Textual and Statistical Analysis of Russian IRA Facebook Advertisements

*The paper is written in the scope of a student-faculty collaborative summer research with professor Richard K. Merritt (merritri@luther.edu).

David Oniani
Luther College
oniada01@luther.edu

November 21, 2019

Abstract

The 2016 United States Presidential Election was targeted by an unprecedented intelligence and influence campaign. Arising out of Russian so-called Internet Research Agency (IRA), it sought to sow discord, attack the fissures of the United States, and ultimately, sway the election results. [1] [2] In 2018, a portion of the IRA-backed Facebook advertisements was released by The United States House Permanent Select Committee on Intelligence. All of the advertisements are in the PDF format. We have scraped the PDF files and present the results obtained by both textual and statistical analysis of the above-mentioned data. Authorship attribution and sentiment analysis were also performed. [1] [5] In addition, we trained a binary text classifier and have made the data publicly available for other researchers and/or interested people in a much nicer and easier-to-manipulate CSV format.

1 Data and Preparation

The dataset was scraped from [2] more than 3,500 Russian IRA Facebook [2] posts made publicly available in the PDF format by the House Intelligence Committee. We used a free and open-source `Python` library [6] `pdftotext` to scrape the data. Many PDF files were formatted in a way that it was virtually impossible for `pdftotext` to parse them correctly. For this reason, we have manually reviewed most of the CSV files for validity. [5] All the CSV files have been made publicly available. It is important to note that the dataset is just a sample of a bigger dataset and, albeit *less likely*, might not be a good representation of the overall campaign.

In total, there are 5 different sets of the datasets. Figure 1 categorizes them.

Set	Dataset 1	Dataset 2	Dataset 3	Dataset 4
all	all.csv	N/A	N/A	N/A
by-year	year-2015.csv	year-2016.csv	year-2017.csv	N/A
year-2015	2015-Quarter-2.csv	2015-Quarter-3.csv	2015-Quarter-4.csv	N/A
year-2016	2016-Quarter-1.csv	2016-Quarter-2.csv	2016-Quarter-3.csv	2016-Quarter-4.csv
year-2017	2017-Quarter-1.csv	2017-Quarter-2-April.csv	2017-Quarter-2-May.csv	2017-Quarter-3.csv

Figure 1: Categorization of the Sets of the Datasets

[1] Please note that the paper does not discuss the political implications of these actions, but attempts to explore the methods of persuasion that were employed in this influence campaign.
[2] In the paper, words “advertisement” and “post” are used interchangeably.

All datasets are of the same "shape." Below is the output of Python's pandas package command `dataset.dtypes`.

```
import pandas as pd

data = pd.read_csv("./data/csv/all/all.csv", na_filter=False, thousands=",")
print(data.dtypes)

Ad ID                       int64
Ad Text                    object
Ad Landing Page            object
Ad Targeting Location      object
Excluded Connections       object
Age                        object
Language                   object
Placements                 object
People Who Match           object
Friends of Connections     object
Ad Impressions              int64
Ad Clicks                   int64
Ad Spend                   object
Ad Creation Date           object
Ad End Date                object
dtype: object
```

Figure 2: Dataset Layout

Every row represents a Facebook advertisement which has 15 different properties and thus, the dataset has 15 columns. Out of these columns, only 3 are integer values, namely "Ad ID", "Ad Impressions", and "Ad Clicks." The rest of the columns are of the string type. This is the case for all datasets.

2 Word Analysis

Among the strategies utilized by this influence campaign, one that stood out the most was the exploitation of internal issues of the United States by leveraging the social, political, and historical backgrounds of the country.

The word choice played a crucial role in the effectiveness of the influence campaign. An overwhelming number of advertisements referenced issues including race, police, immigration, religion, and guns. Figure 2 shows a bar chart with TOP 25 most commonly used words after eliminating linking verbs, prepositions, pronouns, and other [3] non-relevant words.

[3]The list of all eliminated words is provided at https://github.com/oniani/ira-analysis/blob/master/eliminated-words.txt.

