Erasmus University
Erasmus School of Economics

Master Thesis Economics and Business Economics

Neural network approach to Russian botnet

Student name: Nikita Nesterov
Student ID number: 422428
Supervisor: Prof Dr. O. R. Marie
Date: 02.12.2019

Abstract

This paper contributes to the ever-expanding body of literature on machine learning and predictive analysis. The research models the output of Twitter accounts associated with the Russian Internet Research Agency during the Ukranian Crisis. The paper sets a competitive forecast benchmark of 82.75% accuracy with vector autoregressive moving average model with exogenous variables and proceeds to employ recurrent neural networks. The results suggest that the time series can be accurately predicted using various architectures, with bidirectional long short-term memory variation achieving a 92.45% forecast accuracy. The predictions are in part based on tweets' content features extracted with natural language processing.

1 Introduction

The USSR is widely regarded as the world's first propaganda state, a regime that harnessed culture, education, and media in the process of nationwide indoctrination[1]. Political campaigns of the 1930s climaxed with the publication of *History of the Communist Party of the Soviet Union (Bolsheviks)*, which "instead of the Bible" was meant to "give a rigorous answer to many important questions."[2] The Soviet propaganda machine remained on active duty throughout the Cold War, brought to an end by Gorbachev's revolutionary program of glasnost and perestroika which inevitably resulted in the dissolution of the Soviet Union[3]. Boris Yeltsin was elected the first president of the Russian Federation as the country entered the wild 90s, a decade of demographic decline, social stratification, and organized crime. In a 1996 election, characterized by pervasive corruption[4], Yeltsin was re-inaugurated despite the preceding crises. He resigned in 1999 and was followed by Vladimir Putin who remained Russia's undisputed leader ever since, despite having to temporarily entrust presidency to Dmitry Medvedev in 2008, due to legislative reasons, in a "neither free nor fair" election[5].

As of lately, however, once the world's most powerful[6] and influential[7] leader appears to lose support on the home front. According to the state-run Russian Public Opinion Research Centre (WCIOM), frequently accused of having an exceedingly favorable relationship with the Kremlin[8], the presidential approval rating is only 33.4% as of January 2019[9]. High inflation expectations, unpopular reforms and a general disappointment in the regime are a few reasons underlying this historically low index. Formerly seen as a person who brought fame and glory to Russia, Putin is now held responsible for the country's problems. WCIOM's general director Valery Fedorov emphasizes that "As long as people have an adverse outlook, the actions of authorities will be perceived with distrust." Political scientist Nikolay Petrov adds: "There is a turning point for power which requires heroic efforts to support Putin's diminishing legitimacy."[10] The backlash against the presidential 2019 New Year address, detailing the current state of affairs, further exposes genuine public sentiment, as the speech receives in excess of 66 thousand dislikes on Channel One Russia YouTube page[11]. Owing to the lack of meaningful political opposition, Putin can potentially only fall from power as a result of civil disturbance.

State propaganda has proven time and again to be an effective crowd control instrument[12] by employing repetition, isolationism, association building techniques, and exploiting various communication channels. It is defined as a systematic form of purposeful persuasion that attempts to influence the emotions, attitudes, opinions, and actions of specified target audiences for ideological, political or commercial purposes[13]. The rise of social media has enhanced the capabilities of government indoctrination, as evidenced by post-revolutionary Egypt[14], the political landscape of Venezuela[15], and Russian military intervention in Ukraine[16]. It is exceedingly potent in affecting individuals' opinions and behavior[17]. Social media has helped to foster democratic conversation on various public and political issues[18] but has also been used to propagate misinformation[19].

Propaganda messages are commonly spread and amplified by bots, computer algorithms that automatically produce content and interact with humans. Bots have become a common sight on social media[20] with an estimated 48 million Twitter accounts worldwide controlled by computers[21]. Numerous studies claim that Russia's Internet Research Agency (IRA) has been spreading false and politically biased material on Twitter[22,23]. There have been attempts to map out the Russian botnet[24], yet no research has ventured to model and predict the behavior of the government-backed bots. This paper aims to design a framework that could analyze their role in spreading misinformation, and ultimately forecast the volume of messages that the accounts related to the IRA produce as a response to the ongoing events.

The predictive analysis comes in various forms. Statistical models aim to approximate reality in a mathematically formalized way, whereas artificial neural networks (ANNs) are designed to resemble the neurons in the human brain and form the basis for artificial intelligence. A framework once established to resemble a system of electrical circuits[25] is now capable of solving complex problems and uncovering the most subtle of correlations. There exist different ANN architectures, including feed-forward, convolutional, and recurrent. In an effort to model the tweet output volume, this paper employs a recurrent neural network (RNN) architecture that is specifically designed to identify data's sequential characteristics and predict future outcomes. Vector autoregressive moving average model with exogenous variables (VARMAX) sets a performance benchmark for this neural network.

The research draws inspiration from the developments in the field of deep learning and natural language processing (NLP). Adhering to the notion of the importance of data, sentiment and objectivity variables are tailor-made to fit the analysis' needs using support vector machine (SVM) classifiers. It appears that the VARMAX is capable of predicting the output volume with an average accuracy of 82.76%. The recurrent neural network with long short-term memory (LSTM) surpasses this benchmark by 2.68 percentage points. The addition of a bidirectional information flow effectively halves the forecast error for a 92.45% accuracy. The model with gated recurrent unit (GRU) displays a relatively inferior performance and barely misses the benchmark. Altogether, the array of predictive models is highly capable of learning the trend movement and highlights the capabilities of the neural networks in predictive analysis.

The remainder of the paper is organized as follows: section 2 provides an overview of the developments in the field of predictive analytics by discussing the related literature. Section 3 specifies the statistical model that sets a benchmark for the deep learning analysis, whereafter the principles of recurrent neural networks are explained. Besides, this section outlines the construction of NLP-based predictors. Section 4 describes the data as a whole and the nature of each variable individually. Section 5 assesses the models, forecasts the volume of tweets generated by the bot-controlled accounts, and presents the results. Section 6 extends the deep learning framework with variations to cell structure and network architecture as well as comparing the models' performance. Lastly, section 7 concludes while also highlighting the research limitations and providing suggestions for future studies.

2 Related literature

Present-day predictive modeling and analysis find their origin in the dawn of the computer age of the 1940s. At the time, the research primarily aimed to assist the Allies in the war efforts and was characterized by revolutionary works aspired to break the German Enigma code and the designs of automated targeting for anti-aircraft weapons. In the 1950s, the world saw the commercialization of predictive analytics with weather forecasting and optimization of air travel logistics. Decades later, analytics went mainstream and the recurrent neural networks were first introduced.

In his original paper, Hopfield outlined the conceivable capabilities of the framework modeled after the operations of the human brain[26]. Successive research sought to overcome many of the shortcomings of the pioneering work by adopting progressively intricate network architectures. Many studies failed to improve upon the groundwork[27], others proved impractical[28], and only a few successfully resolved the issues inherent to early adoption and caught on. The long short-term memory[29] cell structure proposition was one of these studies. This variant of the recurrent neural network has since been adopted by many researches achieving state-of-the-art results and retains popularity despite the new variations, such as the gated recurrent unit[30], being introduced.

Statistical modeling has likewise greatly evolved throughout the years and is used in many predictive analysis tasks. Works that utilized autoregressive moving average models have successfully predicted out-of-sample power output of electrical grids[31], surface water level fluctuations[32], and tourism revenues[33], among others. Still, purely statistical approach predominantly takes place in the academic world. Backed by continuing technological advancement, artificial intelligence-based analysis overcomes many of the limitations of these traditional prediction methods. Consequently, in today's age of big data, recurrent neural networks are employed by government agencies, research institutions, and corporations in a wide variety of tasks. Machine learning has made it possible to make predictions regarding the ongoing business processes based on past observations. Similarly, the market prices of electricity[34] can now be reliably predicted and the stock market trends can be forecasted weeks into the future[35].

