Determining of The User Attitudes on Mobile Security Programs with Machine Learning Methods

Rıdvan Yayla
Information Technology Department, Bursa Technical University,
Mimar Sinan Dist. Mimar Sinan Boulevard Eflak Str. 16310, Bursa, Turkey
E-mail: ridvan.yayla@btu.edu.tr

Turgay Tugay Bilgin
Computer Engineering Department, Bursa Technical University,
Mimar Sinan Dist. Mimar Sinan Boulevard Eflak Str. 16310, Bursa, Turkey
E-mail: turgay.bilgin@btu.edu.tr

Keywords: security, mobile application, classification, machine learning

Received: April 11, 2021

Security plays an important role in today's virtual world. Cybersecurity software has widely been used by the development of portable virtual environments. Smartphones take place in an important part of our lives. Daily routines are carried out over mobile phones, especially after the Covid-19 pandemic process. Due to its ease of use, compulsory or optional mobile phone use brought also about a lot of security concerns. Mobile security software is used for different purposes such as virus removal and protection of personal information according to user preferences. In the field of natural language processing, user preferences can now be analyzed on the basis of machine learning methods with sentiment analysis. In this paper, the preference reasons for mobile security software are analyzed with machine learning methods based on user comments and sentiment analysis. In the study, all user comments were classified into 10 main categories and the user preferences of mobile security programs were analyzed.

Povzetek:
"Ugotavljanje uporabniškega odnosa do programov mobilne varnosti z metodami strojnega učenja"

1 Introduction

Security played an important role in the virtual world for a lot of web applications over the past decades. People preferred web and mobile applications for their routines because of their ease of use. Mobile phone applications have been widely used by users, especially after the Covid-19 pandemic process. While mobile applications provide convenience for people, they brought also about some security problems. Due to security gaps, users have become the main target of hackers. While the number of mobile phones increased each day, security threats such as capturing personal information with social engineering methods also increase at the same rate.

Nowadays, there are a lot of security threats on smartphones such as advertisement, undesired information sharing with synchronization or fake link. Smartphone users try to prevent security threats with mobile security applications and firewalls. Mobile security programs carry vital importance to the users for providing protection. While some users can prefer an antivirus program, some prefer them for personal information privacy. Over the past decade, many smartphone models have been developed by phone manufacturers and users all over the world bought smartphones for their routines. The hackers have also targeted mobile applications in recent years due to the trending usage. This situation brought also about mobile phone security concerns. Due to the security gaps, the threats increased in vulnerable mobile phones. Against threats, many mobile security programs have been developed by developers and used by users for different reasons.

On the other hand, evaluation scales based on different structure such as questionnaire, comment text or rating star are also widely used among web and mobile users in recent years. A comment may be useful for others for buying anything or installation a program. Users give an idea and help for a purchase or installation process to each other. Google Play Store (GPS) which is an Android software platform gives an evaluation possibility to the users. While user attitudes on mobile security in the literature are handled with classical methods such as mostly survey, the GPS comments on mobile security programs are analyzed with a new approach based on machine learning methods in this paper.

2 Methodology

Most of the applications are downloaded on GPS by the usage of smartphones and the security programs become preferable among the users due to the security concerns. Developers have developed many mobile security applications for different purposes to meet these requirements. At the same time, a lot of research has been also made about the user attitudes on mobile security in recent years. Most of these researches has been made by either being handled a group of people or a local area with classical methods.

In 2008, a study that is made with a survey on 300 mobile-phone users in Oman investigated user attitude towards m-commerce and other mobile devices [1]. On the other hand, Tambe and Kulal developed an offline Android mobile security application when the smartphone is stolen or lost by a thief [2]. In 2016, another survey study realized with 301 attendees is made of awareness of mobile device security [3]. Özkan and Bıçakçı made an analysis for two-factor authentication against account hijacking attacks. They analyzed eleven different Android authenticator application and used different engineering techniques and open-source tools in their study [4]. Ophoff and Robinson made an online survey with 619 South African mobile users for exploring end-user smartphone security awareness and the questions based on mobile security are prepared by these researchers [5]. Ziqiang et al., proposed by combining static and dynamic security detection method to detect client-side. They divided mobile application security detection into two parts server and client security detection. Additionally, they developed an automated platform for mobile security detection [6]. Benenson et al., made an interview with 24 mobile users that are between 18 and 50 years old for satisfying the sentiment analysis requirement. He consulted security experiments and attitudes of these 24 mobile users for his study. As a result of interviews, he suggested some hypotheses [7].

As it is also seen in the literature, mobile security user attitudes have determined by classical survey methods. In our study, we proposed a new approach by using sentiment analysis based on machine learning and classification methods based on word analysis. In the study, the comments at GPS which is a global platform are investigated by confining with Turkish comments as a prototype study but the study can be also extended with different languages such as English, Italian, Hungarian or Slovenian provided that the categories and searched words should be defined in the desired language.

In the application, 249 mobile security programs on Google Play Store are examined and a sentiment analysis based on the combination of user comments and ratings have been made. The information taken with scraped method on GPS is processed and the user comments with ratings are analyzed as positive, negative and neutral with sentiment analysis.

Moreover, the criteria by which the users use the programs were analyzed by being categorized with 10 categories. The aim of the study, instead of the classical methods, a new approach based on the machine learning method that contains sentiment analysis by taking into account user ratings is to propose. The working principle of the proposed method is shown as Figure 1.