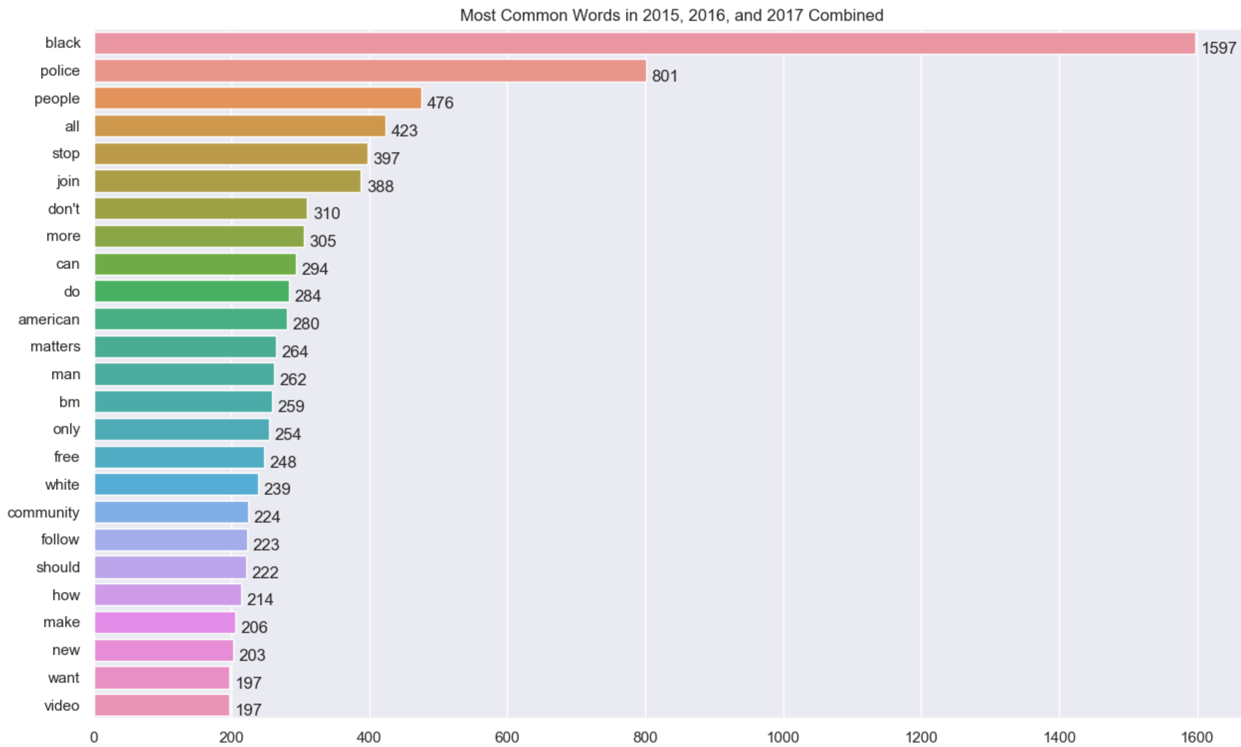


Figure 3: Most common words in 2015, 2016, and 2017 combined.

The vast majority of Facebook ads featured words and phrases related to race and race-based discrimination. More than 50% of the posts referenced race and generated over 25 million ad impressions. Three most commonly used words were "black", "police", and "people."

Words associated with "taking action" such as "join", "stop", "do", "follow" etc. were used extensively for persuasive purposes. These words in conjunction with others such as "American", "matters", "community" etc. further increased the persuasive quality of advertisements by metaphorically connecting them with cultural and social identities of people and fueling hateful rhetoric.

117 posts referenced words "Clinton" and "Trump" (last names of the presidential candidates at the time) accounting for 61,774 ad clicks and 853,855 ad impressions in total. There were 3 posts mentioning word "Donald" and not mentioning "Trump" resulting in 226 clicks and 2,182 impressions, but none mentioning "Hilary" and not "Clinton."

There was only a single ad directly referencing Russia with 135 clicks and 1,236 impressions. The advertisement was denouncing the request for the increase in the US military spendings and making a comparison with the Russian military budget.