In a recent development, machine learning algorithms are used in conjunction with natural language processing. This methodology has unlocked the insights and analytical potential of written[36] and spoken[37] natural language data. Otherwise known as computational linguistics, it plays a pivotal role in the operation of personal assistants akin to Apple Siri and Amazon Alexa[38]. Besides, natural language processing is used to extract and analyze data from social media[39]. Opinion mining and sentiment analysis have a wide range of applications, such as allowing governmental institutions to estimate crime rates[40] and firms to predict market demand for their products[41].

3 Methodology

This section first generalizes the original univariate Box-Jenkins autoregressive moving average framework[42] and adds exogenous variables. Then, it explains the intuition behind the recurrent neural networks following the groundwork of Hochreiter and Schmidhuber[29]. Lastly, it outlines the natural language data processing following the guidelines of the natural language tool kit (NLTK) module, outlined by Loper and Bird[43].

3.1 VARMAX

Vector autoregressive moving average (VARMA) model extends the univariate deisgn by including k time series with their lags as regressors. The VARMA(p, q) process of variable y_i, can be denoted as follows:

$$y_t = \sum_{i=1}^{p} \varphi_i y_{t-1} + \epsilon_t - \sum_{i=1}^{q} \Theta_j \epsilon_{t-j}$$

In this equation, vectors y_i and ϵ_i contain k univariate time series with $k \geq 1$, and ϵ_t is the error term with conditional mean of zero.

The number of tweets generated at any point in time is assumed to be correlated with own past values, as well as current and past values of other variables. The VARMAX process is appropriate in this setting as it allows to capture linear interdependencies of multiple series. Besides, unlike structural models it only requires a list of variables that, in theory, can intertemporally affect each other. Hence, the VARMAX(p, q, s) extends the model by introducing s exogenous variables and is given by the following expression:

$$y_t = \sum_{i=1}^{p} \varphi_i y_{t-1} + \sum_{i=0}^{s} \Theta_i^* x_{t-i} + \epsilon_t - \sum_{i=1}^{q} \Theta_i \epsilon_{t-i}$$

Here, $y_t = (y_{1t}, ..., y_{kt})$ denotes the endogenous variables that can be influenced by exogenous $x_t = (x_{1t}, ..., x_{rt})$ determined outside of the model. $\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{kt})$ represents the unobserved variables of a vector white noise process with $E(\epsilon_t) = 0$, $E(\epsilon_t \epsilon_s') = 0$ for $t \neq s$, and $E(x_t \epsilon_t') = 0$, meaning the residuals are uncorrelated with exogenous variables and have an expected mean value of zero. Besides, the roots of $|\varphi(z)| = 0$ and $|\Theta(z)| = 0$ are presumed to lay outside the unit circle, an assumption needed to ensure the invertibility of the linear process and stationarity of the time series.

The equations for minimal mean squared error l-step-ahead forecast of the endogenous variable y_{t+l} can be expressed as follows:

$$y_{t+l|t} = \sum_{j=1}^{p} \varphi_j y_{t+l-j|t} + \sum_{j=0}^{s} \Theta_j^* x_{t+l-j|t} - \sum_{j=l}^{q} \Theta_j \epsilon_{t+l-j},\ l \leq q$$

$$y_{t+l|t} = \sum_{j=1}^{p} \varphi_j y_{t+l-j|t} + \sum_{j=0}^{s} \Theta_j^* x_{t+l-j|t},\ l > q$$

here, $y_{t+l-j|t} = y_{t+l-j}$ and $x_{t+l-j|t} = x_{t+l-j}$ for $l \leq j$. The VARMAX process outlined in this section is computed with the *statsmoÍdels* Python module.

Upon fitting the model, the Ljung-Box test[44] is used to assess the assumptions by testing for the existence of autocorrelation in the residuals. Given the time series of length n, the test statistic is compiled as follows:

$$Q = n(n+2) \sum_{k=1}^{m} \frac{\hat{r}_k^2}{n-k}$$

here, m is the number of lags, and $\hat{r}_k^2$ is the estimated autocorrelation of the series at lag k. The test indicates the presence of autocorrelation by rejecting the null hypothesis if $Q > \chi^2_{1-\alpha,h}$, where $\chi^2_{1-\alpha,h}$ is the chi-square distribution table value with h model parameter-adjusted degrees of freedom, and a significance level α.

The forecast accuracy can be determined in a number of ways and this paper considers two of the most prevalent. The root mean squared error (RMSE) of the forecast measures how far on average the predicted values are from the true values and is calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (f_i - o_i)^2}{N}}$$

The mean absolute percentage error (MAPE) is the second measure used to determine forecast accuracy and is given by the following expression:

$$MAPE = \frac{1}{N} \sum_{i=1}^{N} |\frac{o_i - f_i}{o_i}|$$

In both instances, f_i and o_i denote respectively the forecasted and observed values, and N stands for the number of predicted data points. Of note, the forecast accuracy of neural networks is calculated in a similar way in order to facilitate an unbiased inter-model comparison.

3.2 RNN

Figure 1 highlights the standard architecture of a recurrent neural network, that instead of relying on statistical assumptions, utilizes layers of interconnected neurons for processing input and generating output. First, the variables are transformed into machine-readable input vectors. Next, the algorithm learns the patterns and correlations present in the data and ultimately produces the output. Formally, given a set of variables $x = (x_1, x_2, ..., x_T)$, the network updates its memory, otherwise known as the recurrent hidden state h_t, by $\phi(h_{t-1}, x_t) \forall t \neq 1$, where ϕ is a nonlinear function. This memory can be denoted as $h_t = g(Wx_t + Uh_{t-1})$, where g is a bounded function such as a hyperbolic tangent. Upon being updated, some input vectors are inclined to upsurge, making other appear insignificant. The *tanh* activation regulates the output by normalizing the values to an $[-1, 1]$ interval. A standard RNN cell is scarse on internal operations as shown in figure 2 and suffers from short-term memory. In long sequences, the information from earlier steps is failed to be carried over, resulting in values, used to update the weights of a neural network, to become exceedingly small. This problem, where initial layers of the network receive a minimal update and stop learning, is known as the vanishing gradients[45]. As a result, RNN disregards the early inputs and suffers from a short-term memory.

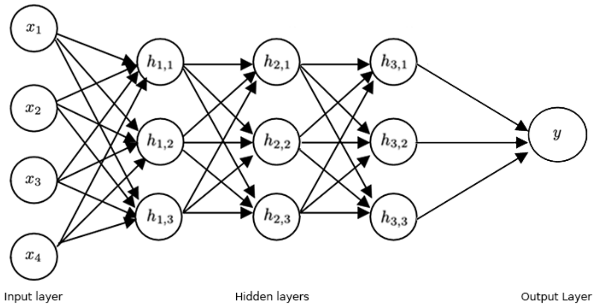


Figure 1: Architecture of RNN

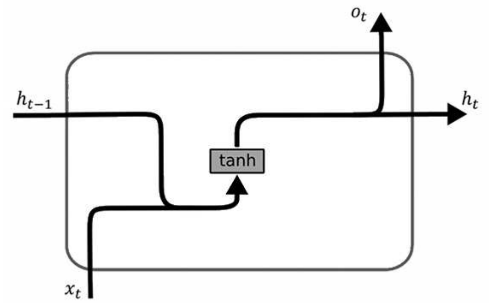


Figure 2: Structure of standard RNN cell

3.3 RNN LSTM

The long short-term memory cell architecture provides a solution to the vanishing gradients problem of standard recurrent neural networks by including internal mechanisms that regulate the information flow. As shown in figure 3, LSTM shares many similarities with a standard RNN, but introduces three gates with sigmoid activations that (partially) retain information c_t at time t. Sigmoid functions are similar to *tanh* functions but regulate to which extent the previous inputs are kept on an $[0, 1]$ interval. A multiplier of 0 makes the network effectively forget earlier information, and 1 preserves it entirely. This dropout rate is regulated by a forget gate f_t with

