A New Approach to Social Engineering with Natural Language Processing : RAKE

Ahmet Furkan Aydogan
axa184@shsu.edu
Dept. of Computer Science
Sam Houston State University
Huntsville 77341 TX - USA

Min Kyung An
an@shsu.edu
Dept. of Computer Science
Sam Houston State University
Huntsville 77341 TX - USA

Mehmet Yılmaz
mehmet@sobiad.com
Asos Information Technologies Company
Firat University Technopole
Elazig - Turkey

Abstract—**Nowadays, with the advancement of technology, the way of thinking of communities has become easily manipulated. Scientific results put forward during the Arab Spring period, which is one of the biggest examples, show that social media is easily used to guide people in different ways. The Rake project has emerged as a method against the aforementioned manipulations. The Rake project, created using Natural Language Processing and Machine Learning algorithms, creates its own dataset and implements the steps in the study. Consisting of three different stages, Rake is a high performance application that can adapt to areas such as community analysis, business area analysis and advertising. As shown in the experiments, it has achieved a machine learning performance of 94%. In addition, the stages that convey how to create the Rake application in large-scale projects were included in the study. Bringing a new perspective to social engineering, Rake can be considered as a starting point for more comprehensive studies.**

Keywords: Social Engineering, Natural Language Processing , Machine Learning, Manipulation Techniques.

I. INTRODUCTION

One of the strangest headlines brought about by technological developments has been the rapid transmission of the socialization numen to online systems. First, the digital relationship building platform GENie (General Electric Network for Information Exchange), which emerged at the beginning of 1985, was laying the trigger seeds of a great formation with 350,000 users [1]. The application named ListServ was intended to establish a pre-business network in 1986, allowing register with an e-mail address and gain access too many e-mail addresses to identify user skills. In 1988, the idea of Internet Relay Chats (IRCs) emerged, making it the first application to share files between users [2]. Its main goal was to keep users in constant communication. The working principle was that the e-mail system provided ease of answering instantly. By 1994, The Place adorned the same system with more graphics, allowing users to create avatars. In 1997, a social media platform called SixDegrees was born. SixDegrees, a great pioneer in shaping today's social media logic, has been a successful application that allows users to create a friend list and accumulates the features of previously created live chat systems [3].

It was obvious that it was a great idea for two people who did not know each other to communicate using only the internet, but the website MoveOn, which emerged in 1998, would change the principle of socialization until now. Even if the application called MoveOn appeared as an open list used to access e-mail addresses, as in the previously mentioned ListServ system, which would soon change. A new movement was about to emerge, starting with the constant sending of e-mails to users opposing Bill Clinton's dismissal. MoveOn would then continue to be the haunt of activists who opposed US actions abroad, and it was the first time that internet activism would emerge [4].

Today, many social media platforms have come to light. With the number of users expressed in billions, social media platforms would prove to the whole world how powerful the digital activism movements that started with MoveOn became in 2011. Social media would be used as a trigger for the start of a string of events that would affect certain parts of the Middle East and Africa, the events that would be called the Arab Spring or the digital revolution [5]. Although the internet usage rate of 4.196.564 (40%) of the population of 10.732.900 is known in Tunisia, the country where the digital revolution was first observed, thousands of people took to the streets as a result of the spread of the news on social media platforms after an unemployed youth burned himself and caused the current authority to change [6]. The second address of the digital revolution was Egypt. The state of Egypt, which has an internet usage rate of 29.809.724 (35.6%) of the population of approximately 83,688,164, has formed the continuation of the process that started in Tunisia. The fact that Egypt was the most used word on Twitter in 2011 shows how high social media is used. In addition, as a precaution, the Egyptian government has tried to reduce the activity towards people to some extent by lowering internet speeds [7]. Libya, the third address of the digital revolution, followed a low level compared to other countries with the use of social media platforms of 954,275 of its 5,613,380 population, but it published the moment of death of the president on the aforementioned platforms and then a large increase in social media usage was observed [8]. In Syria, the last address of the

digital revolution, 5,069,418 of the population of 22,530,746 people use social media platforms and the majority of the mentioned population is the part against the dictatorial regime. Especially, with the emergence of the news on social media platforms that UN representatives and press members were prevented from being admitted to the country, the internal turmoil has increased to the current level [9].