3 Sentiment Analysis

For sentiment analysis, we used a free and open-source `Python` library `TextBlob`. Interestingly, out of all Facebook posts with a non-empty `Ad Text` value, 1643 were positive, 900 negative, and 932 neutral leaving the overall tone of the posts positive. Yet, the claim is not strong or significant as the polarity levels were always near zero. Such low polarity levels, however, demonstrate a highly intelligent design of advertisements.

As the `Ad Texts` did not give us any strong evidences, we looked at the negativity levels across all three years and found a consistent trend (Figure 4).

Subjectivity	Year 2015	Year 2016	Year 2017	2016 (%)
1.0	154	562	176	63.004
0.75	150	547	169	63.164
0.5	123	405	120	62.500
0.25	15	81	22	68.644
0.15	5	33	10	68.750
0.1	2	11	6	57.895
0.05	1	5	2	62.500

Figure 4: Analyzing Negativity

Notice that for any given year and at any subjectivity level, year 2016 consistently comprises around 60% of all posts. The sudden change in strategy and increased levels of negativity could be linked to approaching Presidential Election and making people take an action by voting for candidates promoted by these ads. The year of the Presidential Election was rather negative.

4 Authorship Attribution

Since all of these posts were issued by the same organization/entity, it was interesting to see if there are some common patterns between the Facebook ads of 2015, 2016, and 2017. For this reason, we have performed authorship attribution tests by implementing [3] the paper by Koppel et. al.

We have performed 3 authorship attribution tests:

1. Assume that the Facebook ads of years 2016 and 2017 were written by the same author and assess the similarity for the ads of the year 2015.
2. Assume that the Facebook ads of years 2015 and 2017 were written by the same author and assess the similarity for the ads of the year 2016.
3. Assume that the Facebook ads of years 2015 and 2016 were written by the same author and assess the similarity for the ads of the year 2017.

In all three cases, we had to merge different posts to meet the required minimum text length of 500 words. [4] This was done using randomization to avoid bias. [5] Results are available in the form of JSON files formatted in the manner shown below.

[4]Note that because of randomization, texts were somewhat scrambled and are not available in any format. That said, one can easily redo the authorship attribution with similar accuracy using the algorithm which is already implemented and resides in the GitHub repository, https://github.com/oniani/ira-analysis/tree/master/koppel11.

[5]Results JSON: https://github.com/oniani/ira-analysis/tree/master/koppel11/results

```
{
  "answers": [
    {
      "unknown_text": "2015-unknown1.txt",
      "author": "candidate2016",
      "score": 0.58
    },
    ...
    {
      "unknown_text": "2015-unknown56.txt",
      "author": "candidate2016",
      "score": 0.76
    }
  ]
}
```


Figure 5: JSON example

The first answer tells us that the unknown text was written in year 2015, and that there is 58% chance that it was written by the author of Facebook ads of year 2016. The last answer shows that there is 76% chance that the given ad, which was posted in 2015, was authored by the entity who wrote the posts in year 2016.

Figure 6 shows the results obtained after performing all three above-mentioned authorship attribution tests.

Years	Similarity (%)
2016 and 2017	72.509 (similarity to 2015)
2015 and 2017	68.516 (similarity to 2016)
2015 and 2016	62.392 (similarity to 2017)
Average (2015, 2016, and 2017)	67.806%

Figure 6: Average Similarity

On average, there is approximately 68% similarity across texts suggesting that the design strategy for these posts were roughly similar.

5 General Statistics

The distribution of posts across all three years is of a bimodal nature with two peaks in quarters 2 and 4 of the year 2016. Given the fact that the US Presidential Election was held in the fourth quarter of 2016, it is surprising that the number of posts in the second quarter exceeded that of the fourth quarter.

