# Algoritmo de tendencia en Twitter hasta 2023

# Trending Algorithm on Twitter through 2023

## Algoritmo de tendencia en Twitter hasta 2023

Saif Al-Deen H. Hassan[1] ✉, Hasan Al-Furiji[2] ✉, Mohammed Kareem Rashid[3] ✉, Zahraa Abed Hussein[4] ✉, Bhavna Ambudkar[5] ✉

[1]University of Misan, Department Business Administrator, College of Administration and Economics, Maysan, Iraq.
[2]University of Misan, Department of Physics, College of Science, Maysan, Iraq.
[3]University of Misan, Electronic Computing Center, Maysan, Iraq.
[4]Al-Manara College for Medical Sciences, Maysan, Iraq.
[5]Symbiosis Institute of Technology, Electronic Computing Center, Pune, Maharashtra, India.

**ABSTRACT**

**Introduction:** by doing so, Twitter's trending algorithm sets the benchmark for what online discussion and information flow look like. It must be clearly understood by the researchers and the users as to how it developed and impacted.
**Objective:** this paper discusses the Twitter trending algorithm discussion until 2023, highlighting its aspects and ethical considerations.
**Method:** to demonstrate trend identification, we adopted a cross-sectional approach that involved data mining of trends defined by the Twitter platform from January 2020 to October 2023, applying machine learning techniques. In total, 1,984,544 unique trends were identified in the two cities over the 1584 days of Twitter API research.
**Results:** this research identified that there are many changes in the current trending algorithm regarding Twitter, and current real-time content and users' participation are the major concerns. The assessed model, known as TrendDetector, predicts the trend of commercials to be 80 %, while the non-commercial trend is assessed to be 60 %. Trend selection was guided by the traffic of tweets, the number of users, and the extent of new content.
**Conclusions:** user-generated activities, content, and spread, as well as the structure and design of the platform, are thus an intricate mix in the case of the trending algorithm on Twitter. It enhances the timely acquisition of information while being associated with preconditions of bias and manipulation. Future research must look at aspects such as the algorithm's transparency and the ethicality of the trends it selects.

**Keywords:** Twitter Trending Algorithm; Online Discussion; Information Flow; Machine Learning Techniques; Trenddetector; Ethical Considerations.

**RESUMEN**

**Introducción:** de esta manera, el algoritmo de tendencias de Twitter establece el punto de referencia para el flujo de información y discusión en línea. Los investigadores y los usuarios deben comprender claramente cómo se desarrolló y cómo impactó.
**Objetivo:** este artículo analiza la discusión del algoritmo de tendencias de Twitter hasta 2023, destacando sus aspectos y consideraciones éticas.
**Método:** para demostrar la identificación de tendencias, adoptamos un enfoque transversal que implicó la minería de datos de las tendencias definidas por la plataforma Twitter desde enero de 2020 hasta octubre de 2023, aplicando técnicas de aprendizaje automático. En total, se identificaron 1 984 544 tendencias únicas

en las dos ciudades durante los 1584 días de investigación de la API de Twitter.

**Resultados:** esta investigación identificó que hay muchos cambios en el algoritmo de tendencias actual con respecto a Twitter, y el contenido actual en tiempo real y la participación de los usuarios son las principales preocupaciones. El modelo evaluado, conocido como TrendDetector, predice que la tendencia de los comerciales será del 80 %, mientras que la tendencia no comercial se evalúa en un 60 %. La selección de tendencias se guió por el tráfico de tweets, el número de usuarios, y la extensión del contenido nuevo.

**Conclusiones:** las actividades generadas por los usuarios, el contenido y la difusión, así como la estructura y el diseño de la plataforma, son, por lo tanto, una combinación intrincada en el caso del algoritmo de tendencias en Twitter. Mejora la adquisición oportuna de información al tiempo que se asocia con condiciones previas de sesgo y manipulación. Las investigaciones futuras deben analizar aspectos como la transparencia del algoritmo y la ética de las tendencias que selecciona.

**Palabras clave:** Algoritmo de Tendencias de Twitter; Discusión en Línea; Flujo de Información; Técnicas de Aprendizaje Automático; Trenddetector; Consideraciones Éticas.