By 2014, a Cambridge University Professor named Aleksandr Kogan developed an application and aimed to create profiles about the US voters. As a result of the surveys conducted on Facebook, sentiment analysis of approximately 50 million people was made. As a result of the Cambridge Analytica company using the aforementioned data in the US elections, the debate that will continue to the present day has been ignited [10]. Today, many web pages have been created as a continuation of the MoveOn movement and state authorities have been forced to meet the demands of the people in a short time. By positively shaping the movement of digital activism, which is haunted by state authorities, it possible to remove manipulations and build a mentally happy public base. This article analyzes social media on behalf of the adoption by the people of the basic state policy, and in particular the Gen Z overseas activities in a country for creating the most appropriate forms of expression foreign public and state authorities is a project that informs the social activities at regular intervals. While performing its operations, it maps the emotions, reactions and words of social media accounts.

II. LITERATURE REVIEW

Unfortunately, it is not possible to expect the projects to analyze social media platforms to be published in academic fields in full, because they are generally used by large countries in the field of intelligence. However, there are articles that examine social media analysis under different titles and present them to academic media. One of them is the work called We Feel: Mapping emotion on Twitter. For a year, the team working on Twitter collected data with the API they provided to users, and using the data they collected, they were successful in removing the emotions of users such as love, anger, jealousy, depression, and hope. The app, which collects 22,000 Tweets per minute, predicted that some events over the US would be canceled as a sign of its success [11].

Mapping social activities and concepts with social media (Twitter) and web search engines (Yahoo and Bing): a case study in 2012 another article called US Presidential Election examined search engines and social media platforms. Before the elections, researches were conducted on the analysis of the geographical distribution of the voters and their activities. As in the previous study, the study created using APIs successfully eliminated the names of the candidates over 16.751.33 Tweets and created a distribution map. As an additional area, the word mapping application provided visuals showing what kind of words voters used on behalf of the candidates in which region [12].

Another article called R-Map: A Map Metaphor for Visualizing Information Reposting Process in Social Media was used to calculate the interaction rates of shares on social media platforms. First of all, it was aimed at predicting the impact of the update, which will be sent thanks to the data provided, on people with different opinions and social levels. After making its predictions, the project, which uses visuals that are visually related to the real world, has been a good example of the potential success that can be achieved when ML and AI algorithms are used effectively [[13].

The study, named Mapping the progress of social media research in hospitality and tourism management from 2004 to 2014, revealed the expectations from the tourism sector as a result of the research conducted on approximately 32 thousand people on social media. In particular, the data obtained from social media platforms were divided into small groups and some information about which service should be developed further. As an example, food and accommodation topics were examined separately and some problems in both areas that could potentially pose a problem for the young generation were revealed. The important aspect of the article is that it scientifically demonstrates the necessity of improving the responses of social media platforms based on people's decision mechanisms, using social media analysis [14].

The article titled The Adoption of social media analytics for crisis management - Challenges and Opportunities contains studies on the attitudes of social media platforms in times of crisis. As a result of interviews with experts, especially in the field of fire, the positive and negative aspects of the reactions of social media platforms during a certain crisis have been the subject of research [15].

The article titled Text Analytics in Social Media mentions that word mapping will provide a great advantage in obtaining information from social media platforms when used with the NLP method. Especially, the article stating that the tag system provides a great advantage is complaining about too much data as a negative return of social media as a source. In order to eliminate the negative aspects, the article stated that the data to be processed should be selected well [16].

Some words such as Machine Learning (ML) and Artificial Intelligence (AI) have started to take up more and more of our lives day by day. The two potentially mentioned topics will undoubtedly be used in projects that will affect our entire lives in the coming years. Another reality is that using ML and AI technologies it is possible to analyze human populations and create a community resistant to internal and external manipulations. The Rake Project fully helps government agencies to create this resilient society.

III. A NEW APPROACH TO SOCIAL ENGINEERING WITH NATURAL LANGUAGE PROCESSING : RAKE

The Rake is basically a project that analyzes social media organs for the adoption of state policies by the public, helps to create the most appropriate way of expression among the countries, especially to the Z-Generation, and to the foreign people in their overseas activities, and informs the state authorities about social activities at regular intervals. While performing its operations, it maps the emotions, reactions and words of social media accounts.