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + V_f C_{t-1})$$

The current state C_t is recurrently updated by partially forgetting the existing memory and adding new memory content $\tilde{C}_t$: $C_t = f_t C_{t-1} + i_t \tilde{C}_t$, where $\tilde{C}_t = \tanh(W_C x_t + U_C h_{t-1})$. The extent to which the information from the current step is added to the memory cell is modulated by an input gate i_t with

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + V_i C_{t-1})$$

The activation h_t of the LSTM cell can then be denoted as $h_t = \sigma_t \tanh(c_t)$ where σ_t is the output gate that contains data from previous inputs and establishes what the next hidden state should be. It is as follows:

$$\sigma_t = \sigma(W_o x_t + U_o h_{t-1} + V_o C_t)$$

Unlike the standard RNN that overwrites its content at each step, LSTM is able to determine which existing memory should be retained and add additional information. Intuitively, if the LSTM unit detects an important feature at an early stage of the sequence, it is able to capture long-term dependencies by retaining it. The network is compiled using *scikit-learn* and *keras* Python libraries.

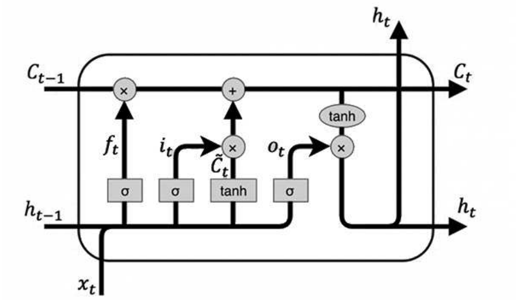
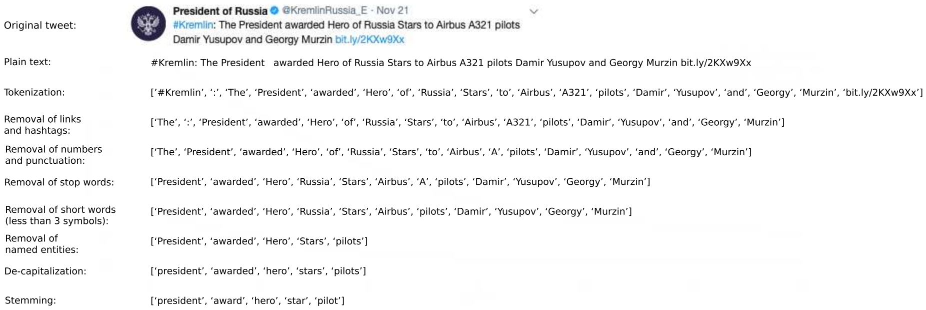


Figure 3: Structure of RNN LSTM cell

3.3 NLP

It is entirely possible to base a forecast on statistical data exclusively, but ignoring the potential information gains from the natural language contained in tweets is imprudent. The natural language tool kit module makes it possible to convert this unstructured data to a format readable by machine with an objective to label the messages based on sentiment and objectivity. First, the classifier is trained on two labeled corpora designed specifically to be used in natural language analysis. The sentiment dataset[46] contains 5331 positively and 5331 negatively labelled sentences, whereas the objectivity dataset[47] contains 5000 objective and 5000 subjective sentences. Consequently, the support vector machine algorithm is tasked to label each tweet based on the knowledge obtained from the labeled corpora. SVM is argued to be a superior alternative to a Naïve Bayes classifier[48] but can exhibit a negative bias due to a higher positive label threshold[49].

The training datasets are structured and preprocessed, whereas the data acquired from Twitter archives is not. An English language-based classifier would not be capable of labeling Russian sentences. Hence, these are translated with Google translate API, inasmuch as compiling Russian training data is beyond the scope of this research. Be that as it may, translated but otherwise unprocessed tweets cannot be supplied to the algorithm if any meaningful insights are to be gained. Accordingly, the sentences are tokenized by words and stripped off punctuation marks, hashtags, and links. Next, the stop words and short words are discarded, as those generally convey little meaningful information. Then, the named entities are removed and other words are de-capitalized. Finally, the remaining words are reduced to their stem, thereby removing any variations. This process is illustrated step-by-step in figure 4 using a tweet recently published on the president of Russia official Twitter page. Consequently, negative or subjective tweets are assigned a value of -1 and positive or objective receive a 1 instead. Lastly, the obtained statistics are aggregated to a daily level.



Original tweet:	President of Russia @KremlinRussia_E · Nov 21 #Kremlin: The President awarded Hero of Russia Stars to Airbus A321 pilots Damir Yusupov and Georgy Murzin bit.ly/2KXw9Xx
Plain text:	#Kremlin: The President awarded Hero of Russia Stars to Airbus A321 pilots Damir Yusupov and Georgy Murzin bit.ly/2KXw9Xx
Tokenization:	['#Kremlin', ':', 'The', 'President', 'awarded', 'Hero', 'of', 'Russia', 'Stars', 'to', 'Airbus', 'A321', 'pilots', 'Damir', 'Yusupov', 'and', 'Georgy', 'Murzin', 'bit.ly/2KXw9Xx']
Removal of links and hashtags:	['The', ':', 'President', 'awarded', 'Hero', 'of', 'Russia', 'Stars', 'to', 'Airbus', 'A321', 'pilots', 'Damir', 'Yusupov', 'and', 'Georgy', 'Murzin']
Removal of numbers and punctuation:	['The', 'President', 'awarded', 'Hero', 'of', 'Russia', 'Stars', 'to', 'Airbus', 'A', 'pilots', 'Damir', 'Yusupov', 'and', 'Georgy', 'Murzin']
Removal of stop words:	['President', 'awarded', 'Hero', 'Russia', 'Stars', 'Airbus', 'A', 'pilots', 'Damir', 'Yusupov', 'Georgy', 'Murzin']
Removal of short words (less than 3 symbols):	['President', 'awarded', 'Hero', 'Russia', 'Stars', 'Airbus', 'pilots', 'Damir', 'Yusupov', 'Georgy', 'Murzin']
Removal of named entities:	['President', 'awarded', 'Hero', 'Stars', 'pilots']
De-capitalization:	['president', 'awarded', 'hero', 'stars', 'pilots']
Stemming:	['president', 'award', 'hero', 'star', 'pilot']

Figure 4: NLTK preprocessing

4 Data

4.1 Background

The availability and consequent acquisition of relevant data is the main concern of any empirical research. Even though "Russia has been at the forefront of trying to shape the online conversation using tools like bots and trolls"[50], and plenty of data is available, determining whether a social media account is governed by a human or bot has proven to be a challenging task[51,52]. Fortunately, a 2018 Twitter transparency report has released archives containing accounts that are believed to be connected to state-backed information operations[53]. The dataset adopted in this paper contains information on 3613 accounts associated with the IRA along with over eight million tweets for a timeframe spanning from 2014 to 2017.