3 Materials and Methods

3.1 Support Vector Machine (SVM)

The sentiment analysis is based on natural language processing (NLP) and machine learning algorithms and it is made by Support Vector Machine (SVM). SVM is a supervised machine learning algorithm that can be used for both classification or regression challenges [8]. The comment sentiment is investigated as three main emotions that are positive, negative and neutral by SVM. Additionally, the Stochastic Gradient Descent (SGD) algorithm is used for the text classification in each comment of determined sentiment.

3.2 Stochastic Gradient Descent (SGD)

SGD is a fundamental machine learning approach that can be applied to large-scale and diluted machine learning problems frequently encountered in text classification and natural language processing [9]. SGD is also a good optimization algorithm by the ease of application and its productivity. The advantage of SGD is to update each sample in each step by decreasing optimization calculations, especially in big datasets. SGD refers to calculating the derivative from each training data instance and calculating the update immediately [10]. SGD algorithm is used based on Term Frequency – Inverse Document Frequency (TF-IDF) weight factor in text classification.

TD-IDF is a calculated weight factor of a word which shows importance in a text by using a statistical method .Thanks to TF-IDF, it is determined in which reason the GPS users write their ratings about mobile security programs. TF-IDF weights [11] are calculated as follow:

$$weight_{w,t} = \begin{cases} \log(tf_{w,T}+1)\log\frac{n}{x_w}, & f_{w,T} \geq 1 \\ 0, & otherwise \end{cases} \quad (1)$$

3.3 N-gram Model

N-gram basically means a sequence of N words. This model helps to predict the next item in a sequence. Although the n-gram model has more than one method, it basically consists of 3 main methods. Unigram express n-gram of size 1, bigram refers to n-gram of size 2 and trigram means n-gram of size 3 [12]. The size of n-gram can be also increased such as four-gram, five-gram. The n-gram model can be explained with a simple sentence as following:

Example sentence:
"*The mobile security application is very good.*"

1- Unigram: Each single word is considered for the recurrent calculation.
   "The", "mobile", "security", "application", "is", "very", "good"
2- Bigram: Each pair of words are considered for the recurrent calculation.

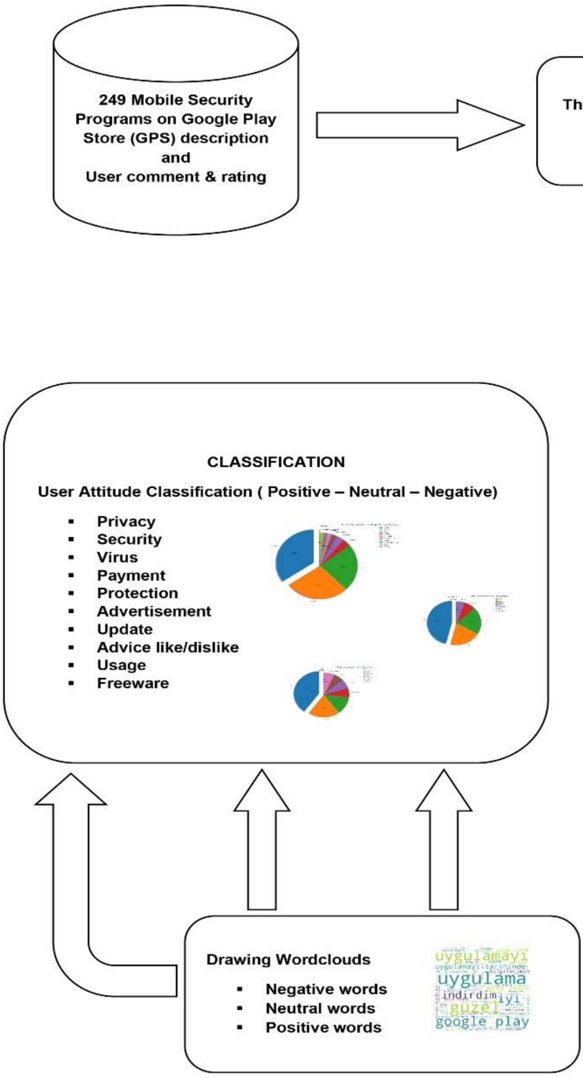

Figure 1: Working principle of the proposed method.

"The mobile", "mobile security", "security application", "application is", "is very", "very good".

3- Trigram: Each three sequence of words are considered for the recurrent calculation.
"The mobile security", "mobile security application", "security application is", "application is very", "is very good"

In the study, the n-gram analysis is applied as unigram, bigram and trigram for all comments. Additionally, the keyword processing algorithm is separately used at positive, negative and neutral comments. Keyword processing assigns a rate to word and it gives a score. Finally, it gives a percentage score of the text as a whole. The words are drawn by starting from the highest score word by word cloud [13]. This process is executed for all positive, negative and neutral text separately and the most used words are shown in each sentiment group as a word diagram.

3.4 Model Accuracy

The classification model that follows the working principle was evaluated with the standard metrics called accuracy, precision, recall and F1-score where TP is true positive, TN is true negative, FP is false positive and FN is false negative [14].