A. Emotion Mapping

Emotion map creates reports of the nerve endings of societies and analyzes the potential problems that state institutions will face in the future. While performing the Rake operations, it can make comprehensive analysis from the smallest group of people to the whole country. For example, a nuclear power plant to be placed in the Black Sea region will naturally be viewed negatively by the Black Sea people due to the Chernobyl event. The Rake project is used to report daily requests made by the public to government channels against foreign manipulator institutions that make the most preferred emotion mapping in countries. In terms of intelligence, it helps to produce counter-arguments by producing public statistics about countries that make negative propaganda against the homeland, primarily neighboring countries.

B. Reaction Mapping

Reaction map analyzes how and to what extent the negative and positive thoughts on an issue by human populations and how and to what extent human communities react to policies For example, what positive or negative reactions is the World Health Organization (WHO) exposed to by the public? Usually, the first 3 negative or positive perspectives are important when doing reaction mapping. As will be mentioned later, reaction mapping, which can be used to draw regional inferences, specifically examines the effects of domestic policies on the public. However, as a negative feature, reaction mapping becomes available for examination after the implementation of policies. The aforementioned negative feature, on the other hand, can be turned into positive by guiding the decisions to be taken by the state channels in the later stages. In terms of intelligence, it analyzes how effective the reports and news originated abroad are and by using which methods the negative effects can be turned into positive.

c) Word Mapping

Word mapping is used to analyze the shares of individual social media accounts and report the most used words and supported headings. As an example, the most commonly used words by Bill Gates are Africa, Climate Change, Sickness and Energy. As can be guessed, the aforementioned main topics can be used in social engineering or social engineering processes in order to approach the masses appropriately. Additionally, reports on social media content in regional areas can be submitted. For example, before a park area planned to be created by the government channels in the Houston region, it can be inferred as to how much negative or positive support it receives with social media analysis. The Rake project, can be used to structure speech texts created by politicians by analyzing the words that are most used by individual users and can have an important place in government policies. For example, if, as a result of daily reports, a significant increase is observed in the use of the word "Racist", the word specified by the politicians can be removed from the text of that day. In addition, as mentioned above, before the regional studies, the reaction of the society can be evaluated and measures can be taken against possible problems or the service quality can be increased regionally. In terms of intelligence, prior to visits to neighboring countries, titles that have a reputation in the eyes of the public or the reporting of words that will have a positive impact can be submitted to government channels.

In order to put the three main headings mentioned above into action, four different work packages have been put forward and detailed. The correct and complete execution of the work packages is the main element of the Rake project.

IV. WORK PACKAGES OF RAKE PROJECT

A. Academic Infrastructure and Feasibility Reporting

It aims to create the financial reporting that will emerge in the creation of the Rake project and to work more efficiently by having an academic scientific background. In order to create an academic infrastructure, it is necessary to make reports and inferences after visiting universities with advanced education levels determined throughout the country, interviews with academicians.

Tasks to be Performed within the Scope of the Work Package :

1) *Social Analysis Rates*
2) *Average Data and Storage Accounts*
3) *Financial Average of the Data Center to be Created*
4) *UPS and Accent Design of the Created Data Center*
5) *Cooling and Installation*
6) *Performing Legal Examinations of the Whole System*
7) *Searching for Employment Opportunities*

As a purpose of under the aforementioned title, by reporting the researchers conducted in an academic language and order and submitting them to the state media, creating an infrastructure to be used in future projects.

B. Steps for Emotion Mapping

With the reports that will be provided to the state channels as soon as possible, the emotion mapping made by foreign organizations on the homeland is reported daily to determine the impact areas and evaluate them for analysis. In the future, it is a topic that aims to analyze and report the titles that affect or show the potential to affect external relations.

Tasks to be Performed within the Scope of the Work Package :

1) *Determining the organizations that make emotion mapping and designing appropriate algorithms.*
2) *Determination and implementation of appropriate technologies for the realization of algorithm plans.*
3) *Creating encryption methods to be developed for the safe transmission of reports.*
4) *Following the technical work and performing the daily maintenance and tests required to fix any problems.*
5) *Reporting the political, legal, economic and social aspects of the reports to be prepared about the analysis of the reports obtained and their social effects.*

C. Steps for Reaction Mapping

It is a step that aims to analyze the effects of the policies that are expected to come into force in the future by the state channels or that are currently in effect on the public and to present them in the form of a daily report. In a later stage, it is a title that can generate reports on how foreign policies have been met by foreign people.