4.2 Variable of interest

The number of tweets generated by the accounts related to the IRA is the main variable of interest in this research. Figure 5 plots its development over a three-year period. Numerous activity surges are clearly visible with a local maximum of 57646 tweets on a day following the shooting of MH17 over Eastern Ukraine. These messages pressed the Ukranian government to "tell the truth" by proclaiming that "Kiev shot the Boeing". The signing of lustration legislation by former president Petro Poroshenko has likewise attracted the agency's attention, and so did the first anniversary of the annexation of Crimea. Tweets advocated the notion that Crimea is rightfully Russian, in spite of various legal issues. The messages are published when the circumstances demand intervention and predicting the output volume may prove challenging without the incorporation of exogenous variables related to the ongoing events. Hence, the array of variables specified below should not be seen as exhaustive, yet adequately comprehensive to establish a solid forecast foundation.

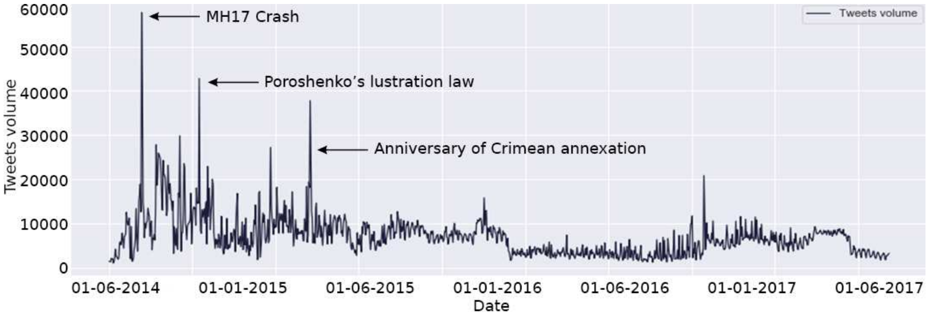


Figure 5: Tweets volume generated by IRA-associated accounts

4.3 Predictor variables

The Twitter dataset aims to provide information on a diverse array of variables, although many are either unavailable or incomplete. While the forecast precision can be positively influenced by including a wider range of metrics, the opposite is just as likely to happen. Accordingly, the research constrains itself by solely including the variables that are complete and are likely to positively influence the model performance.

4.3.1 Accounts

Twitter flagged over three thousand accounts due to their affiliation with the Internet Research Agency. Figure 6 illustrates the course of their creation. As the average output remains stable throughout the period, peaks frequently occur in the days prior to surges of output volume. The maxima are observed in preparation for the first anniversary of the annexation of Crimea and in early July 2014. A possible connection to the crash of the Malaysian Boeing could be hypothesized, but this activity is more likely attributed to the ruling out of ceasefire agreement in war-torn Ukraine and the ensuing Ukraine government's offensive[54]. Numerous lesser pronounced increases are observed throughout the period and all signal greater message output in the near future.

4.3.2 Retweets

Strong public engagement is essential in getting the word of Kremlin across. An average number of retweets measures the number of times that a message has been reposted by a non-IRA account. The plot is provided in figure 6 and similarly displays occasional increases. These are mostly observed in the second half of 2014 following the Zelenopillya rocket attack in July[55], Russian government forces invasion in August, and the escalation of conflict in November. A greater volume of tweets is likely to captivate attention and increase both the total and average number of retweets.

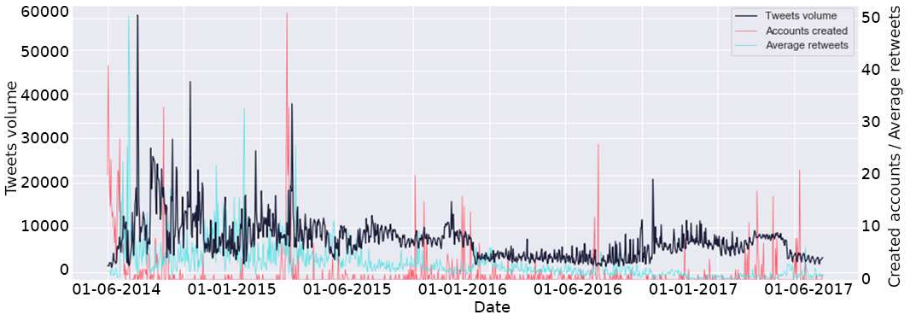


Figure 6: Number of created accounts and average retweets

4.3.3 Sentiment

Classifier-based sentiment captures the emotion inherent to the messages. Figure 7 plots the daily average sentiment values as they struggle to surpass an upper bound of −0.5, thereby signaling that most if not all tweets are labeled as negative. At a glance, the movement exhibits little observable correlation with the output volume, but if there exists any, the neural network would uncover it. The sentiment reaches the local minimum of −0.94 in the aftermath of the airliner shootdown and gradually improves in the following months. Not long after, the values take a plunge as the conflict leads up to the deadliest battle in Ukraine's fight against the separatists[56]. This variable accentuates Russia's engagement in a two-front war with efforts at shaping public opinion often surpassing military engagement. Government propaganda needs not to necessarily be negative in nature however, as reinforcing the public's belief in the righteousness of Crimean annexation is equally misleading as declaring Ukraine responsible for the MH17 tragedy.

4.3.4 Objectivity

Similar to the sentiment, objectivity stems from the classification algorithm. The degree to which tweets are influenced by beliefs and opinions bears relatively greater variance, as witnessed in figure 6. Based on these observations, objectivity is believed to be negatively associated with the number of tweets. The values continuously fluctuate around the zero mark and drop with increases in the output volume. The classifier has correctly labeled the disinformation campaign and like the sentiment, objectivity reaches minimum values in the second half of July 2014. By the same token, a precipitous drop is observed a year after the takeover of the Crimean peninsula. Dependability of the classifier can assuredly be questioned since propaganda is subjective by definition and not all tweets are labeled accordingly. It then must be noted that the IRA does not exclusively spread deceit but also concerns itself with less deceptive reporting.

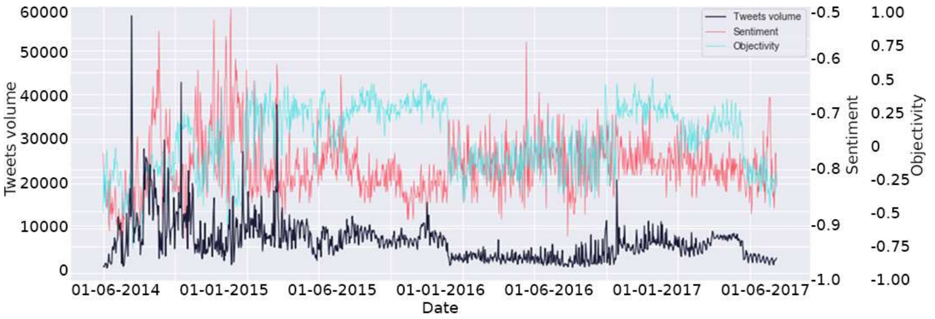


Figure 7: Average values of sentiment and objectivity

4.3 Descriptive statistics

Table 1 summarizes the descriptive statistics of the variables and substantiates previous observations. A high degree of volatility is striking with some values as high as eight times the mean. The average sentiments values remain negative throughout the period and the number of accounts created per day is close to its minimum value but with a pronounced maximum. Similarly, there are large discrepancies between the lowest and highest values in the number of retweets. This array of variables is prone to prove satisfactory for the ensuing forecast.

4.4 Stationarity

Unlike the LSTM model, VARMAX requires variance and correlation structure of the time series to remain unchanged, otherwise known as the notion of weak stationarity[57]. This condition is assessed with an augmented Dickey-Fuller test[58] that determines how strongly a time series is defined by its trend. The null hypothesis states that the series has a time-dependent structure and can, therefore, be represented by a unit root. The alternative hypothesis states that the variables are stationary. Table 2 provides the test statistics with their respective significance values. The series of accounts, retweets, and sentiment appear to be stationary in their original form, whereas tweets and objectivity require first differencing for the condition to hold.