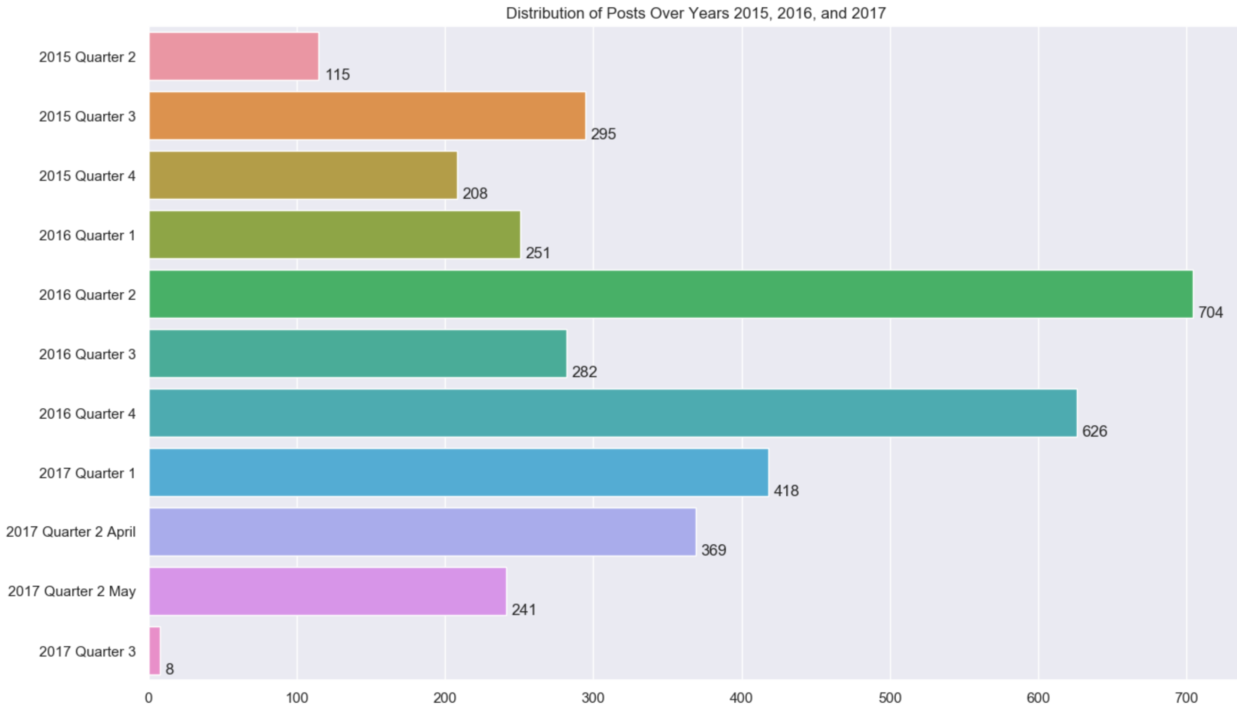


Figure 7: Distribution of Posts Over Years 2015, 2016, and 2017.

As for ad spendings, the fourth quarter of 2016 exceeds that of any quarter, with the second quarter coming next. Interestingly, most of the ads were paid in the Russian currency (ruble) with two exceptions (4 posts in total) in 2016 quarter 3 and 2017 quarter 1 when IRA spent $74.000 and $35.330 respectively.