Accuracy is a statistical measure which is defined as the division of the correct predictions (TP & TN) made by a classifier divided by the sum of all predictions made by the classifier, including FP and FN [15]. The accuracy is computed as follow:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

Precision is defined as the ratio pf the correctly identified positive cases in all the predicted positive cases [16]. Precision is computed as follow:

$$precision = \frac{TP}{TP+FP} \tag{3}$$

Recall is the sensitivity of the model and it is defined as the ratio of the correctly identified positive cases to all the actual positive cases, which is the sum of FN and TP [17]. Recall is also shown as follow:

$$recall = \frac{TP}{TP+FN} \tag{4}$$

F1 Score is the harmonic meaning of the precision and recalled by taking into FP and FN cases [18]. It shows a good performance in an unbalanced data set. It is calculated as follow:

$$F_1 score = \frac{2 \times (precision \times recall)}{(precision \pm recall)} \tag{5}$$

4 System Design and Components

4.1 Dataset

In our study, the mobile security software on Google Play Store are examined with Turkish user comments and votes. The user comments and votes are evaluated for the sentiment analysis together. Initially, 249 mobile security programs and related comments are handled as metadata. Google play scraper based on PHP composer is used for extracting metadata and all program names, application ids and user comments and votes are written in an excel file [19]. Thanks to Google Play scraper, 45617 comments that belong to 249 mobile security programs have been examined with a written python script in the application.

4.2 Sentiment Analysis

Sentiment analysis has been realized based on three main emotions as negative, positive and neutral. Moreover, the user votes as rating stars are also counted in sentiment analysis examination. In literature, the sentiment analysis is made in English and there are a few sentiment analyses studies with other languages such as Spanish or Turkish [20][21]. The sentiment analysis studies based on other languages don't reflect sentiment analysis as successfully as English [22]. The diagram used in this study for the sentiment analysis is shown in Figure 2.

Instead of the usage of non-English sentiment analysis, the comments that are written without English are translated to English and the sentiments are supported with the user votes related to mobile security program for decreasing translation errors. Thanks to the translation process, the application can be executed with the other languages provided that the category names and searched words used for classification are written in the desired language.

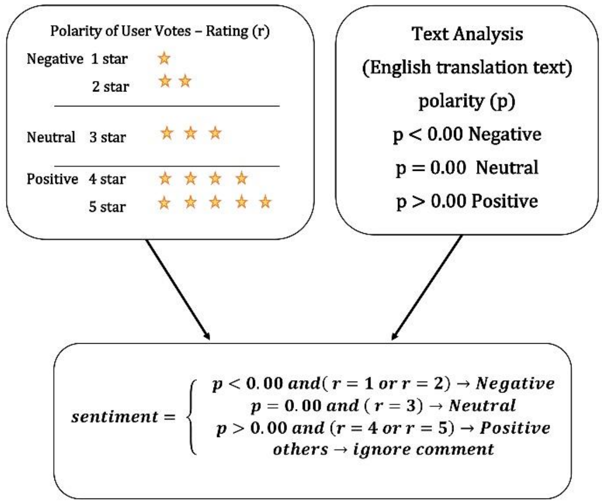

Figure 2: Sentiment analysis of the proposed method.

The sentiment analysis of each comment is basically determined with sentiment polarity. Additionally, the user votes are also taken into account for a realistic analysis in this study. As a hypothesis, the user ratings are accepted 1 or 2 stars as negative, 3 stars as neutral and 4 or 5 stars as positive [23]. Furthermore, when the comment text is evaluated with sentiment polarity, a polarity less than 0 is considered negative, a polarity equal to 0 is considered neutral, and a polarity greater than 0 is considered positive [24]. If any user evaluation doesn't follow both rules, that comment is ignored for the sentiment analysis and

