*Avrupa Bilim ve Teknoloji Dergisi*
*Özel Sayı, 34, S. 541-545, Mart 2022*
*© Telif hakkı EJOSAT'a aittir*
***Araştırma Makalesi***

www.ejosat.com ISSN:2148-2683

*European Journal of Science and Technology*
*Special Issue 34, pp. 541-545, March 2022*
*Copyright © 2022 EJOSAT*
***Research Article***

# Analyzing of Cyber-Security Concepts on Twitter

Nazmiye Eligüzel*[1], Lana Manla Ali[1]

[1] Gaziantep University, Industrial Engineering, 27310 Gaziantep, Turkey, (ORCID: 0000-0001-6354-8215), nazmiye@gantep.edu.tr
[1] Gaziantep University, Industrial Engineering, 27310 Gaziantep, Turkey (ORCID: 0000-0002-8456-2638), manlalana6@gmail.com

(2st International Conference on Applied Engineering and Natural Sciences ICAENS 2022, March 10-13, 2022)

(**DOI:** 10.31590/ejosat.1083377)

**ATIF/REFERENCE:** Eligüzel, N. & Ali, L. M. (2022). Analyzing of Cyber-Security Concepts on Twitter. *European Journal of Science and Technology*, (34), 541-545.

### Abstract

Today's buzzwords are cyber-security and social media. With the advancement of technology and the increased usage of the internet, these concepts are becoming more prominent and research on this subject is expanding. This paper addresses the issue of cyber-security and which professions in different business sectors discuss the topic of cyber-security on Twitter. Therefore, cyber-security concepts and contributors on Twitter are analyzed separately. In that way, business sectors that contributes to cyber-security are tried to be deduced with respect to different concepts. In addition, focuses of business sectors on cyber-security concepts are analyzed. The proposed paper demonstrates the application of the K-means clustering method to cyber-security concepts on Twitter and experts who posted tweets about these concepts.

**Keywords:** Clustering, Cyber-security, Social media, Twitter.

# Twitter'de Siber Güvenlik Kavramlarının Analizi

### Öz

Günümüzün popüler kelimeleri siber güvenlik ve sosyal medyadır. Teknolojinin ilerlemesi ve internet kullanımının artmasıyla birlikte bu kavramlar daha çok ön plana çıkmakta ve bu konudaki araştırmalar genişlemektedir. Bu makale, siber güvenlik konusunu ve Twitter'de siber güvenlik konusunu ele alan farklı iş sektörlerindeki meslekleri ele almaktadır. Bu nedenle, Twitter'deki siber güvenlik kavramları ve ve bu kavramlara katkıda bulunan meslekler ayrı ayrı analiz edilmiştir. Böylece siber güvenliğe katkı sağlayan iş sektörleri farklı kavramlara göre çıkarılmaya çalışılmıştır. Ayrıca, siber güvenlik kavramlarına odaklanan iş sektörleri de analiz edilmiştir. Önerilen makale, Twitter'deki siber güvenlik kavramlarına K-araç kümeleme yönteminin uygulanmasını ve bu kavramlar hakkında tweet'ler gönderen iş uzmanlarını göstermektedir.

**Anahtar Kelimeler:** Kümeleme, Siber güvenlik, Sosyal medya, Twitter

*Corresponding Author: nazmiye@gantep.edu.tr

# 1. Introduction

Cyber-security is defined as an activity or process, ability or capability, or state whereby information and communications systems and the information contained therein are protected from and/or defended against damage, unauthorized use or modification, or exploitation[†]. Cyber-security techniques are often outlined in published materials that aim to protect a user's or organization's cyber environment. Protecting information and systems against serious cyber threats is part of the scope of cyber-security operations. There are several cyber-security approaches available to combat cyber-security threats. Some of the common threats are cyber spionage , cyber warfare, and cyber terrorism (Seemma  et al., 2018). Some of the popular techniques to combat the cyber-attacks are authentication, encryption, digital signatures, antivirus, firewall, and steganography (Pande, 2017). Cyber-security concept is important issue for organizations. There are various ways in which organizations are tackling cyber-security. The topics cover these ways are risk management, cyber insurance, technical controls, training and awareness raising, staffing and outsourcing, and governance approaches and policies (Johns, 2021). Cyber-security is an issue that concerns experts in different business sectors and organizations. According to the report (Johns, 2021) the following business sectors have a higher focus on cyber-security:

- finance and insurance (72% say it is a *very* high priority, vs. 37% of all businesses)
- information and communications (62%)
- health, social work and social care (56%).