Tasks to be Performed within the Scope of the Work Package :

1) *Determining the necessary social media channels for reaction mapping and designing appropriate algorithms.*
2) *Determination and implementation of appropriate technologies for the realization of algorithm plans.*
3) *Execution of previously created secure communication protocols.*
4) *Determining the suitability of the data obtained by making statistical inferences.*
5) *Following the technical works and performing the daily maintenance and tests required to solve the problems that may occur.*
6) *Reporting the political, legal, economic and social aspects of the reports to be prepared about the analysis of the reports obtained and their social effects.*

D. Steps for Word Mapping

The main goal of the title is to create an automation that can create understandable, impressive and most appropriate texts in the eyes of the public by using social media channels to daily analysis, in the creation of keywords for the topics preferred by the public and in the editing of texts created by state institutions. In the later stages, it aims to implement the three main factors (understandable, impressiveness and appropriateness) mentioned above by using the system in external relations.

Tasks to be Performed within the Scope of the Work Package

1) *Determining the necessary social media channels for word mapping and designing appropriate algorithms.*
2) *Determination and implementation of appropriate technologies for the realization of algorithm plans.*
3) *Execution of previously created secure communication protocols.*
4) *Listing the new words and expressions preferred in social media areas by analyzing the data obtained.*
5) *Following the technical works and performing the daily maintenance and tests required to solve the problems that may occur.*
6) *Reporting the political, legal, economic and social aspects of the reports to be prepared about the analysis of the reports obtained and their social effects.*

V. EXPERIMENTAL DESIGN AND RESULTS

Rake is a project that can generate high costs as can be expected. In order to envision the financial report in the following areas, an initial implementation for the project and examples of the potential success of the project are given.

A. Creating Dataset

The Rake project aims to give more descriptive results about its features using a data set of its own. At this stage, the mentality and reactions of another country and the strategy followed will be explained. While creating the dataset, Turkey was chosen as a subject and used to create the data, "Mynet", the most widely used news website in the country [17].

1) Obtaining Links from Mynet's Twitter Page with Twint: At the beginning of the project, it has been observed that the website has implemented a strategy of hiding news links. However, it has been observed that the links on the news site are open to sharing on Twitter. Twint application, which is used by the Rake project to collect data, helped in receiving news links on Mynet's Twitter page. Twint application, as an open source application, allows to obtain data via Twitter [18]. Twint has the capacity to automatically file the shares of the

profiles as well as the location, phone, links on tweet, numbers, followed and followed profiles.

Algorithm 1 : Link Scratch Algorithm Mynet's Twitter Page

```
1: import twint
2: Config = twint.Config ()
3: Config.Username = "Mynet"
4: Config.Custom["user"] = ["bio"]
5: Config.Custom["url"] = ["url"]
6: Config.Limit = 1000
7: Config.Store_csv = True
8: Config.Output = "tweets.csv"
9: twint.run.Search (Config)
```

---

2) Obtaining Data from URLs:

After the Tweets containing the news links of the Mynet website were shot, the data obtained were transferred to a CSV file, and then the news content, the news headline, and the reactions of the readers were obtained.

Reactions obtained from users are presented with angry, unapproved, dislike, happy, approved, like, laughter and shock emoji. Laughter and shock emojis are not included in the list because it is hard to understand whether it is a negative or a positive reaction. Angry, unapproved, and dislike expressions were marked negative, while the remaining emojis were marked as positive. While making inference, the difference in number between the total negative and positive expressions themselves was used.

Algorithm 2 : Taking Data from Mynet

```
 0: Targets ⇐ Text - Reactions
 1: Inputs : URLs.txt ⇐ List of Mynet Links
 2: Outputs : Text[] , Reactions [] ⇐ From Web Page
 3: for lines ⇐ URLs.txt do
 4:     Connecting Server ⇐ Request Library
 5:     Parsing Content ⇐ BeautifulSoup Library
 6:     Article_Content[] ⇐ Parsing Content [i]
 7:     Article_Content(Split) ⇐ Parsing
 8:     Article_Text ⇐ Article_Content[ii:0]
 9:     Reactions [Article_Text[0:i]] ⇐ Taking Reactions
10:     Reactions[] ⇐ Drop Reactions[,:6:8]
11:     Counter_Positive ⇐ 0
12:     Counter_Negative ⇐ 0
13:     Counter_Positive += Reactions[,:0:-3]
14:     Counter_Negative += Reactions[,:3:6]
15:     if Counter_Positive > Counter_Negative then
16:         Article_Text[,:0] ⇐ "Positive"
17:     end if
18:     if Counter_Positive < Counter_Negative then
19:         Article_Text[,:0] ⇐ "Negative"
20:     end if
21: end for
22: Total = [] ⇐ Sum of Data
23: Total.Append(Article_Text and Reactions)
```

---

3) Translate Obtained Text with NLP:

The links obtained with Twint were used to record the news content and reactions on the web page. Language separation can be given as a weak point of the algorithm that allows the content and reactions obtained in the last stage to go to the examination stage. The contents of the web addresses belonging to Turkey, which is the target country, must be translated into English. However, today's translate algorithms do not yield very stable results.