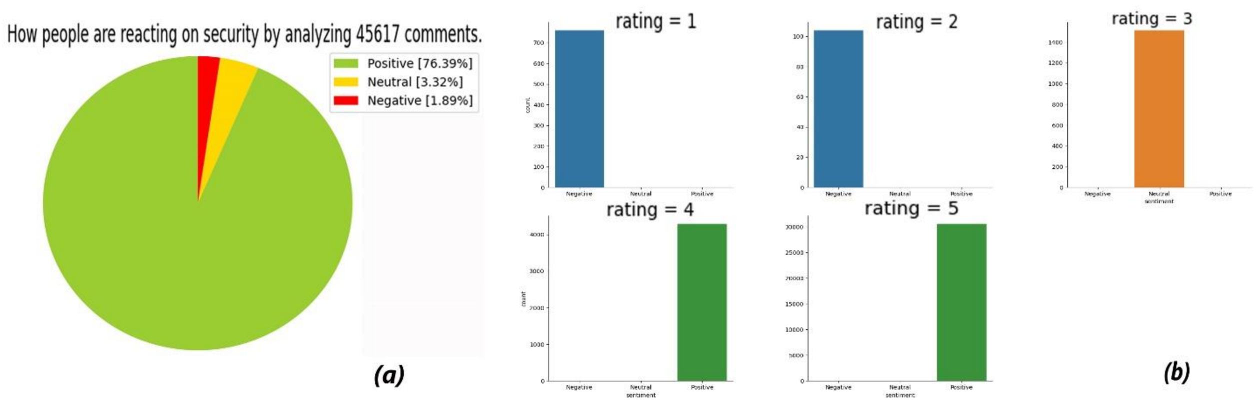

Figure 3: (a) User reactions (b) User votes

Class No	*Class Name* **(Turkish Class Name)**	***Searched Words* (Translation)**
1	Virus (Virüs)	*Temiz* (clean), *virüs, virus* (virus), *yok* (not), *bahis* (bet), *reklam* (advertisement), *sil* (delete)
2	Security (Güvenlik)	*Temiz* (clean), *güven* (secure), *kişisel* (personal), *bilgi* (info), *siber* (cyber), *saldırı* (attack), *koruma* (protection), *virüslü* (infected)
3	Privacy (Gizlilik)	*Kişisel* (personal), *bilgi* (info), *şahıs* (person), *paylaş, paylas* (share)
4	Freeware (Ücretsiz)	*Ücretsiz*, *ucretsiz* (freeware), *bedava* (free) ,*parasız* (no charge), *free*
5	Paid (Ücretli)	*Para* (money), *ücret* (fee), *kazan* (win), *reklam* (advertisement), *yıllık* (annual), *uyelik, üyelik* (membership), *abone* (subscriber), *deneme* (trial), *satış* (sales), *satma* (sell), *kart* (card), *sürüm* (version), *premium, ödeme* (payment), *satin* (buy), *fiyat* (price)
6	Advertisement (Reklam)	*Reklam* (advertisement), *afis*, *afiş* (banner), *pano* (board), *sürekli* (permanent), *izle* (watch), *gereksiz* (unnecessary)
7	Protection (Koruma)	*Virus* (virus), *sil* (delete), *tehdit* (threat), *koru* (secure), *etkili* (effective), *karşı* (against)
8	Update (Güncel)	*Performans* (performance), *güncel* (update), *otomatik* (automatic), *son* (final), *indirme* (download)
9	Usage (Kullanım)	*Performans* (performance), *kullanışlı* (useful), *yararlı* (benefit), *zararlı* (harmful), *işlevsel* (functional), *kasmıyor* (twitch), *ısınma* (warming),*şarj* (charge), *gereksiz* (unnecessary), *tüketim* (consumption), *bildirim* (notification)
10	Advice like/dislike (Tavsiye beğeni)	*öneri* (offer), *beğen* (like), *tavsiye* (advice), *çok iyi* (very well), *maşallah*, *harika* (wonderful), *kötü* (bad), *hiç* (any)

Table 1: The categories of user attitude.

classification. Moreover, some users can make a negative comment and give a positive vote. These kinds of comments have been ignored by means of the proposed method.

Due to the user comments and star ratings, the user reactions are also shown in Figure 3(a). According to the evaluation, the GPS users thought about security programs as positive 76.39 %, neutral 3.32% and negative 1.89%. In addition, the number of positive, negative and neutral ratings according to the user ratings is shown in Figure 3 (b).

In each comment, the most used words (unigrams), bigrams and trigrams are determined by n-gram analysis. In addition to n-gram analysis, word clouds that contain positive, negative and neutral comment words have been also separately drawn. Finally, all comments are classified with 10 user attitude categories according to the sentiment and it has been determined whether the users like or not the security programs according to which criteria. The algorithm accuracy is evaluated with SVM algorithm.

4.3 Classification

When the mobile security comments are observed as a whole, 10 main Turkish classes are determined for the user attitude classification. The classification is made based on these classes. The class names (categories) based on user attitude are shown in Table 1.

The class names weren't only searched, but also other words denoting the class name in each comment. A searched word can belong to more than two classes. These searched words were taken into account for each related class. For example, let a Turkish comment is as follow:

Turkish Comment:
"*Virüsü iyi temizliyor, kişisel bilgilerimi kurtardım, koruma sağlıyor, kullanışlı bir program. tavsiye ederim*."

English Translation:
"*It cleans the virus well; I recovered my personal information and it provides protection. It is a useful program. I advise everybody*"

In this study, all words are searched as Turkish.

1- "*Virus*" and "*clean*" words will belong to virus class,
2- "*personal*" and "*information*" words will belong to privacy and security classes
3- "*protection*" word will belong to protection class,
4- "*useful*" word will belong to usage class
5- "*advise*" word will belong to advice like/dislike class.

In addition to searched words, the root of the word such as "protect" word in protection class or "pay" word in paid class is also evaluated within the classes. Thereby, the frequency of the word is determined in the comment text. Even if this prototype study is prepared in the Turkish language, the word search process can be made with different language such as Spanish or Slovenian languages. If the searched words and classes are prepared in Slovenian, the search can also be made in Slovenian.

5 Experimental Results

5.1 User Attitudes

When the experiment on user ratings and comments is observed, it is stated that most of the users thought positive about mobile security programs. Moreover, users have installed the programs for different reasons such as privacy, security or freeware. When the comments have been separated by the sentiment analysis, the most used words have been showed by word cloud at each sentiment group. The most used positive (a), negative (b) and neutral (c) Turkish words are shown in Figure 4.

positive, neutral and negative comments are shown in Figure 5(a), Figure 5(b) and Figure 5(c) .
The association probability of the words is also examined by n-gram analysis as unigram, bigram and trigram. The stop words such as “Google”, “Play”, “Store”, “very”, “for” are ignored at n-gram analysis. For all comments, the unigram, bigram and trigram analysis are also shown in Figure 6, Figure 7 and Figure 8 respectively.

5.2 Model Evaluation

In the study, the sentiment analysis based on SVM is evaluated with standard metrics which are precision, recall, F1 score. In our study, the comment text that has

(a) Positive (b) Neutral (c) Negative

Figure 4: (a) Positive (b) Neutral (c) Negative words.

When all comment texts are also investigated based on 10 categories that are shown in Table 1, the user attitudes are also shown in a pie chart for each sentiment group. In the positive comments, it is observed that the users mostly preferred mobile security programs for privacy, security and cleaning virus. In the negative comments, it is observed that the users mostly complained about mobile security programs for privacy, security and cleaning virus. In the neutral comments, it is observed that the users mostly stayed neutral on mobile security programs about privacy and virus. The pie charts of the

the positive, neutral and negative polarity are examined for the comment texts and the accuracy of the classifier is calculated as 97%. The accuracy of the model is shown in Table 2.