In recent years, these three sectors have consistently given cyber-security a greater priority. By contrast, and similarly to previous year, the food and hospitality sector and the construction sector also see cyber-security as a lower business priority (only 62%) (Johns, 2021).

The proposed study aims to analyze cyber-security concepts from the business perspectives on Twitter. In the analysis, cyber-security concepts, its contributors, who are the experts in different business branches and net sentiments scores are considered. In the proposed study, Twitter data is used because Twitter is an important platform where people share their feelings and thoughts. On Twitter, a variety of topics have been discussed and clustered together. Some of these studies are as follows:

Sadasivuni and Zhang (2020) proposed a study that applied Learning Quotient and Text mining methods in order to cluster gathered tweets with specific hashtags such as bombing, depressed, and anti-depressed during the 'Bomb' blasts in April 2019. Amati et al. (2021) introduced a massive text document clustering algorithm by using community detection methods on the weighted hashtag graph instead of any traditional clustering algorithm, such as LDA. Following the assignment of hashtags to clusters, the most popular clusters and hashtags were associated with topics of broad interest, such as sports, politics, and health. Yoshida et al.(2021) looked at a network of retweets regarding former Japanese Prime Minister Shinzo Abe to see how conservative and liberal disparities in the extent to which political tweets reach less partisan moderate people in a nonwestern culture. The usage of hashtags did not differ between conservatives and liberals in the examination of tweet content, but there were variations in the use of emotion terms and linguistic expressions. During the 2018 Brazilian presidential election, Soares and Recuero (2021) looked at how political misinformation disseminated during discursive conflicts on Twitter. These were disputes for the dominant narrative between two storylines based on opposing hashtags: one based on mainstream media content and the other on misinformation, especially from hyper partisan sources. Hassan et al. (2021) proposed a study to see how the #datasaveslives hashtag was used on social media, how often it was used, and by whom; this way, they could have a better idea of the impact of a large social media campaign in the UK health informatics research community and beyond.

As seen above mentioned studies, social media, especially Twitter is the significant platform to analyze various topics under the pre-determined hashtags. In the proposed study, the hashtags related to the cyber-security field are evaluated with together contributors on Twitter.

When we consider academic studies in the field of cyber-security, it can be said that there are too much studies. In recent years, cyber-security has been a popular issue in the academic. By using "cyber-security" search string, publications are evaluated in "Web of Science" database (February 08, 2022). Search is resulted in 11042 publications. The number of publications according to the study areas is given in Fig 1.

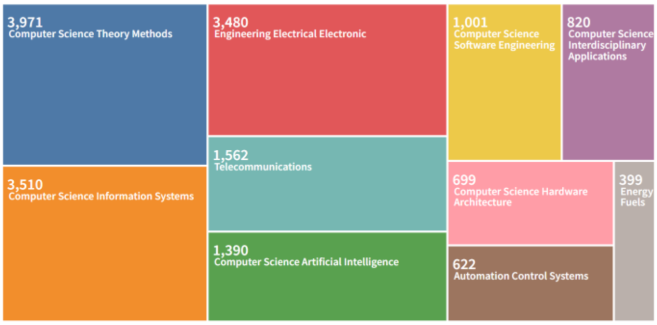
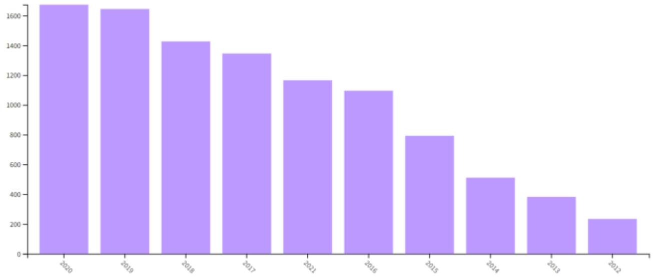


Fig 1. Cyber-security study areas

As seen from Fig. 1, cyber-security concept is handled with mostly in the areas of "computer science theory methods" and "computer science information systems". After that, the fields of "engineering electrical electronic", "telecommunications", "computer science artificial intelligence", and so on are seen, respectively. When we consider the years of studies, it is seen in Fig 2,  the number of studies increases every year except 2021.