Table 1: Descriptive statistics of the variables

Variable	Average	Median	Std. Dev.	Min	Max
Tweets	7058.93	6430.00	4712.52	1154.00	57646.00
Accounts	1.21	0.00	3.94	0.00	51.00
Retweets	3.18	2.25	3.67	0.11	50.34
Sentiment	-0.79	-0.80	0.06	-0.94	-0.51
Objectivity	0.03	0.05	0.24	-0.77	0.46

Table 2: Augmented Dickey-Fuller test statistics

Variable	Original	1st difference
Tweets	-2.5496	-5.742***
Accounts	-9.6795***	-
Retweets	-3.4755***	-
Sentiment	-4.2649***	-
Objectivity	-2.5222	-5.971***

*p-value<0.1, **p-value<0.05, ***p-value<0.01

5 Results

5.1 VARMAX

Upon compiling a number of VARMAX models, an appropriate lag order is chosen based on the information criteria. Akaike information criterion (AIC)[59], Bayesian information criterion (BIC)[60] and Hannan-Quinn information criterion (HQIC)[61] are considered. Of the three, the AIC is the least strict in penalizing the degrees of freedom loss. Table 3 shows that the model with a lag length of 3 has an overall lowest value and is, therefore, chosen for the analysis.

Table 4 provides the parameter estimates of the model for the output volume of tweets, thereby indicating how well the model fits the data. Every variable has a significant effect in at least one of the three lags. The coefficients are challenging to interpret in their numeric values but if statistically significant, signal the existence of predictive causality, meaning that a particular variable contains useful information for predicting the other, rather than causing it directly.

Table 3: Values of information criteria

Order	AIC	BIC	HQIC
1	28058.82	28285.01	28144.30
2	27608.81	27960.66	27741.76
3	27355.25	27922.37	27625.29
4	27411.73	28084.08	27630.65
5	27444.86	28014.90	27639.66

Table 4: Tweets equation VARMAX estimates

Variable	Lag 1	Lag 2	Lag 3
Tweets	0.599***	0.301	0.130***
	(0.023)	(0.026)	(0.022)
Accounts	-56.491	19.168*	41.043**
	(0.036)	(0.016)	(0.004)
Retweets	109.705	49.895	32.061*
	(24.935)	(31.074)	(28.472)
Sentiment	-803.392**	-2266.069**	-1140.940***
	(385.915)	(436.021)	(380.892)
Objectivity	1960.153**	134.769	-930.898*
	(736.544)	(120.617)	(501.596)

*p-value<0.1, **p-value<0.05, ***p-value<0.01

In line with the previous observations, lower sentiment values are associated with higher tweets volume. This also applies to objectivity, but only in the third lag, whereas the first indicates the opposite effect. The average number of retweets is positively associated with the dependent variable and so it is with its own lagged values. The coefficients of account creation indicate that increase is indeed associated with a higher number of tweets in the following periods. The selection of variables appears suitable for forecasting, which is true for both the statistical, as well as the deep learning model.

In preparation for the forecast, the assumptions underlying the model are assessed with the Ljung-Box test for autocorrelation of residuals, outlined in the methodology. The null hypothesis states that the time series of residuals is white noise and the alternative hypothesis states that the residuals are correlated. Table 5 provides Q test statistics that all appear lower than their respective lag length-related critical values. As there exists no correlation in the residual time series, model assumptions are considered to hold.

Figure 8 plots the development of true tweets output values and contrasts those with the model prediction. The forecast follows the same general trend as the observed output but not without misestimates, especially visible in the 24th time period with a deviation from the true value as high as half of the mean at that point. Similar outliers are not uncommon and the model sets a benchmark for the neural network with RMSE and MAPE values of respectively 673.52 and 17.24%.

Table 5: Ljung-Box test Q statistics

Lag	Q	Critical value
1	1.172	2.706
2	1.241	4.605
3	1.836	6.251

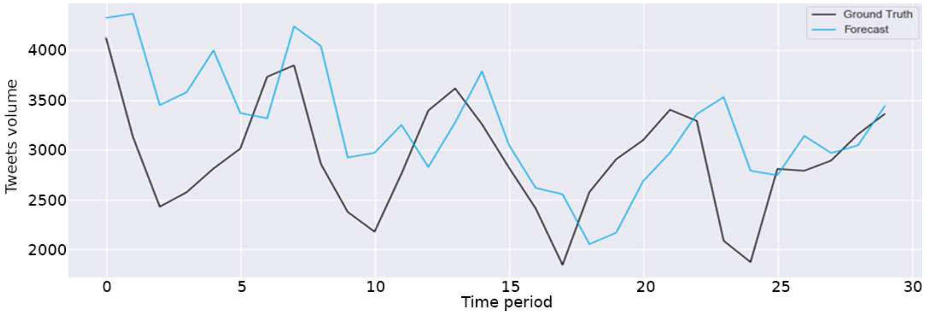


Figure 8: VARMAX forecast

5.2 LSTM

Fitting the data into the format suitable for machine learning and tuning of network's hyperparameters are aspects that can either make or break the model. The network produces a one-day-ahead prediction of tweet volume based on a 30 day rolling window of observations and four predictor variables as features. Below, the criteria of the network are specified. First, the train-validate-test split determines the sequences used to respectively train, validate and test the algorithm. The final month of the time frame is assigned to testing and the remaining data is divided in accordance with the k-fold cross-validation[62]. This method addresses the bias-variance trade-off and requires an input value for k. Empirically, $k = 5$ produces estimates that suffer neither from excessively high bias nor variance[63]. The resulting 78-19-3 split places both local extremes in the training sequence thereby greatly benefiting the learning process.

Epochs define the number of times that the network is presented an entire training set and gets a chance to update its parameters. There is no rule of thumb for the number of epochs, but as long as the validation and training loss keep dropping, the training should continue. The number of hidden nodes, known as neurons, determine the power of the network, but exponentially increase the training time. This parameter is likewise determined by k-fold cross-validation. The batch size determines the number of fragments to split the data into and is adjusted automatically to meet hardware restrictions.

The dropout rate determines the extent to which earlier information is retained and prevents the model from overfitting. An optimal value for hidden layers is between 0.5 and 0.8 and the input layer requires a higher dropout[64]. The learning rate determines how much the weights are updated with a 0.01 baseline[65]. The choice of optimization algorithm may increase or decrease training efficiency and *Adam* is recommended as it compares favorably to other optimization methods[66]. Lastly, the advised loss function for the training of recurrent neural networks is the mean squared error[67].

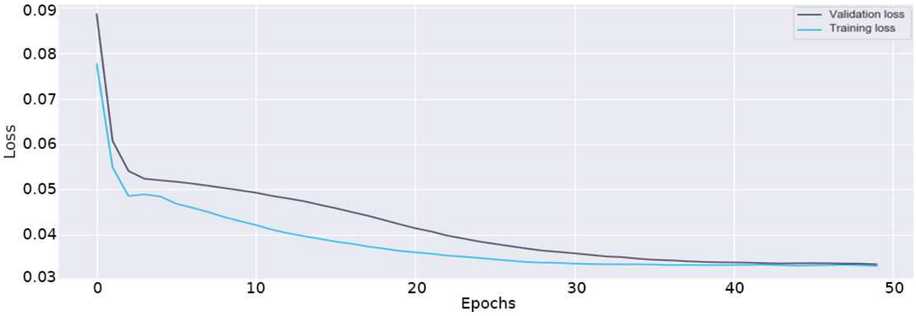


Figure 9: Validation and training loss of RNN LSTM

How well the model fits the data ultimately determines whether it will perform well on a sequence prediction problem. An underfit model achieves high accuracy on the training split, but performs poorly on the validation split. This is observed in a plot where the training loss is lower than the validation loss and trends suggest that additional improvements are conceivable by continued training. Alternatively, the performance could have leveled off without converging, indicating that it is beneficial to update the network's power with additional neurons. An overfit model shows continuously improving performance on the training set but degrading validation performance after a particular epoch. A well-fit model, on the other hand, shows good performance on both sets with decreasing training and validation loss that converge around the same point. Figure 9 illustrates the recalls and establishes the goodness of fit as the losses converge at a sub-0.04 level around the 40$^{\text{th}}$epoch.