Accuracy Evaluation	Precision	Recall	F1-Score	Support
Negative	0,96	0,63	0,76	83
Neutral	0,90	0,44	0,56	143
Positive	0,97	1,00	0,98	3497
Model accuracy			0,97	3723
Macro avg	0,94	0,69	0,78	3723
Weighted avg	0,97	0,97	0,96	3723

Table 2: The accuracy of the model.

According to Table 2, while the weighted average represents weights of the performance metrics negative, neutral and positive, the macro average represents an average of the performance metrics without weights. The support is the number of actual user attitude classes in the comment texts that follow the proposed rule.

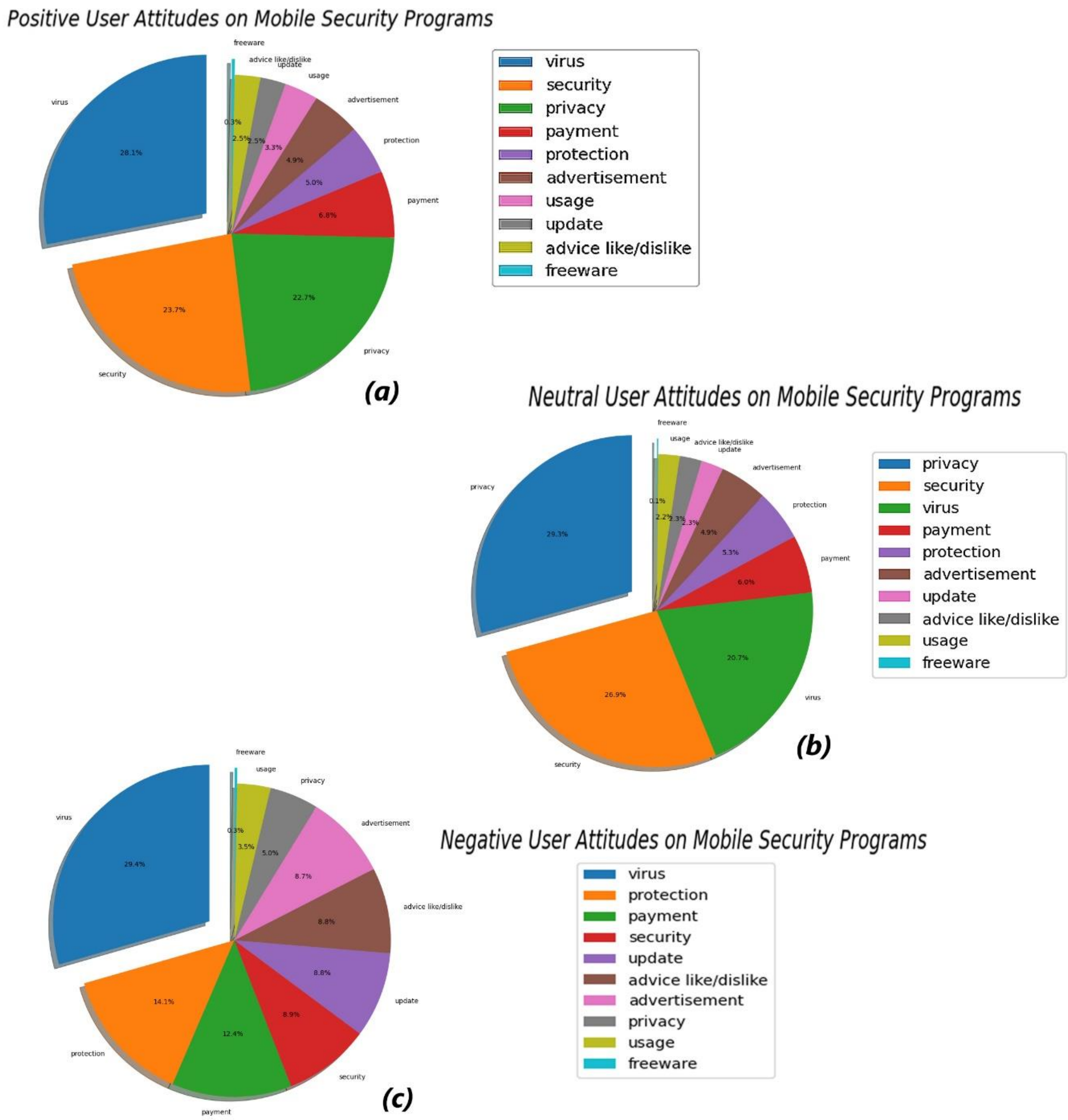

Figure 5: The positive (a), neutral(b) and negative (c) classified user attitudes.

6 Conclusion

Nowadays, security plays an important role in every virtual platform. With the development of mobile phones, security concerns have become more important in mobile devices. The advantages such as usage of ease, fast accessibility of the information, and portability caused to use phones by the people at more.

After the Covid-19 pandemic, phone usage noticeably increased all over the world. As a result of this situation, the requirement for mobile security is also increasing at the same rate. Mobile phone users share their ratings about phone features and mobile application at application store markets and other platforms. The shared information has become of vital importance for the other users. By observing the user attitudes, both the developers can develop different security mobile software and the users can use the security application for the most desired request. Moreover, the user attitudes on the virtual world are not only important for security but also becomes an indispensable element for marketing, politics and society.

In our study, the usage of mobile security programs has been examined with machine learning methods and the user attitudes on mobile security are also investigated by predetermined criteria. A user analysis is made by considering the Turkish user attitudes on mobile security software as a prototype study. The study can be also expanded with user attitudes that are in different languages and topics.