[†] NICCS, 'Explore Terms: A Glossary of Common Cybersecurity Terminology,' https://niccs.us-cert.gov/glossary

Fig 2. The years of cyber-security related publications

Fig. 2 demonstrates that studies in the cyber-security field are on an increasing trend between the years of 2012 and 2020. However, in 2021, a decreasing trend is seen in this field.

The topic of cyber-security is prominent in academia. In addition, cyber-security is an issue that is of interest to experts in different business sectors. The proposed study aims to analyze cyber-security topic on Twitter from the expert perspectives.

The rest of this paper is organized as follows: Section 2 demonstrates the methodology; Section 3 presents results and discussions, and finally the paper is concluded in section 4.

## 2. Methodology

By using two different data-sets, the steps in the implementation of the K-means algorithm are demonstrated. One of them is the contributor type of social media users who post tweets about cyber-security issues, and the other data is about cyber-security concepts. Data is retrieved from GlobalDtata[‡] between the dates of November 1, 2020 and October 31, 2021. In Appendix 1, cyber-security concepts, the number of total posts of the mentioned concepts, the number of contributors who posted tweet about that concept, and their net sentiments are provided. Net sentiment scores are given at GlobalData platform. It is calculated according to the positive, neutral, and negative mentions. Appendix 2 demonstrates the contributor types of social media users about cyber-security issues with the number of total posts, contributors, and net sentiments.

In the proposed study, K-means algorithm is utilized. The K-means clustering approach is part of a group of techniques known as partitioning-based techniques, which are based on the repeated repositioning of data points between clusters. It is used to divide a data set's events or variables into non-overlapping clusters (Morissette & Chartier, 2013). Namely, it uses similarity matrix. The general process of K-means algorithm is given as follows:

**Algorithm 1** General process of the K-means algorithm

1. ***Select*** the value of K (number of clusters using Elbow method)
2. ***Choose*** K random points to act as centroids.
3. ***Assign*** each data point to the nearest centroid that will build the predetermined clusters depending on their distance from the randomly selected centroid
4. ***Place*** a each cluster's new centroid
5. ***Repeat*** step 3 to reassign each data point to the cluster's new nearest centroid.
6. If there is a reassignment, proceed to Step 4; otherwise, proceed to Step 7.
7. ***Finish***

As seen from the general process of the K-means algorithm, in order to determine the number of clusters, Elbow method is utilized. Elbow method is the one of the popular cluster optimization methods. This approach compares the difference in the sum of square error (SSE) of each cluster to assess the consistency of the optimal number of clusters. The best cluster number is formed by the most extreme difference establishing the elbow angle [10]. The major aim is to reduce the distance between the data points and the cluster's centroid. The operation is repeated until the sum of distances reaches a minimal value. These processes are implemented by utilizing Python 3.9 software.

## 3. Results and Discussions

First of all, cyber-security concepts are analyzed. In Fig 3. the optimal number of clusters is demonstrated through Elbow method.