Figure 10 plots the development of the volume of tweets produced by the IRA in July of 2017. Unlike three years ago, no extraordinary surges are observable and the number of daily tweets fluctuates around a considerably lower three thousand mark. The general movement appears downward sloping, but the last few days indicate a possible increase that could be related to Russian president's announcement of sending back hundreds of US diplomats in retaliation of announced sanctions[68]. The forecast of RNN LSTM almost always follows the trend set by true observations and tracks its trajectory ever so more closely compared to its statistical counterpart. Similarly to the VARMAX, however, the model fails to predict the trend movement changes prior to their occurrence and appears to be one step ahead, albeit in a negative sense. Considering the inherent unpredictability of the series and the fact that the network could only learn from the past observations during the training period, a forecast RMSE of 483.89 is adequate. Besides, it manages to surpass the baseline by 2.68 percentage points with an 85.44% forecast accuracy. As it stands, the accuracy is comparable to results achieved by other papers in the field that aimed to predict seemingly erratic time series[69], but further improvements are attainable.

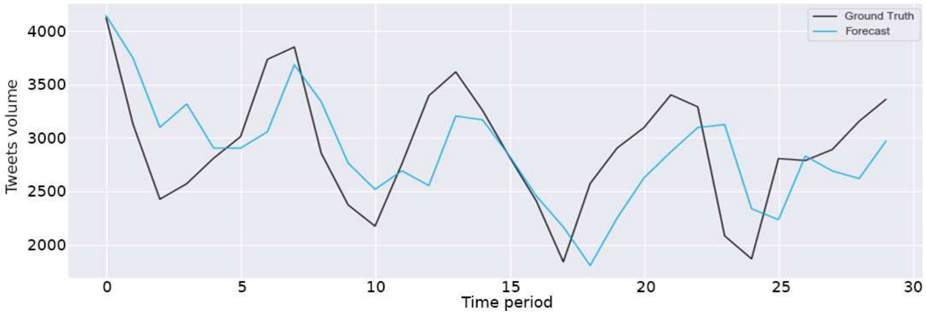


Figure 10: RNN LSTM forecast

6 Extensions

Unlike the traditional forecasting techniques, recurrent neural networks provide ample customization and extension opportunities with alternative architectures and cell structures both in the realm of possibilities. This section considers two extensions that build forth an established foundation and aims to improve the network performance.

6.1 Bidirectional architecture

The forecast trajectory indicates that the model could potentially benefit from an additional information flow in the training period, so it would be able to determine the changes in trend direction prior to it taking place and follow the actual values more closely. This can be done by ensuring that the model learns from both past, as well as the future values during the training. A bidirectional architecture achieves this by duplicating the first recurrent layer and placing it next to the existing one. It provides the original series as input to the first layer and the reversed copy to the duplicated one, while keeping the cell structure intact. The procedure effectively doubles the flow of information and could potentially lead to more efficient training, a better understanding of the problem by the network and ultimately, an improved forecast accuracy.

The model is fitted, tuned, trained, and validated in a procedure similar to that of a standard LSTM and figure 11 plots the forecast. Enabling the network to look into the future during the training has allowed it to uncover previously unobserved patterns and correlations. In contrast to the original, the bidirectional model tracks the true values of the tweets volume far more closely and is able to foresee the trend changes well in advance. Besides, under no circumstance do the forecasted values fail to follow the true trajectory, all the while staying close to the ground truth. The gains in performance are reflected by a greatly increased accuracy of 92.45% with both RMSE and MAPE being nearly half of their initial values at 274.36 and 7.55%, respectively.

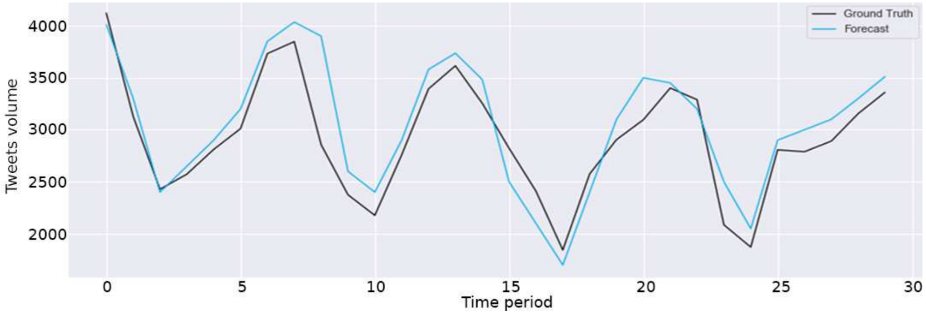


Figure 11: Bidirectional RNN LSTM forecast

6.2 GRU

Gated recurrent unit has a cell structure similar to the LSTM but only includes two gates to regulate the flow of data. An update gate combines the forget and input gates and determines to what extent the current input should be kept and what new information should be added. The reset gate determines how much of the previous information is to be retained on a $[0, 1]$ interval. Zero indicates that the gate is closed and as a result, the cell acts as if it processes the first value of the input sequence allowing it to forget the previously computed state.

The GRU variation is younger and has fewer internal processes, generally making it more efficient to train at the cost of precision. The model is compiled in a process similar to the LSTM and figure 12 plots the forecast. The results are comparable to the original neural network but technically fail to beat the benchmark with a forecast accuracy of 81.71%. However, as the prediction follows the true output values relatively closely, the RMSE is lower than that of the VARMAX at 594.43.

Figure 13 compares the accuracy of the considered models and their variations, each with its own set of advantages and drawbacks. Of note, while the bidirectional LSTM does show a superior forecasting performance, it endures a comparatively longer training time and requires more computational resources.