The biggest reason why text translation algorithms do not work stably is due to spelling rules and errors. In reality, while the correct words and sentences written according to the rules are translated successfully, it is possible to encounter big problems even with the smallest spelling mistakes. Punctuation marks and grammar rules, which can be ordered differently in each language, are among the problems encountered by text translation algorithms. At this stage of the project, it is not possible to avoid the aforementioned problems. However, Natural Language Process (NLP) algorithms can be used to detect the specified problems and make improvements [19].

NLP is a discipline that allows processing and editing of written data using algorithms. While performing NLP operations, it is subjected to many sub-processes. The first of the NLP directives used in two parts in the Rake project is to ensure compliance of the texts obtained in Turkish with the language rules. Tokenization, Stemming, Lemmatization, Stop-Words, and Tagging processes were used in the adaptation stages of Turkish texts.

1) Tokenization

Tokenization offered by Natural Language Tool-Kit can be applied to words or phrases. Natural Language Toolkit (NLTK) is a library in the Python programming language [20]. NTLK is a powerful library for studying and processing human language. In addition, NLTK features include classification, tokenization, stemming, tagging, parsing, and semantic reasoning. Finally, NLTK contains close to 50 corpus and can be accessed free of charge. Word tokenization is used to break up words in a sentence [21]. The word tokenize method is applied in Rake application. After the tokenization process, it is easier to develop the words by the methods mentioned in the following headings.
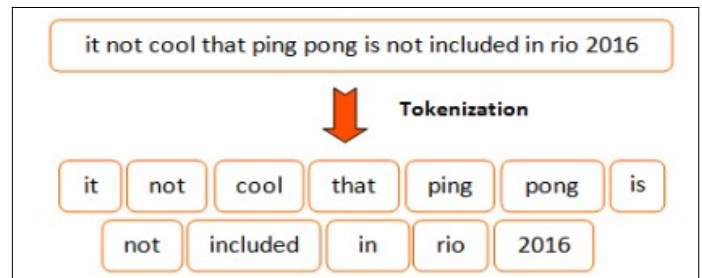

Fig. 1: Illustration of Tokenization Processing.

2) Stemming

Stemming algorithms have been used in the computer science department since the 1960s. Used to access the root of

words. Accessing the roots of words will provide additional performance and accuracy for Machine Learning [22].
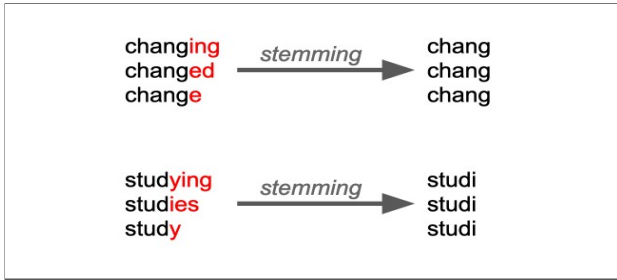

Fig. 2: Illustration of Stemming Processing.

3) Lemmatization

Lemmatization, although ambiguity with Stemming, is different. Lemmatization algorithms take care of preserving the meaning of words. Also, the good Lemmatization algorithm selects the most meaningful word by checking its predecessor and goes to its root. After the use of the stemming algorithm,it is beneficial to perform Lemmatization process [23].
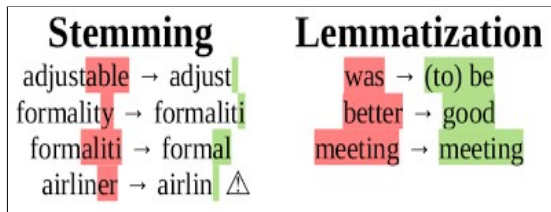

Fig. 3: Illustration of Lemmatization Processing.