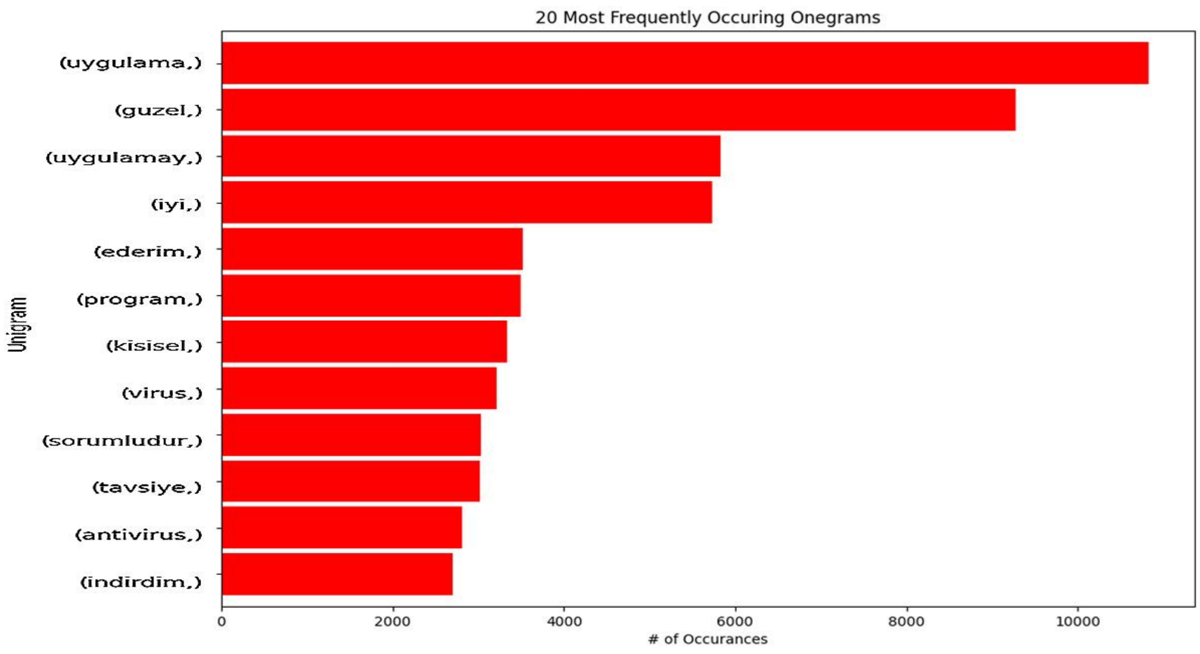

Figure 6: Unigram n-gram analysis.

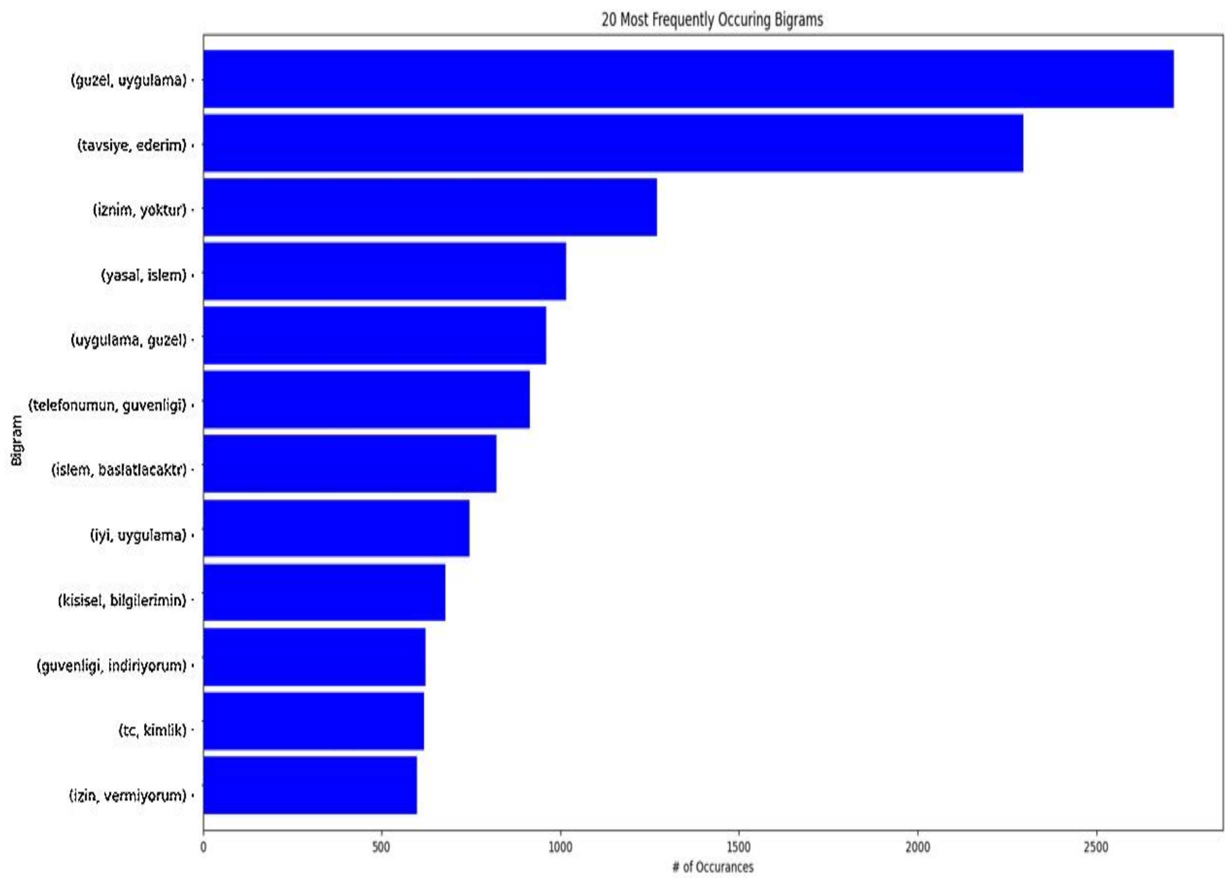

Figure 7: Bigram n-gram analysis.