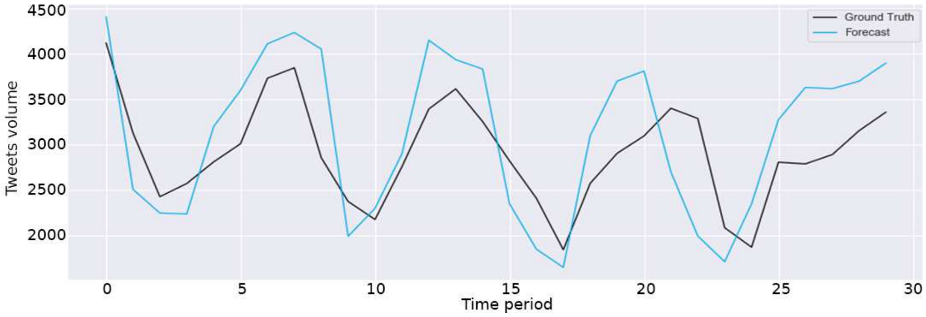


Figure 12: GRU RNN forecast

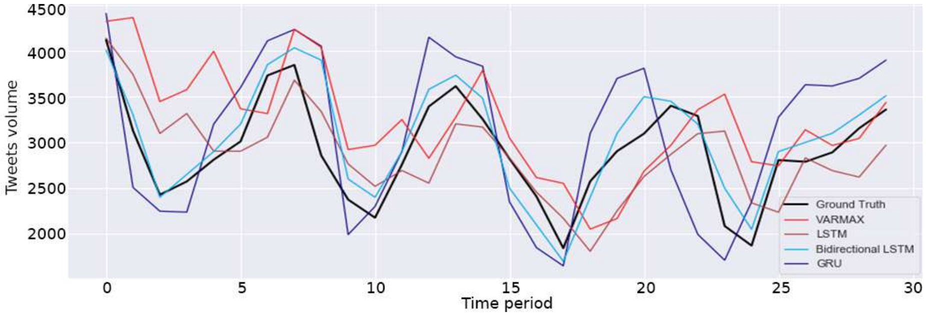


Figure 13: Forecast comparison

7 Discussion and conclusions

7.1 Main findings

This paper has utilized an autoregressive moving average model with exogenous variables and an array of recurrent neural network architectures to predict the development of tweet output produced by the accounts associated with the IRA. The statistical model has set an 82.76% forecasting accuracy baseline with statistically significant contributions of all exogenous variables. The neural network with long short-term memory has surpassed the performance benchmark with a 2.68 percentage point accuracy increase. An addition of bidirectional information flow has further allowed this model to reduce the forecast error and achieve an accuracy of 92.45% with a deviation from the true value of 274.36 on average. The gated recurrent unit has failed to yield any additional improvement with an accuracy of 81.71%, just shy of the baseline. The high degree of accuracy observed in the forecast is made possible by the sentiment and objectivity data extracted from the tweets.

Nonetheless, each framework does come with its own set of limitations. The autoregressive model is restricted to numerical data, demands stationarity and a pre-specified window of lagged observations. Recurrent neural networks, on the other hand, have a far wider range of applications but require an involved finetuning of hyperparameters, rigorous data preprocessing and have a long training time.

This paper simultaneously contributes to the field of machine learning-driven predictive analysis, as well as the research of bots utilized in social interactions. It can conclude that each of the models was successful in modeling and predicting the time series in its own way and proves that the behavior of Twitter bots as captured by the sequence of their output can be forecasted with a relatively high degree of accuracy.

From a practical perspective, were this framework to exist in the turbulent times of the Russia-Ukraine crisis, it would be capable of persistently being one step ahead of the regime's efforts to shape public opinion. This is potentially beneficial, as contrary to the Western world, a sizable part of the Russian community is still unaware of the deceitful government practices. This paper draws attention to their existence, demonstrates their true extent and facilitates the prediction. Forewarned is forearmed and it would make an ordinary man take the deceitful reporting with a grain of salt and make him less likely to fall victim to indoctrination. A nation-wide implementation of the framework is, of course, wishful thinking, considering that the government would certainly not facilitate a scheme directly aimed at counteracting their efforts. That being said, this model would prove to be a valuable asset in the hands of the political opposition.

7.2 Limitations and further research

One of the limiting factors of the framework in its current state is its static nature. In the interest of live analysis and forecast, additional modifications and enhancements need to be considered. Besides, this work should be seen as the first attempt at predicting the behavior of social media bots and is restricted by limited programming knowledge and experience of the author. Likewise, the dataset derived from the Twitter archives could be expanded by identifying potentially harmful accounts as they appear and retrieving associated messages. Besides, the variables related to the ongoing developments that are answered by the tweets, could be included. The provision of additional information is likely to improve the performance of the algorithm and enhance accuracy. Finally, employing the established framework in a different, but otherwise, similar setting could potentially be insightful. Cases of government intervention that shape the political conversation in countries as Egypt or Venezuela could be examined.

8 Acknowledgements

I would like to express my sincere gratitude to my supervisor, Prof. Dr. Olivier Marie for his invaluable guidance and advice. My deep appreciation also goes out to my friends, the student of LSE, Anqi Shan, and the student of ESE, Philip van Moll, for their priceless feedback and suggestions. Besides, I would like to say a heartfelt thank you to my mother for always believing in me and staying by my side throughout this journey. Finally, I'm indebted to all who showed their support and understanding during this challenging period.

References

[1] Richard Pipes. *Russia under the Bolshevik regime.* Vintage, 2011.

[2] Robert Service. *A History of Modern Russia: From Tsarism to the Twenty-first Century.* Harvard University Press, 2009.

[3] Ronald Suny. *The revenge of the past: Nationalism, revolution, and the collapse of the Soviet Union.* Stanford University Press, 1993.

[4] Matthew Lantz. *Russian Election Watch.* Harvard University, 1996.

[5] Blomfield Adrian and Hooper Duncan. *Russian election "neither free nor fair".* 2008. URL: `https://www.telegraph.co.uk/news/worldnews/1580598/Russian-election-neither-free-nor-fair.html`.

[6] Forbes Magazine. *Vladimir Putin.* 2019. URL: `https://www.forbes.com/profile/vladimir-putin/#1b9741f26fc5`.

[7] Richard Stengel. *Choosing Order Before Freedom.* 2007. URL: http://content.time.com/time/specials/2007/article/0,28804,1690753_1690757,00.html.

[8] Natalia Morari. *WCIOM: Corruption in exchange for loyalty.* 2007. URL: https://newtimes.ru/articles/detail/7767/.

[9] WCIOM. *Trust in Politicians.* 2019. URL: https://wciom.ru/news/ratings/doverie_politikam/.

[10] Elena Mukhametshina. *Vladimir Putin's rating has ceased to be invulnerable.* 2019. URL: https://www.vedomosti.ru/politics/articles/2019/01/20/791905-reiting-putina.

[11] Echo of Moscow. *New Year's address of the President of Russia received a lot of negative feedback on the Internet.* 2019. URL: https://echo.msk.ru/news/2344415-echo.html.

[12] Jacques Ellul. *Propaganda.* Knopf New York, NY, 1966.

[13] Richard Alan Nelson. *A chronology and glossary of propaganda in the United States.* Greenwood Publishing Group, 1996.

[14] Sara El-Khalili. "Social media as government propaganda tool in post-revolutionary Egypt". In: *First Monday* 18.3 (2013).

[15] Michelle Forelle et al. "Political bots and the manipulation of public opinion in Venezuela". In: *arXiv preprint arXiv:1507.07109* (2015).

[16] Simon Hegelich and Dietmar Janetzko. "Are social bots on Twitter political actors? Empirical evidence from a Ukrainian social botnet". In: *Tenth International AAAI Conference on Web and Social Media.* 2016.

[17] Eytan Bakshy et al. "Everyone's an influencer: quantifying influence on twitter". In: *Proceedings of the fourth ACM international conference on Web search and data mining.* ACM. 2011, pp. 65–74.

[18] M Conover, J Ratkiewicz, and M Francisco. *Political polarization on twitter.* 2011.

[19] Alessandro Bessi and Emilio Ferrara. *Social bots distort the 2016 US Presidential election online discussion.* 2016.

[20] Yazan Boshmaf et al. "Design and analysis of a social botnet". In: *Computer Networks* 57.2 (2013), pp. 556–578.

[21] Onur Varol et al. "Online human-bot interactions: Detection, estimation, and characterization". In: *Eleventh international AAAI conference on web and social media.* 2017.

[22] Yuriy Gorodnichenko, Tho Pham, and Oleksandr Talavera. *Social media, sentiment and public opinions: Evidence from #Brexit and #USElection.* Tech. rep. National Bureau of Economic Research, 2018.

[23] Massimo Stella, Emilio Ferrara, and Manlio De Domenico. “Bots sustain and inflate striking opposition in online social systems”. In: *Proceedings of the National Academy of Sciences* (2018).

[24] Denis Stukal et al. “Detecting bots on Russian political Twitter”. In: *Big data* 5.4 (2017), pp. 310–324.

[25] Warren McCulloch and Walter Pitts. “A logical calculus of the ideas immanent in nervous activity”. In: *The bulletin of mathematical biophysics* 5.4 (1943), pp. 115–133.