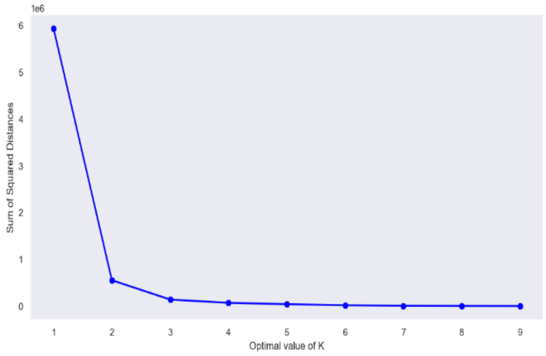


Fig. 3. Elbow method for cyber-security concepts

As seen from the Fig.3 , the elbow point is seen at the optimal value of K equals to 2. Therefore, the number of clusters is chosen as 2. After that, K-means algorithm is applied by considering the number of clusters. All the features, that are concepts, total posts, contributors, and net sentiment, are considered by making clustering. According to the given indexes in Appendix 1, index of points in the clusters are demonstrated in Fig. 4

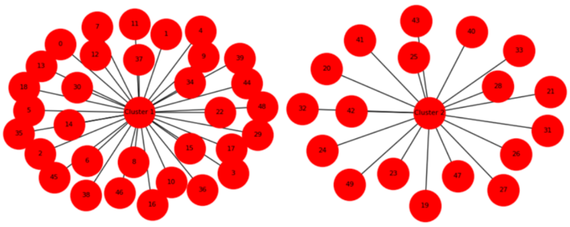


Fig. 4. Indexes of the clusters

In Fig. 4 the relationship between cyber-security concepts and total posts are demonstrated. As seen from the Fig. 5, there are two centroids as follows:

Centroid 1: [[ 27.9375 55.9054375 13.2595 0.9290625 ]

Centroid 2: [ 18.38888889 16.81166667 697. 0.96444444]]

[‡] https://explorer.globaldata.com/ThemesView/DashboardAnalytics/1768?keyword=8627

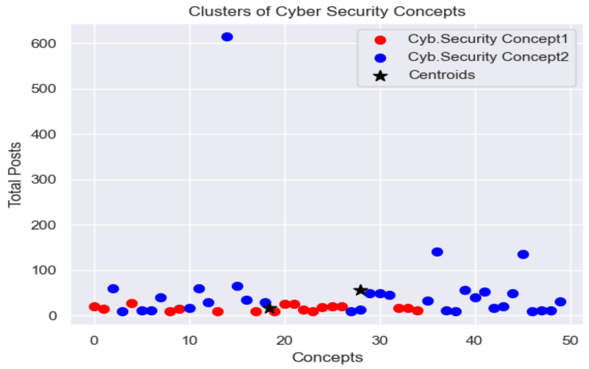


Fig. 5. Representation of the cyber-security concepts

As seen from the Fig. 5, cluster 1 (Cyber-security Concept 1) includes lesser concepts and lesser total posts. However, in cluster 2 (Cyber-security Concept 2), there are much more concepts than cluster 1 and total post about some concepts are quite high.

The concepts which are seen in cluster 1 are “CyberSecurity”, “Information Security”, “Securities”, “Cyberattack”, “Cyber Attack”, “Artificial Intelligence”, “IoT”, “Malware”, “Ransomware”, “Hacker”, “Hackers”, “Hacking”, “Cloud”, “Machine Learning”, “Cybercrime”, “Information Privacy”, “Technologies”, “Data Breach”, “Cyber Crime”, “Phishing”, “Coronavirus Disease 2019”, “Networks”, “Government”, “Breach”, “COVID-19”, “Innovation”, “Startups”, “Startup Company”, “Spyware”, “Automation”, “Firms”, and “Investments”. It is seen that cluster 1 is mostly related to the security issues and investments. İt considers information security and privacy, cyber-security, threat-related problems, new investments. In addition, cluster 1 includes pandemic issue which effect the life at recent times.

Cluster 2 covers the concepts of “Big Data”, “Data Protection”, “Data Science”, “5G”, “Fintech”, “Financial Technology”, “Digital Transformation”, “IIoT”, “Industrial Internet Of Things”, “Analytics”, “Cloud Security”, “Deep Learning”, “Infographic”, “Cloud Computing”, “Data Privacy”, “Denial Of Service Attack”, “DDoS”, “Cyber Threat”. It can be concluded that cluster 2 includes data-related, financial, and analytical issues.