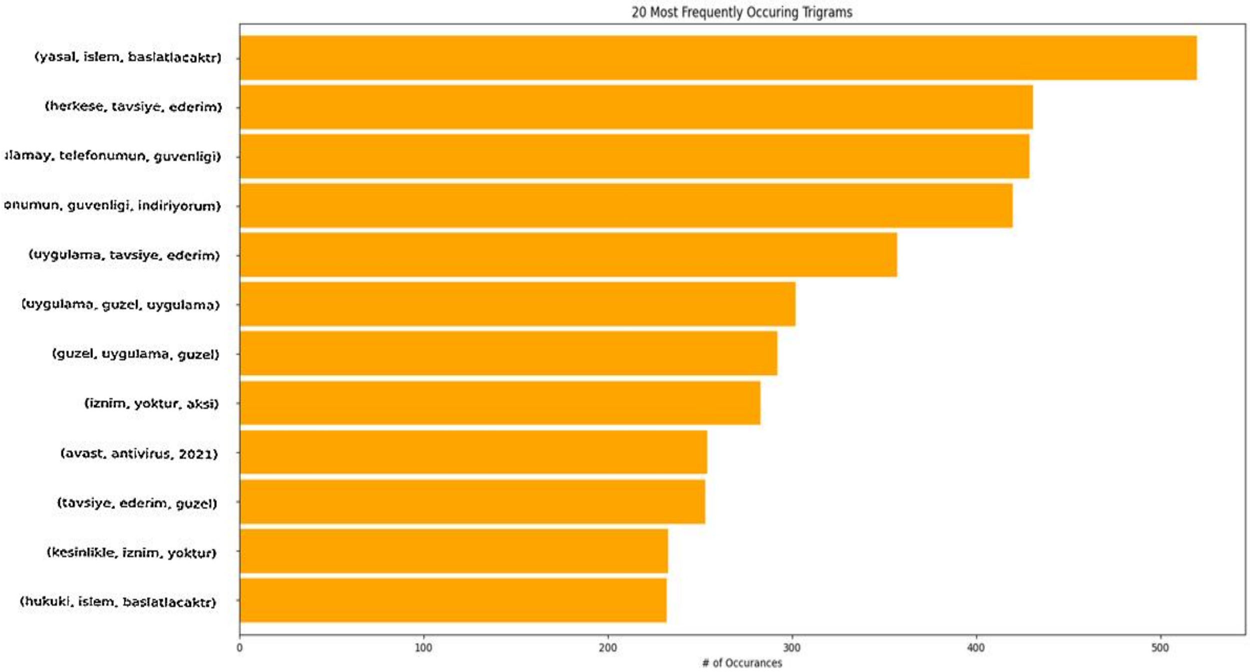

Figure 8: Trigram n-gram analysis.

Additionally, the user attitudes can be evaluated with higher accuracy by the development of the machine learning techniques and a lot of prediction can be made by using the previous user attitudes. In the near future, mobile security will become an indispensable element of our lives with the minimization of communication tools and this situation will more required investigation of the user attitudes.

References

[1] N. Manochehri and M. Y. Alhinai (2006). "Mobile phone users attitude towards Mobile Commerce (m-commerce) and Mobile Services in Oman". *2nd IEEE/IFIP International Conference in Central Asia on Internet*, Tashkent, Uzbekistan, pp. 1-6, https://doi.org/10.1109/CANET.2006.279277.

[2] V. Tambe, D. Chauhan, S. Kulal and S. Sherkhane (2018). "*Offline Mobile Security*". 2018 International Conference on Smart City and Emerging Technology (ICSCET), IEEE, Mumbai, India, pp.1-4, https://doi.org/10.1109/ICSCET.2018.8537303.

[3] N. Clarke, J. Symes, H. Saevanee and S. Furnell, S. (2016). “Awareness of Mobile Device Security: A Survey of User's Attitudes”. *International Journal of Mobile Computing and Multimedia Communications (IJMCMC)*, IGI Global Publisher of Timely Knowledge7(1), pp.15-31. https://doi.org/10.4018/IJMCMC.2016010102

[4] C. Ozkan and K. Bicakci (2020). "Security Analysis of Mobile Authenticator Applications," *2020 International Conference on Information Security and Cryptology (ISCTURKEY)*, IEEE, *Ankara, Turkey*,pp.18-30, https://doi.org/10.1109/ISCTURKEY51113.2020.9308020.

[5] J. Ophoff and M. Robinson (2014). "Exploring end-user smartphone security awareness within a South African context," *2014 Information Security for South Africa*, IEEE, Johannesburg, South Africa, pp. 1-7, https://doi.org/10.1109/ISSA.2014.6950500.

[6] Z. Zhou, C. Sun, J. Lu and F. Lv (2018). "Research and Implementation of Mobile Application Security Detection Combining Static and Dynamic", *2018 10th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, IEEE, Changsha, China, pp. 243-247, https://doi.org/10.1109/ICMTMA.2018.00065.