4) Stop-Words

Stop words is another commonly used pre-processing method in Natural Language Processing. NLTK has a Stop-Word list of 16 different languages. Stop Words is used to extract the most used words in a language. In addition, many search engines use this method in terms of performance. It has an important place for Machine Learning algorithms because it eliminates a large load [24].
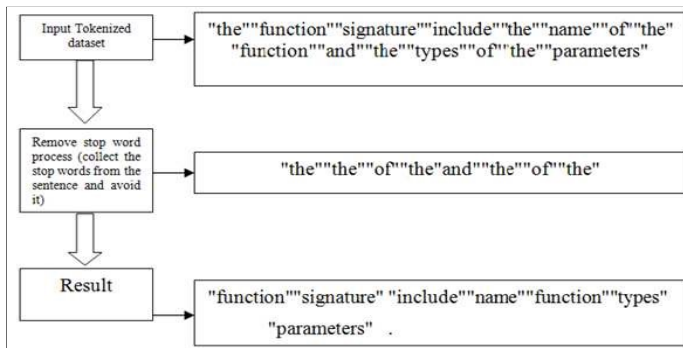

Fig. 4: Illustration of Stemming Processing.

After the above-mentioned processes, the text is now at the stage of translating it into English language. Text translation process was performed using Google API. After receiving the translated text into English, Rapid Automatic Keyword Extraction library was used to identify the keywords and outputs were obtained as a summary of the entire text. In the last stage, the keywords of a news and reactions to the news are now left. The aforementioned data is recorded in a .CSV file and made ready for Machine Learning stages.

Algorithm 3 : Translation of Obtained Text

1: Inputs : Total[] ⇐ List of Text and Reactions
2: Outputs : Translated_Text[] , Reactions [] ⇐ Total_Translated []
3: **for** lines ⇐ Total[i:0] **do**
4:     Sentences ⇐ Total[i:0]
5:     Sentences ⇐ Turkish Spell Checker ⇐ Zemberek Library
6:     Sentences ⇐ Turkish Sentence Normalizer ⇐ Zemberek Library
7:     Sentences ⇐ Turkish Sentence Extractor ⇐ Zemberek Library
8:     Sentences ⇐ Turkish Sentence Morphology ⇐ Zemberek Library
9:     Sentences ⇐ Turkish Sentence Tokenizer ⇐ Zemberek Library
10:     Translator ⇐ Google_Translator Library
11:     Sentences ⇐ Translator ⇐ TR TO ENG
12:     Sentences ⇐ Tokenization ⇐ NLTK Library
13:     Sentences ⇐ Stemming ⇐ NLTK Library
14:     Sentences ⇐ Lemmatization ⇐ NLTK Library
15:     Sentences ⇐ Stop Words ⇐ NLTK Library
16:     Sentences ⇐ Punctuation ⇐ String Library
17:     Sentences ⇐ Rapid Automatic Keyword Extraction
18:     Translated_Text[Sentences, Total[:i]]
19: **end for**
20: Translated_Text = [] ⇐ Sum of Data with Translated Text - Keywords
21: Dataset.CSV ⇐ Translated_Text = []

---

4) Applying Machine Learning Algorithms to Dataset:

After the above processes, the dataset was created and the equivalent of each keywords was equalized by the public whether it was negative or positive. Among the Machine Learning algorithms used, Logistic Regression, SVM, and K Neighbors Classifier were used. Before using algorithms, NLTK operations were repeated and finally, numerical relationships between words were revealed using TF-IDF Vectorizer. The next steps are to go through the final checks of the data set and make experiments with three different machine learning algorithms after 80% of it is machine learning and the remaining 20% is used as prediction. The outputs of the results of the algorithms are given below;

Fig. 5: Result of Logistic Regression Model.

```
KNeighbors Classifier
                precision    recall  f1-score   support

     negative      0.83        0.81      0.82       309
     positive      0.89        0.91      0.90       549

     accuracy                            0.87       858
    macro avg      0.86        0.86      0.86       858
 weighted avg      0.87        0.87      0.87       858
```

Fig. 6: Result of K-Neighbors Classifier Model.

```
                precision    recall  f1-score   support

     negative      0.94        0.93      0.93       309
     positive      0.96        0.97      0.96       549

     accuracy                            0.95       858
    macro avg      0.95        0.95      0.95       858
 weighted avg      0.95        0.95      0.95       858
```

Fig. 7: Result of SVM Model.