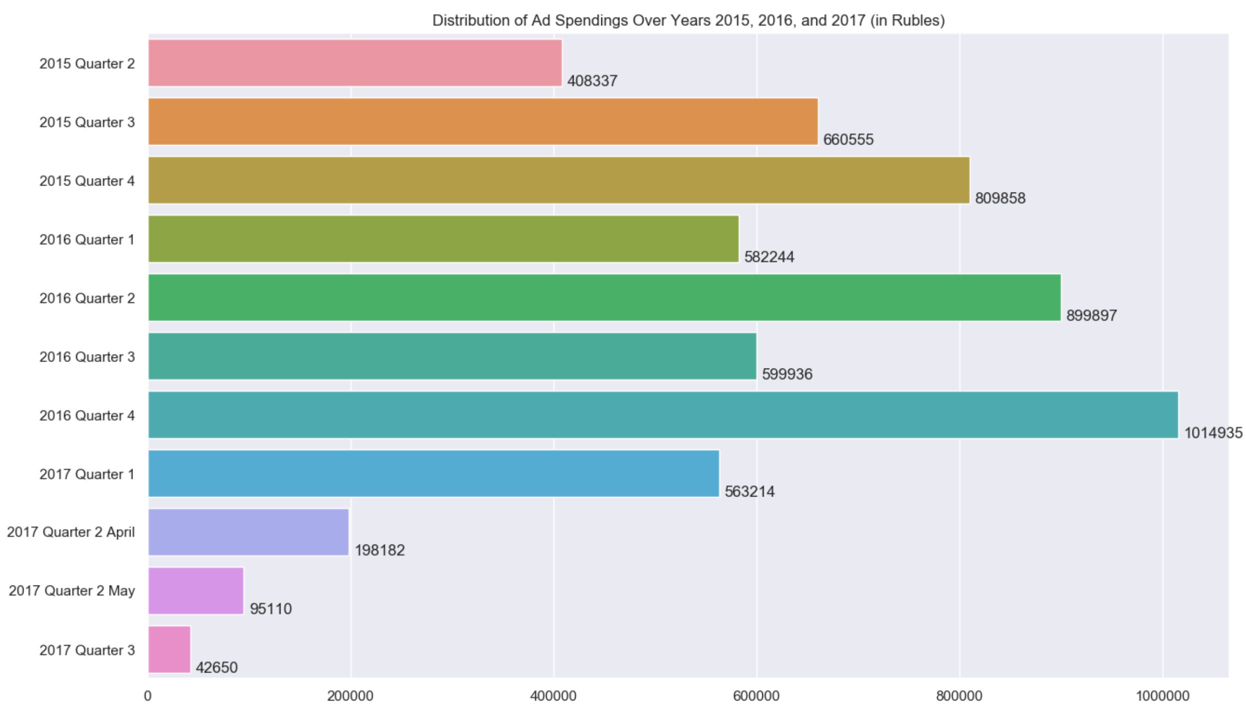


Figure 8: Distribution of Ad Spendings Over Years 2015, 2016, and 2017 (in rubles).

Furthermore, 99.8% of all *paid* ads across all three years were paid in rubles. Figure 9 shows a chart with the number of posts based on currency.

Currency	Total (All Years)
RUB	2549
USD	5
[6] None	787
0	176

Figure 9: Number of Posts Depending on the Currency.

As the information for the reader, the Russian ruble is used only in Russia, Belarus, and two regions of Georgia, which are considered by Russia as partially recognized states of Abkhazia and South Ossetia.

IRA was somewhat liberal toward age, with 69.542% of advertisements targeting all individuals of age 18 or more. Figure 10 shows the distribution of these ads.

