,400": 5, "$518,400+": 6}
politicalStanding = {"Left": 0,"Left-Leaning": 1, "Moderate": 2,
"Right-Leaning": 3, "Right": 4}
realFake = {"Real":0,"Fake":1}
columnsRF = ['Real or Fake','Real or Fake.1', 'Real or Fake.2', 'Real or
Fake.3', 'Real or Fake.4', 'Real or Fake.5', 'Real or Fake.6', 'Real or
Fake.7', 'Real or Fake.8', 'Real or Fake.9']
data['Gender'] = data['Gender'].replace(gender)
data['Education'] = data['Education'].replace(education)
data['Age'] = data['Age'].replace(age)
data['Main source of news'] = data['Main source of news'].replace(news)
data['Social Media Use'] = data['Social Media Use'].replace(socialMedia)
data['Income'] = data['Income'].replace(income)
data['Political standing'] = data['Political
standing'].replace(politicalStanding)
for column in columnsRF:
 data[column] = data[column].replace(realFake)
```

Figure 2. Code of data preparation

After the data is all processed, the machine learning part of the code can finally start. Since there are multiple attributes, it is simpler to use a function. In the function, the data is split randomly so that 70% is used to train and 30% is used to check the answers, though only one attribute is used at a time. Once the data is split, the training data is fed into a linear classification algorithm. The remaining 30% of the data is then fed into the algorithm, the answers checked, and the resulting scores printed. This is repeated for each and every attribute.



```
def training(column):
 #split data between X and Y
 x_Data = data[['Real or Fake','Real or Fake.1', 'Real or Fake.2', 'Real or
Fake.3', 'Real or Fake.4', 'Real or Fake.5', 'Real or Fake.6', 'Real or
Fake.7', 'Real or Fake.8', 'Real or Fake.9']]
 y_Data = data[[column]]

 #split data between train and test
 X_train, X_test, Y_train, Y_test = train_test_split(x_Data, y_Data, test_size
= 0.3, random_state=1)

 #load model
 model = svm.SVC()
 model.fit(X_train, Y_train)

 #train model
 print(model.score(X_test,Y_test))