[7] Z. Benenson, O. Kroll-Peters and M. Krupp (2012). "Attitudes to IT security when using a smartphone," *2012 Federated Conference on Computer Science and Information Systems (FedCSIS)*, IEEE,Wroclaw, Poland, pp. 1179-1183.

[8] S. Ray (2017)., "Understanding Support Vector Machine (SVM) algorithm from examples", Analytic Vidhya, Retrieved from https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/

[9] N. S. Huda, M. S. Mubarok and Adiwijaya (2019)."A Multi-label Classification on Topics of Quranic Verses (English Translation) Using Backpropagation Neural Network with StochasticGradient Descent and Adam Optimizer," *2019 7th International Conference on Information and Communication Technology (ICoICT)*, IEEE, Kuala Lumpur, Malaysia, pp. 1-5, https://doi.org/10.1109/ICoICT.2019.8835362.

[10] M. S. Alsadi, R. Ghnemat and A. Awajan (2019)."Accelerating Stochastic Gradient Descent using Adaptive Mini-Batch Size," *2019 2nd International Conference on new Trends in*

Computing Sciences (ICTCS), IEEE, Amman, Jordan, pp. 1-7, https://doi.org/10.1109/ICTCS.2019.8923046.

[11] F. Kabir, S. Siddique, M. R. A. Kotwal and M. N. Huda (2015). "Bangla text document categorization using Stochastic Gradient Descent (SGD) classifier", *2015 International Conference on Cognitive Computing and Information Processing (CCIP)*, IEEE, Noida, India, pp. 1-4, https://doi.org/10.1109/CCIP.2015.7100687.

[12] A. Tripathy, A. Agrawal and S. Kumar Rath (2016). "Classification of sentiment reviews using n-gram machine learning approach", *Expert Systems with Applications*, Pages 117-126, https://doi.org/10.1016/j.eswa.2016.03.028.

[13] I. N. Dewi, R. Nurcahyo and Farizal (2020). "Word Cloud Result of Mobile Payment User Review in Indonesia," *2020 IEEE 7th International Conference on Industrial Engineering and Applications (ICIEA)*, Bangkok, Thailand, pp. 989-992, https://doi.org/10.1109/ICIEA49774.2020.9102048.

[14] M. Hasnain, M. F. Pasha, I. Ghani, M. Imran, M. Y. Alzahrani and R. Budiarto (2020). "Evaluating Trust Prediction and Confusion Matrix Measures for Web Services Ranking," *IEEE Access*, vol. 8, pp. 90847-90861, https://doi.org/10.1109/ACCESS.2020.2994222.

[15] C. Liu, Y. Sheng, Z. Wei and Y. Yang (2018). "Research of Text Classification Based on Improved TF-IDF Algorithm," *2018 IEEE International Conference of Intelligent Robotic and Control Engineering (IRCE), IEEE*, Lanzhou, China, pp. 218-222, https://doi.org/10.1109/IRCE.2018.8492945.

[16] V. Sundaram, S. Ahmed, S. A. Muqtadeer and R. Ravinder Reddy (2021). "Emotion Analysis in Text using TF-IDF," *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) IEEE*, Noida, India, pp. 292-297, https://doi.org/10.1109/Confluence51648.2021.9377159.

[17] T. Hasan, A. Matin and M. S. R. Joy (2020). "Machine Learning Based Automatic Classification of Customer Sentiment", *2020 23rd International Conference on Computer and Information Technology (ICCIT) IEEE*, Dhaka, Bangladesh, pp.1-6, https://doi.org/10.1109/ICCIT51783.2020.9392652.

[18] M. Kumar Jain, D. Gopalani, Y. Kumar Meena and R. Kumar (2020). "Machine Learning based Fake News Detection using linguistic features and word vector features," *2020 IEEE 7th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON) IEEE*, Prayagraj, India, pp. 1-6, https://doi.org/10.1109/UPCON50219.2020.9376576.

[19] Ne-Lexa (2020, March 8). *google-play-scraper*, https://github.com/Ne-Lexa/google-play-scraper

[20] S. M. Jimenez Zafra, M. T. Martin Valdivia, E. Martinez Camara and L. A. Urena Lopez (2019). "Studying the Scope of Negation for Spanish Sentiment Analysis on Twitter," *IEEE Transactions on Affective Computing*, 10(1), pp. 129-141, https://doi.org/10.1109/TAFFC.2017.2693968.

[21] M. Rumelli, D. Akkuş, Ö. Kart and Z. Isik (2019). "Sentiment Analysis in Turkish Text with Machine Learning Algorithms," *2019 Innovations in Intelligent Systems and Applications Conference (ASYU), IEEE*, Izmir, Turkey, pp. 1-5, https://doi.org/10.1109/ASYU48272.2019.8946436.

[22] T. T. Zin, "Sentiment Polarity in Translation" (2020). 2020 IEEE Conference on Computer Applications (ICCA) IEEE, Yangon, Myanmar, pp. 1-6, https://doi.org/10.1109/ICCA49400.2020.9022831.

[23] Fang, X., Zhan, J. (2015). "Sentiment analysis using product review data". *Journal of Big Data* vol,2, 5 pp. 1-14. https://doi.org/10.1186/s40537-015-0015-2

[24] F. Calefato, F. Lanubile, F.Maiorano, et al. (2018). "Sentiment Polarity Detection for Software Development". *Empir Software Eng*, Springer Link, 23, pp. 1352–1382 . https://doi.org/10.1007/s10664-017-9546-9