VI. REMARKS AND FUTURE RESEARCH

Rake achieved 94% success with only weekly news analysis. Rake, which can achieve higher performance with a multi-scope dataset, can be used in many different areas. For example, people, titles and areas can be selected that can analyze a country before the investments to be made by companies, and then provide positive benefits. Since the Rake project can already analyze positive keywords, negative words that can be included in a text can be made more suitable and used as an automatic text generation tool.

VII. CONCLUSIONS

People's thoughts and behaviors are at levels that can be easily manipulated. Acting on community psychology, individuals are good at imposing positive and negative thoughts on others easily. The Rake project determines a community's reactions to words and combines them with Machine Learning algorithms. Within the framework of the results obtained, words that can create negative thoughts can easily be replaced by positive ones. The Rake project specified how the above-mentioned methods should be used and the stages one by one and created a short-scale example. The dataset and programming outputs of the project are presented.

REFERENCES

[1] P. H. Lewis, "A boom for on-line services," *The New York Times*, pp. 1–8, July 1994.
[2] P. Christensson, ""listserv definition.," *TechTerms. Sharpened Productions*, July 2017.
[3] Weeks, Brian E., et al., "Incidental exposure, selective exposure, and political information sharing: Integrating online exposure patterns and expression on social media," *Journal of Computer-Mediated Communication*, pp. 363–379, 2007.
[4] Carty, Victoria, "Multi-issue, internet-mediated interest organizations and their implications for us politics: A case of moveon.org," *Social Movement Studies*, pp. 265–282, Jan 2011.
[5] Khondker, Habibul Haque, "Role of the new media in the arab spring," *Globalizations*, pp. 675–679, 2011.
[6] K. L. Netterstrøm, "After the arab spring: The islamists' compromise in tunisia.," *Journal of Democracy*, pp. 110–124, 2015.
[7] B. Korany and R. El-Mahdi, "Arab spring in egypt: revolution and beyond," *American University in Cairo Press*, 2012.
[8] A. Garwood-Gowers, "The responsibility to protect and the arab spring: Libya as the exception, syria as the norm," *UNSWLJ*, p. 132, 2013.
[9] Heydemann, Steven, "Tracking the arab spring : Syria and the future of authoritarianism," *Journal of Democracy*, pp. 59–73, 2013.
[10] Isaak, Jim, and Mina J. Hanna, "User data privacy: Facebook, cambridge analytica, and privacy protection," *Computer*, pp. 56–59, 2018.
[11] Larsen, Mark E., et al., "We feel: mapping emotion on twitter," *IEEE journal of biomedical and health informatics*, pp. 1246–1252, 2015.
[12] Tsou, Ming-Hsiang, et al., "Mapping social activities and concepts with social media (twitter) and websearch engines (yahoo and bing): a case study in 2012 us presidential election.," *Cartography and Geographic Information Science*, pp. 337–348, 2013.
[13] Chen, Shuai, et al., "R-map: A map metaphor for visualizing information reposting process in social media.," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1204–1214, 2019.
[14] Lu, Ying, ZeChen, and Rob Law., "Mapping the progress of social media research in hospitality and tourism management from 2004 to 2014.," *Journal of Travel Tourism Marketing*, pp. 102–118, 2018.
[15] Stieglitz, Stefan, et al., "The adoption of social media analytics for crisis management," *Challenges and Opportunities*, 2018.
[16] Hu, Xia, and Huan Liu. , "Text analytics in social media mining text data," *Springer,Boston, MA*, pp. 385–414, 2012.
[17] E. Umut, "Ankara haberleri - mynet'te yeniçerili strateji oyunu - yerel haberler," *Hurriyet News*, 2021.
[18] C. Zacharias, "twintproject/twint," *GitHub*, 2021.
[19] D. Klein and C. D. Manning, "Natural language grammar induction using a constituent-context model," *Advances in neural information processing systems*, vol. 1, pp. 35–42, 2002.
[20] J. Perkins, "Python text processing with nltk 2.0 cookbook," no. 1, 2010.
[21] J. J. Webster and C. Kit, "Tokenization as the initial phase in nlp," 1992.
[22] A. G. Jivani *et al.*, "A comparative study of stemming algorithms," *Int. J. Comp. Tech. Appl*, vol. 2, no. 6, pp. 1930–1938, 2011.
[23] S. Wilson, "Natural language processing for personal values and human activities," 2019.
[24] H. Saif, M. Fernandez, Y. He, and H. Alani, "On stopwords, filtering and data sparsity for sentiment analysis of twitter," 2014.