In the second part of the study, contributor types, that is, professions who posted about cyber-security subject are analyzed. K-means clustering technique is used by considering all features in Appendix 2, such as contributor type, total posts, the number of the contributors, and net sentiment scores. Firstly, the number of clusters is determined by using Elbow method as in the first part of analysis. In Fig. 6 optimal number of clusters are given.

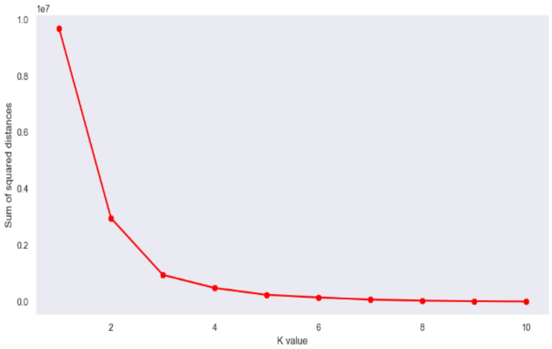


Fig. 6. Elbow method for contributor types

According to the Fig. 6 the number of the clusters are taken as 3. After that, K-means algorithm is applied. According to the given indexes in Appendix 2, index of points in the clusters are demonstrated in Fig. 7.

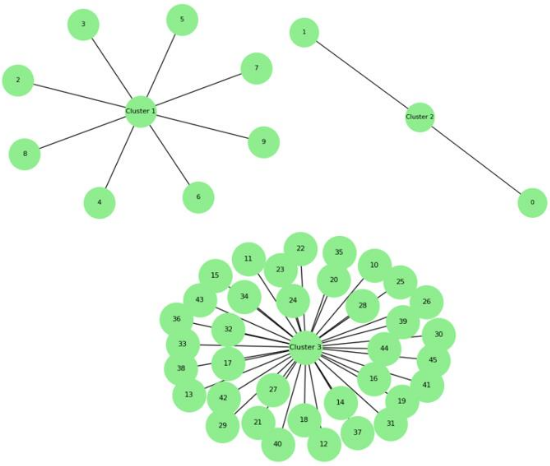


Fig. 7. Indexes of the clusters

In Fig. 7 the relationship between contributor types and total posts are demonstrated. As seen from the Fig. 8, there are three centroids as follows:

Centroid 1: [[1.98750000e+01 9.55895000e+04 3.07875000e+02 9.41250000e-01]

Centroid 2: [2.75000000e+01 3.56295500e+05 1.56300000e+03 9.30000000e-01]

Centroid 3: [2.28055556e+01 5.42333333e+03 1.88361111e+02 9.18888889e-01]]

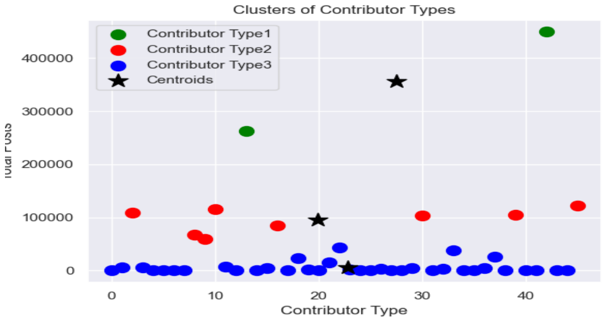


Fig. 8. Representation of the contributor types

As seen from Fig. 8, there three types of clusters. Cluster 1 includes only two elements but much more total posts. Cluster 2, includes 8 elements and total posts for each element looks average. Cluster 3 includes much more elements but total post for each element is low when compared to the other two clusters.

Cluster 1 includes the concepts of “Technology Expert” and “Cybersecurity Expert”. It is said that technology experts, in addition to cyber-security experts, have sent a significant number of posts about cyber-security topic.

Cluster 2 includes the concepts of “WearableTech Expert”, “CXO/Founder”, “AI Expert”, “Robotics Expert”, “IoT Expert”, “EV Expert”, “Big Data Expert”, and “Blockchain Expert”. Namely, Experts operating in the subject of disruptive technologies can be found in cluster 2.