training('Gender')
training('Education')
training('Age')
training('Main source of news')
training('Social Media Use')
training('Income')
training('Political standing')
```

Figure 3. Code of machine learning

4. EXPERIMENT

The purpose of our study is to determine the factors that contribute to a person’s belief in misinformation. In order to determine this, we need to first determine the relevant factors of the participant and whether or not they would believe in misinformation. The easiest way to accomplish the first was to simply ask them for it. This way is the most reliable and it is also easy to get plenty of information from it. The basic information section has ten questions, asking the user their gender, education, age, main source of news, social media use, income and political standing. The second one is slightly more challenging. We decided, instead of simply asking, to ask the participant to determine which ones of the news articles are fake and which ones are real. This way, we get a clearer picture of their decision making process than if we had asked them outright, allowing biases to skew the results. There are ten questions divided into 5 categories: Science, Health, Trivia, Politics and Entertainment/Sports. Each of these categories will have 2 screenshots of a news article including the headline, an image and perhaps a small fragment of the first paragraph or two. Each of these will have 2 possible answers: true or false. This way, it is possible to tell numerous things from the survey. It would be able to tell which area the participant is most interested in, which area is most vulnerable to misinformation and even which political party the participant may be aligned with in the politics section. This study does not take much advantage of this but a future, more in depth study could. The screenshots and news headlines are chosen to be intentionally ridiculous as well to make it harder to distinguish between the real and the fake news articles. Unfortunately, we were not able to get mass responses so we generated our own dummy data instead.

Below is a chart of our experiment results. After generating the dummy data, we needed to develop the best algorithm to process it. As a result, we decided to process the data using different methods to determine the best one. We used SVC, Random Forest Classifier and a Linear Regression Classifier. From the below chart, it would seem that all three methods fared incredibly closely. Since, again, this is dummy data and in no way reflects the real world, the results don’t matter much but it is still very clear that all three models agree to an extent. It would seem that, from the dummy data, ender, main source of news and social media use all play a fairly large role in determining a person's decisions to either believe or discount misinformation.

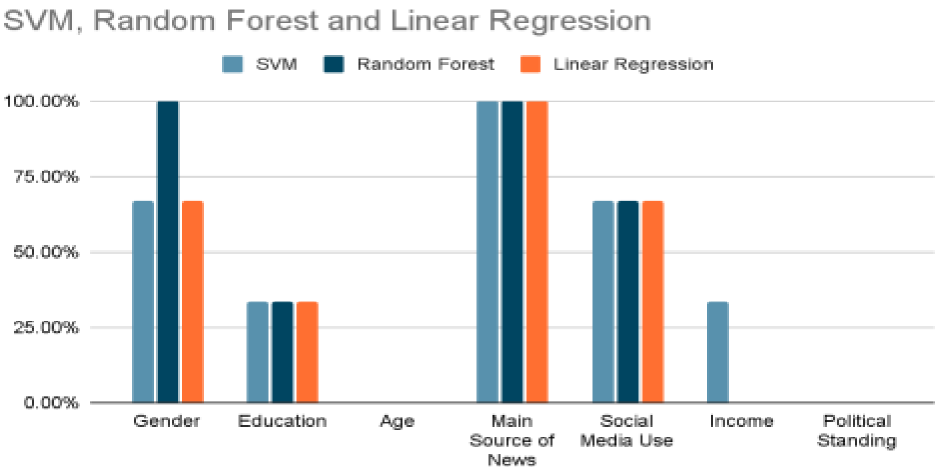


Figure 4. Result of experiment

In the above experiment, we solved the problem of getting participants, getting the basic information of the participants and getting their behavioral patterns. We got around the first by generating dummy data to start. We got around the second by using the simple method of asking them for it and we got around the final one by asking them to classify real and fake news. The experiment was constructed around the goal of training the algorithm to identify which attributes contribute to a participants decision making so the dummy data was generated with clear trends in mind. The main attributes that were focused on were education, main source of news and social media use. These were the ones that were not randomized and instead heavily targeted. They had their answers modified to produce a clear result to prove the algorithm was working. Do keep in mind that this is still dummy data generated for the purpose of developing this algorithm. From the above chart, it is also clear that it worked. The algorithm was able to predict a clear trend in those areas with a few minor deviations.

5. Related Work

This study focuses on how many people actually read an article after they have seen the headline [4]. Instead of a survey, they simply divided the amount of people who actually clicked on the url by the people that saw the post. Through this method, it is much easier to collect a large sample of data and it will have little to no influence from biases. Our method is much more detailed but is otherwise subject to the personal biases of the participant. An example of this would be on the political standings question. The participant may feel like they belong in one group but may instead normally be classified in another.

This study functions in an incredibly similar fashion to this one [5]. They were asked whether or not they would believe health information if they received it from a source such as WeChat. This is very direct and very simple so it is hard to mess up on. The problem with this is also one that we faced: how do we know they are being honest? Many can claim to not believe in information unless provided with a credible source but putting it into practice is another thing altogether. This survey does cut down on that slightly by using more indirect methods to probe the participant for information rather than asking outright but certain questions, such as the basic information, still requires honesty.

This study uses a very similar approach as this one [6]. It also uses a survey of sorts and focuses on WhatsApp. The information of the participants (Age and Occupation) are taken and their responses filed under those two categories. The questions themselves ask the participant to identify which messages contain real information and which ones don’t. One of the messages will also have a link of some kind to source materials while the other won’t. A “score” is then calculated using what the participant thought were true or false.

6. CONCLUSIONS

These algorithms have numerous applications in the real world. With enough data, we will be able to safely and accurately predict which groups are most vulnerable to misinformation. The structure of the survey itself could help with future studies [1]. The method by which the news articles are presented and the news articles itself. The articles are intentionally misleading and ridiculous so that it cannot be immediately determined which one is real and which is not.

One large challenge that we faced in this study was actually getting responses to the survey. We did not have the resources to send this out at a large-scale and collect that many replies [2]. As a result, we decided to use dummy data as a proof of concept and to develop the algorithm. Another limitation is the reliance of the participant itself to provide information. While the rest of the survey attempts to get around this by using indirect methods, the basic information still relies on the honesty of the participants. This also means that any information they give us will be subject to bias as well.

The issue of getting the survey out is not too big of an issue to solve [3]. There are plenty of ways to get many people from taking a survey ranging from paid survey companies to sending out a post on social media. The issue of honesty is harder to solve without violating privacy concerns. Another simple solution would be to ask their friends or family to describe them or ask them to describe why they wrote down what they did.

REFERENCES

[1] Fioranelli, M., et al. "5G Technology and induction of coronavirus in skin cells." (2020).

[2] Lal, P., et al. "Edible vaccines: current status and future." Indian journal of medical microbiology 25.2 (2007): 93-102.

[3] Stribling, Jeremy, Max Krohn, and Dan Aguayo. "Scigen-an automatic cs paper generator." (2005).

[4] Gabielkov, Maksym, et al. "Social clicks: What and who gets read on Twitter?." Proceedings of the 2016 ACM SIGMETRICS international conference on measurement and modeling of computer science. 2016.

[5] Pan, Wenjing, Diyi Liu, and Jie Fang. "An Examination of Factors Contributing to the Acceptance of Online Health Misinformation." Frontiers in Psychology 12 (2021): 524.

[6] Bapaye, Jay Amol, and Harsh Amol Bapaye. "Demographic Factors Influencing the Impact of Coronavirus Related Misinformation on WhatsApp: Cross-sectional Questionnaire Study." JMIR public health and surveillance 7.1 (2021): e19858.

[7] Carnegie, Andrew. "Wealth." The North American Review 148.391 (1889): 653-664.

[8] Schmidt, Eric, and Jonathan Rosenberg. How google works. Grand Central Publishing, 2014.

[9] Godfrey-Smith, Peter. "Misinformation." Canadian Journal of Philosophy 19.4 (1989): 533-550.

[10] O’Hara, Maureen. "What is a quote?." The Journal of Trading 5.2 (2010): 10-16.

[11] Iarovici, Edith, and Rodica Amel. "The strategy of the headline." (1989): 441-460.

[12] Cox, David R. "Interaction." International Statistical Review/Revue Internationale de Statistique (1984): 1-24.

[13] Moschovakis, Yiannis N. "What is an algorithm?." Mathematics unlimited—2001 and beyond. Springer, Berlin, Heidelberg, 2001. 919-936.

[14] Norvig, P. Russel, and S. Artificial Intelligence. A modern approach. Upper Saddle River, NJ, USA:: Prentice Hall, 2002.

[15] Coombs, Clyde H. "A theory of data." (1964)