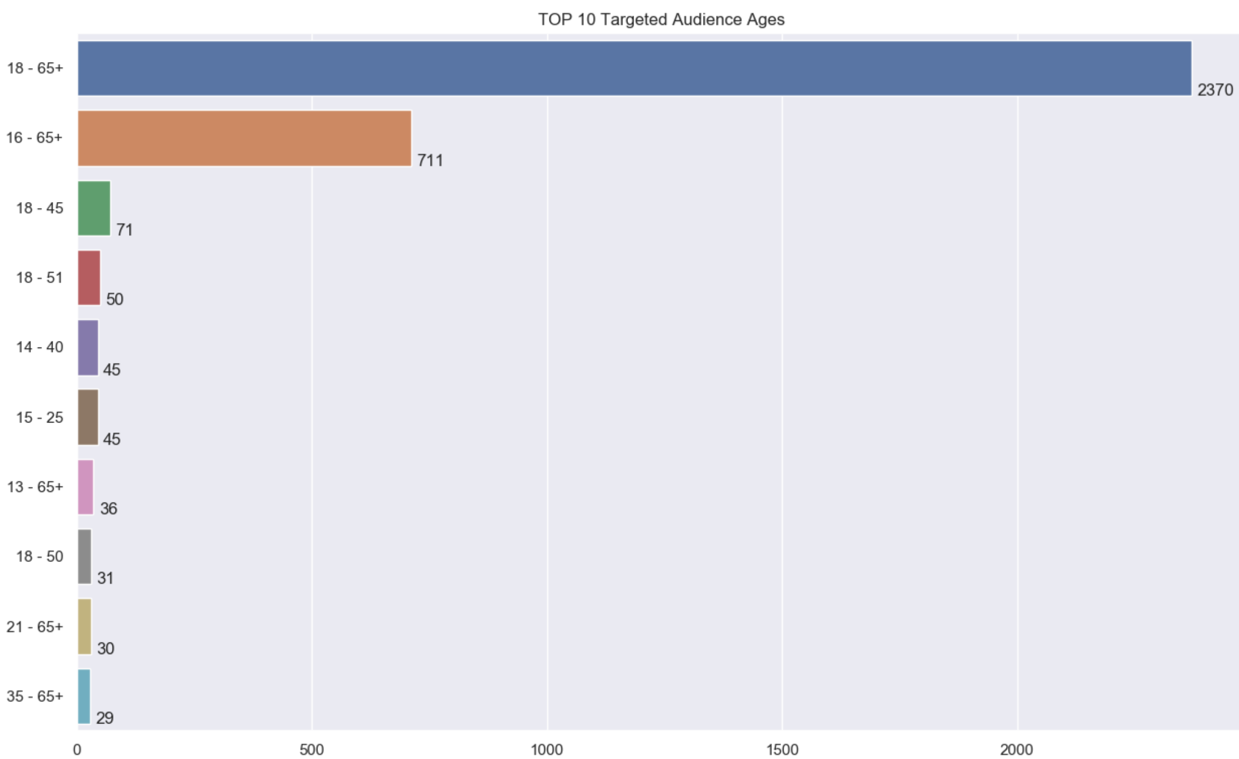


Figure 10: Distribution of Posts Based on Age (Years 2015, 2016, and 2017).

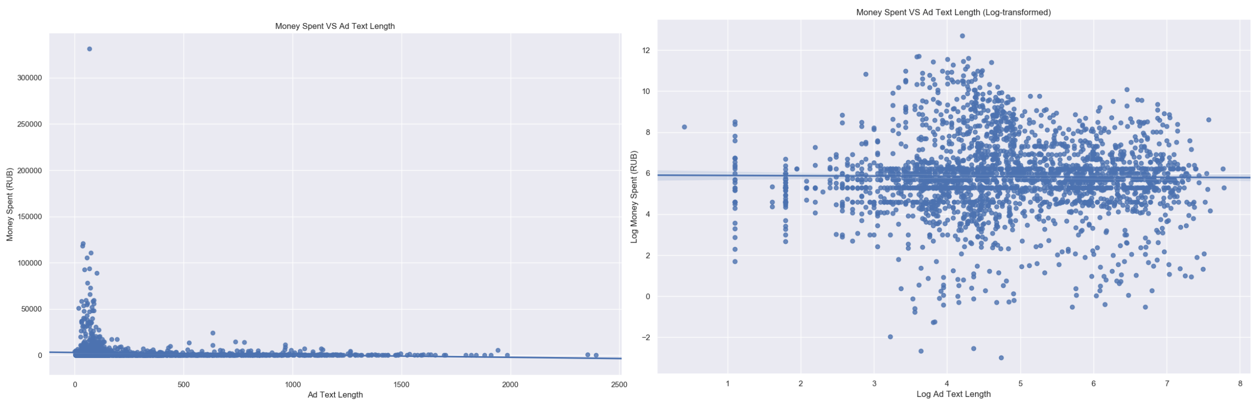
Surprisingly, nearly all advertisements that targeted the 15 – 25 age category featured a website `musicfb.info` which [4] appears to be link to a Chrome extension (currently unavailable) and was registered in Saint Petersburg, Russia (location of IRA). This is a notable fact since music is strongly linked to emotions and can be used as a powerful weapon of manipulation.

[6] None values can be considered 0.

6 Money Spent VS Ad Spend Length – Regression Analysis

We proceed by determinining whether there is a statistically significant relationship between two quantitative variables: the textual length of an advertisement and the money spent on it. Finding the relationship between these variables would give us an idea if, depending on the textual length, some posts had a higher priority than the others.

For this task, we use a linear regression approach. Note that we perform the analysis only on the ads paid in rubles, the primary reason being not having enough data points for USD (only 5 values).



(a) Money Spent VS Ad Spend Length (b) Log Money Spent VS Log Ad Spend Length

Figure 11: Simple VS Log-transformed

The first plot showed no strong relationships between the response and the predictor. That said, the shape of the plot hinted us to apply the log transformation. The log transformation, once again, verified our assumption of no relationship. The numerical data in Figure 11 further reinforces our observations.