The concepts which are seen in cluster 3 are “Fintech Expert”, “Media and publishing”, “Researchers and Consultants”, “Editor/Journalist”, “Finance Expert”, “Cloud Expert”, “Academician”, “ADS Expert”, “Pharma & Medical Expert”, “EGS Expert”, “Investors/PE VC”, “Marketing Expert”, “Healthtech Expert”, “Energy & Mining Expert”, “Food & Beverage Expert”, “Doctors”, “Automotive Expert”, “Retail & Ecommerce Expert”, “Activists”, “Equity Analyst/Trade Analyst”, “Healthcare Expert”, “Sustainability Expert”, “Legal Expert”, “5G Expert”, “Travel & Tourism Expert”, “Construction Expert”, “Economist”, “Packing Expert”, “Beauty & Fashion Expert”, “Health & wellness Expert”, “Insurtech Expert”, “Online Payments Expert”, “BIM Expert”, “Tobacco Expert”, “Sports Expert”, and “Insurance Expert. In cluster 3, it can be noticed that there is a diverse variety of profession groups. This demonstrates that numerous experts from different fields have sent out tweets about cyber-security.

## 4. Conclusions and Recommendations

This paper takes a look at which experts in various business sectors post tweets about the cyber-security field and what the related concepts are. The proposed study comprehends the utilization of K-means clustering approach to analyze cyber-security concepts on Twitter and the experts who tweeted about them. Firstly, cyber-security concepts are analyzed, and then contributors, namely experts in business sectors, on cyber-security concepts are evaluated from the perspective of cyber-security field. In this method, many business sectors that contribute to cyber-security are attempted to be deduced using various concepts. Furthermore, the focus of business sectors on cyber-security concepts is examined.

In the future study, cyber-security issues on Twitter can be analyzed by considering all contributors from various sectors.

## 5. Acknowledge

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## References

Amati, G., Angelini, S., Cruciani, A., Fusco, G., Gaudino, G., Pasquini, D., & Vocca, P. (2021). Topic Modeling by Community Detection Algorithms. OASIS 2021 - Proceedings of the 2021 Workshop on Open Challenges in Online Social Networks, 15–20. https://doi.org/10.1145/3472720.3483622

Hassan, L., Nenadic, G., & Tully, M. P. (2021). A social media campaign (#datasaveslives) to promote the benefits of using health data for research purposes: Mixed methods analysis. Journal of Medical Internet Research, 23(2). https://doi.org/10.2196/16348

Johns, E. (2021). Cyber Security Breaches Survey 2021: Statistical Release. Department for Digital, Culture, Media and Sport. https://assets.publishing.service.gov.uk

Morissette, L., & Chartier, S. (2013). The k-means clustering technique: General considerations and implementation in Mathematica. Tutorials in Quantitative Methods for Psychology, 9(1), 15–24. https://doi.org/10.20982/tqmp.09.1.p015

Seemma, P. S., S. Nandhini, and M. Sowmiya. (2018). Overview of Cyber Security. Ijarcce, 7(11), 125–128. https://doi.org/10.17148/ijarcce.2018.71127

Pande, J. (2017). Introduction to Cyber Security ( FCS ). http://uou.ac.in

Sadasivuni, S. T., & Zhang, Y. (2020). Clustering Depressed and Anti-Depressed keywords Based on a Twitter Event of Srilanka Bomb Blasts using text mining methods. Proceedings - 2020 IEEE International Conference on Humanized Computing and Communication with Artificial Intelligence, HCCAI 2020, December 2014, 51–54. https://doi.org/10.1109/HCCAI49649.2020.00014

Soares, F. B., & Recuero, R. (2021). Hashtag Wars: Political Disinformation and Discursive Struggles on Twitter Conversations During the 2018 Brazilian Presidential Campaign. Social Media and Society, 7(2). https://doi.org/10.1177/20563051211009073

Yoshida, M., Sakaki, T., Kobayashi, T., & Toriumi, F. (2021). Japanese conservative messages propagate to moderate users better than their liberal counterparts on Twitter. Scientific Reports, 11(1), 1–9. https://doi.org/10.1038/s41598-021-98349-2