[26] John J Hopfield. “Neural networks and physical systems with emergent collective computational abilities”. In: *Proceedings of the national academy of sciences* 79.8 (1982), pp. 2554–2558.

[27] Scott E Fahlman. “The recurrent cascade-correlation learning algorithm”. In: *Advances in neural information processing systems* 3 (1991), pp. 190–196.

[28] Kevin J Lang, Alex H Waibel, and Geoffrey E Hinton. “A time-delay neural network architecture for isolated word recognition”. In: *Neural networks* 3.1 (1990), pp. 23–43.

[29] Sepp Hochreiter and Jürgen Schmidhuber. “Long short-term memory”. In: *Neural computation* 9.8 (1997), pp. 1735–1780.

[30] Junyoung Chung et al. “Empirical evaluation of gated recurrent neural networks on sequence modeling”. In: *arXiv preprint arXiv:1412.3555* (2014).

[31] Yanting Li, Yan Su, and Lianjie Shu. “An ARMAX model for forecasting the power output of a grid connected photovoltaic system”. In: *Renewable Energy* 66 (2014), pp. 78–89.

[32] Abdüsselam Altunkaynak. “Forecasting surface water level fluctuations of Lake Van by artificial neural networks”. In: *Water resources management* 21.2 (2007), pp. 399–408.

[33] Mustafa Akal. “Forecasting Turkey’s tourism revenues by ARMAX model”. In: *Tourism Management* 25.5 (2004), pp. 565–580.

[34] S Anbazhagan and Narayanan Kumarappan. “Day-ahead deregulated electricity market price forecasting using recurrent neural network”. In: *IEEE Systems Journal* 7.4 (2012), pp. 866–872.

[35] Jung-Hua Wang and Jia-Yann Leu. “Stock market trend prediction using ARIMA-based neural networks”. In: *Proceedings of International Conference on Neural Networks (ICNN’96)* 4 (1996), pp. 2160–2165.

[36] Alex Graves and Jürgen Schmidhuber. “Offline handwriting recognition with multidimensional recurrent neural networks”. In: *Advances in neural information processing systems* (2009), pp. 545–552.

[37] Santiago Fernández, Alex Graves, and Jürgen Schmidhuber. “An application of recurrent neural networks to discriminative keyword spotting”. In: *International Conference on Artificial Neural Networks* (2007), pp. 220–229.

[38] Matthew B Hoy. “Alexa, Siri, Cortana, and more: an introduction to voice assistants”. In: *Medical reference services quarterly* 37.1 (2018), pp. 81–88.

[39] Brett Duncan and Yanqing Zhang. “Neural networks for sentiment analysis on Twitter”. In: *2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing* (2015), pp. 275–278.

[40] Xiaofeng Wang, Matthew S Gerber, and Donald E Brown. “Automatic crime prediction using events extracted from twitter posts”. In: *International conference on social computing, behavioral-cultural modeling, and prediction* (2012), pp. 231–238.

[41] Bo Pang, Lillian Lee, et al. “Opinion mining and sentiment analysis”. In: *Foundations and Trends® in Information Retrieval* 2.1-2 (2008), pp. 1–135.

[42] George EP EP Box, Gwilym M Jenkins, and G Reinsel. “Time series analysis: forecasting and control Holden-day San Francisco”. In: *BoxTime Series Analysis: Forecasting and Control Holden Day1970* (1970).

[43] Edward Loper and Steven Bird. “NLTK: the natural language toolkit”. In: *arXiv preprint cs/0205028* (2012).

[44] Greta M Ljung and George EP Box. “On a measure of lack of fit in time series models”. In: *Biometrika* 65.2 (1978), pp. 297–303.

[45] Yoshua Bengio, Patrice Simard, Paolo Frasconi, et al. “Learning long-term dependencies with gradient descent is difficult”. In: *IEEE transactions on neural networks* 5.2 (1994), p. 1994.

[46] Bo Pang and Lillian Lee. “Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales”. In: *Proceedings of the 43rd annual meeting on association for computational linguistics* (2005), pp. 115–124.

[47] Bo Pang and Lillian Lee. “A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts”. In: *Proceedings of the 42nd annual meeting on Association for Computational Linguistics* (2004), p. 271.

[48] Johan AK Suykens and Joos Vandewalle. “Least squares support vector machine classifiers”. In: *Neural processing letters* 9.3 (1999), pp. 293–300.

[49] Chu-Hong Hoi et al. “Biased support vector machine for relevance feedback in image retrieval”. In: *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541)* 4 (2004), pp. 3189–3194.

[50] Joshua Tucker. *How Do You Spot a Russian Bot? Answer Goes Beyond Kremlin Watching.* 2017. URL: https://www.nyu.edu/about/news-publications/news/2017/december/how-do-you-spot-a-russian-bot--answer-goes-beyond-kremlin-watchi.html.

[51] Zi Chu et al. "Who is tweeting on Twitter: human, bot, or cyborg?" In: *Proceedings of the 26th annual computer security applications conference.* ACM. 2010, pp. 21–30.

[52] David M Cook et al. "Twitter deception and influence: Issues of identity, slacktivism, and puppetry". In: *Journal of Information Warfare* 13.1 (2014), pp. 58–71.

[53] Twitter. *Transparency report.* 2018. URL: https://transparency.twitter.com/en/information-operations.html.

[54] BBC News. *Ukraine crisis: Donetsk rebels in mass withdrawal.* 2014. URL: https://www.bbc.com/news/world-europe-28177020.

[55] BBC News. *Ukraine conflict: Many soldiers dead in 'rocket strike'.* 2014. URL: https://www.bbc.com/news/world-europe-28261737.

[56] BBC News. *Ukraine's deadliest day: The battle of Ilovaisk, August 2014.* 2019. URL: https://www.bbc.com/news/world-europe-49426724.

[57] Ionut Florescu. *Probability and stochastic processes.* John Wiley & Sons, 2014.

[58] David A Dickey and Wayne A Fuller. "Distribution of the estimators for autoregressive time series with a unit root". In: *Journal of the American statistical association* 74.366a (1979), pp. 427–431.

[59] Hirotugu Akaike. "A new look at the statistical model identification". In: *Selected Papers of Hirotugu Akaike* (1974), pp. 215–222.

[60] Eduardo S Schwartz. "The stochastic behavior of commodity prices: Implications for valuation and hedging". In: *The journal of finance* 52.3 (1997), pp. 923–973.

[61] Edward J Hannan and Barry Quinn. "The determination of the order of an autoregression". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 41.2 (1979), pp. 190–195.

[62] Payam Refaeilzadeh, Lei Tang, and Huan Liu. "Cross-validation". In: *Encyclopedia of database systems* (2009), pp. 532–538.

[63] Gareth James et al. *An introduction to statistical learning.* 2013.

[64] Nitish Srivastava et al. "Dropout: a simple way to prevent neural networks from overfitting". In: *The journal of machine learning research* 15.1 (2014), pp. 1929–1958.

[65] Yoshua Bengio. “Practical recommendations for gradient-based training of deep architectures”. In: *Neural networks: Tricks of the trade* (2012), pp. 437–478.

[66] Diederik P Kingma and Jimmy Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014).

[67] Russell Reed and Robert J MarksII. “Neural smithing: supervised learning in feedforward artificial neural networks”. In: (1999).

[68] BBC News. *Russia’s Putin orders 755 US diplomatic staff to be cut.* 2019. URL: `https://www.bbc.com/news/world-europe-40769365`.

[69] Rui Fu, Zuo Zhang, and Li Li. “Using LSTM and GRU neural network methods for traffic flow prediction”. In: *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)* (2016), pp. 324–328.