RUB (Without Log Transform)	
Slope	-2.594
Intercept	2980.780
R-squared	0.007
P-value	$3.680 * 10^{-5}$
Standard Error	0.627

RUB (With Log Transform)	
Slope	-0.015
Intercept	5.904
R-squared	0.0001
P-value	$0.594 * 10^{-5}$
Standard Error	0.028

Figure 12: Linear Regression Analysis Results.

The chart above, R-squared equals 0.007, which tells us that only 0.7% of the variation in the money paid for advertisements is explained by the number of characters. The residual standard error is 0.627 which is very small compared to the intercept whose value is 2980.780 (RUB). The p-value equals $3.680 * 10^{-5}$ which is a lot smaller than 0.05 and makes the conclusion statistically significant.

Even after performing a log transformation, the results show no evidence of a strong or even a weak relationship between the variables. The p-value reassures us that the results are statistically significant.

Although having no relationship is usually not the desired result, in our case, we can conclude that posts were designed with no noticeable priorities depending on the number of words in the ad text section.

7 Binary Classifier

For the text classification purposes, we have built a binary classifier. The program can, with the high accuracy, tell apart IRA and non-IRA texts. Model has three features: text length, number of punctuation marks, and the frequency of TOP 25 most common words discussed previously in the paper. We took the ensemble learning approach and utilized the random forest classifier (`RandomForestClassifier`) from the *Scikit-learn* package with 180 estimators and unlimited depth. The non-IRA text for the separate label was obtained from the `textgen` package (https://github.com/minimaxir/textgenrnn). Figure 13 shows the metrics which are the measures of robustness of the model.

Criterion	Value (%)
Precision	96.2
Recall	94.7
Accuracy	95.5

Figure 13: Precision, recall, and accuracy of the model

Besides, we have provided a simple interface for the classification of the textual data. The program is called `predict.py` and is available at https://github.com/oniani/ira-analysis/blob/master/nlp-model. For predictions and testing purposes, we recommend the minimum textual length of 43 words and the maximum textual length of 420 words.

It should be noted that, despite the high accuracy of prediction, the NLP model is rather sensitive to the input data and in some cases, even slight changes could lead to a misclassification.

8 Conclusions

We have scraped PDF files made publicly available by The United States House Permanent Select Committee on Intelligence and performed several different analyses including textual and statistical. The obtained results gave us important insights into a variety of approaches and strategies used in this influence campaign. We have also built a binary classifier for telling apart IRA and non-IRA texts. We conclude that the advertisements were designed in a complex and intelligent manner, targeting minorities, promoting hateful rhetoric, and attempting to sway the election results. Much could be analyzed and inferred from the provided datasets and any researcher and/or interested person is free to use them for their research.

References

[1] The United States House Permanent Select Committee on Intelligence. *Exposing Russia's Effort to Sow Discord Online: The Internet Research Agency and Advertisements*. 2018. URL: https://intelligence.house.gov/social-media-content/ (visited on 11/18/2019).

[2] The United States House Permanent Select Committee on Intelligence. *Social Media Advertisements*. 2018. URL: https://intelligence.house.gov/social-media-content/social-media-advertisements.htm (visited on 11/18/2019).

[3] Moshe Koppel, Jonathan Schler, and Shlomo Argamon. "Authorship attribution in the wild". In: *Language Resources and Evaluation* 45 (Mar. 2011), pp. 83–94. DOI: 10.1007/s10579-009-9111-2.

[4] Issie Lapowsky. *Russia-Linked Facebook Ads Targeted a Sketchy Chrome Extension at Teen Girls*. 2018. URL: https://www.wired.com/story/russia-facebook-ads-sketchy-chrome-extension/ (visited on 11/18/2019).

[5] David Oniani and Richard Merritt. *CSV data scraped from the PDF files of IRA Facebook posts*. 2019. URL: https://www.davidoniani.com/datasets (visited on 11/18/2019).

[6] Jason Alan Palmer. *pdftotext*. 2018. URL: https://pypi.org/project/pdftotext/ (visited on 11/18/2019).