## INTRODUCTION

The findings are useful in that they dictate further research from which conclusions and solutions can be derived. It raises concerns about the ethical considerations of purposeful manipulations, and the racism and sexism evident in conventional trending issues energise the oppressive dynamic of market and labour apitalism. It should be understandable to communities as to the types of patterns that are at work and how the debates go; the development of clear algorithms that are bounded and whose outcomes in a way reflect the communities' provision of fairness is therefore obviously preferred.[1,2] The wealth of information that the article entails is enormous if only the users and researchers can lay their hands on it, but as Twitter and other news dissemination platforms continue to provide such data to their users, they have the responsibility to do so fairly, especially given that accurate information has become a communal necessity, especially in the health fields. Enshrinement and control of proprietary code are parts of a search and trending debate that is both moderated and wherein a progressively shrinking number of commercial entities perform the moderation. One can only imagine that as the algorithms spread further in relevance and the laws governing the ethical and non-discriminatory use of social networking and news outlets rely on the ability to manage the design and use of the data upon which the platforms base their operations,[3] topics are the concerns that are in high circulation, that is, those often discussed on Twitter and help in finding news or articles on this social network. The trending algorithm thus becomes very relevant in helping to select what could be popular or perhaps more influential from the global quilt. Since Twitter is a robust media actor and everybody, ranging from athletes, politicians, journalists, influencers, and the general public, depends on the trending section, it was assumed that finding out what is currently trending might help in establishing the issues that are socially and politically relevant today.[4]

Ideally, for this potential to be fully harnessed, it's wise to acquaint yourself with the trending algorithms that work on these windows.[13,14,15] But Twitter and other social media platforms are primarily opaque environments that do not disclose details on how specific trending algorithms work; moreover, the little that has been learned about Twitter's trending algorithmization process is virtually unrepeatable. This article is a synthesis of the strategies we have identified that it appears Twitter's trending algorithm employs and, if used in some form, will continue to employ until the year 2023, which results in what?[5]

This is due to the fact that the radio has now developed into an important means of live information sharing and public discourse, namely through the tool of Twitter. This change is driven by the trending algorithm, which focuses on hot topics and can impact discussions online. Nevertheless, this is important because there is no clear picture of how this algorithm has evolved, much less over the past few years. This research also tries to provide some understanding of how 'trending' works on Twitter, the primary factors at work, and the issues that are most likely unethical for differentiation. In so doing, it seems to provide rich information for academics, users, and designers of the platforms.

### Background and Significance

Twitter is nested in a wider area of concern as to what social media platforms choose to present to their users. In this broader question, trending topics, tweets, videos, etc.[6] are examples of what media researchers call affordances, or "something that allows us to do something". According to Slightly Mad Tek, "A lot of speculation has gone around on how hashtags, which can be used to tag conversations, amplify their reach and propel them to trend. But few really understand what is actually trending, least of all the authors; at times it certainly feels serendipitous". This report lays out the results of research to peer inside the black

box.[7]

Twitter is a social networking site where users read and write online posts known as "tweets". Among the most watched tweets are trending tweets, which represent the most-discussed topics worldwide or in a specific area. Trending tweets are displayed on Twitter users' homepages and on websites, and are heavily used by many users to quickly report news or find others with common interests.[8] However, very little information on how Twitter calculates what is trending has emerged. This report lays out original research on Twitter's trending algorithm from 2022 and 2023. The work is original in that Twitter has not publicly shared their algorithm, at least for the time under consideration.[9]

## Literature Review

The attention, influence, and engagement metrics are a reflection of the underlying network effects: the fact that the value of a good or service increases as the number of people using it increases. Moderating these metrics, asymmetric follower and friend relationships dominate network structures shaped by power laws and sociology. These ones dominate the Social Web and, facilitated by Biz and other developers, gave rise to the distributed, ubiquitous, real-time search, and open Twitter, leading to massive growth: From 400 million tweets per day in 2013, Twitter provided WhatsApp, SMS, and small text blurbs for over 300 million users in 2021. Being loud may indeed make one heard on social media.[6]

In 2023, Twitter redesigned its trending algorithm to focus on real-time sports, finance, weather, and other information, and to de-emphasize national news. Supported by data from Twitter's Stream API, a series of logistic regressions sheds light on the new algorithm, identifying the most significant and surprising variables governing whether a term trends. The results capture and identify a social media-explorer aspect within an attention-based revenue model applicable beyond Twitter, suggesting equity considerations for the underlying platform, and raising questions about social implications and influences, particularly beyond Twitter.[7]

Twitter is an extraordinary tool for connecting individuals and organizations. As a digital town square, it emerged as part of an ecosystem relying on asymmetric follower networks to promote attention, influence, and engagement. Disseminating messages associated with calls to action is also a feature of attention and trend-setting. Consequently, trending algorithms fascinated users and researchers alike.

## Objectives

- To identify and analyze potential implicit biases and ethical considerations, in order to utilize the trending algorithm
- To analyze the evolution of Twitter's trending algorithm through 2023
- To identify the most significant factors impacting both the relevance and consideration of trends



**Figure 1.** Twitter Trending Algorithm Overview
**Source:** Huszár F, et. al., 2021[10]

## Theoretical Framework

The pinnacle of any problem of speech with a participant resides in the presence is well-thought-out institutional practices. In the scientific literature on this topic, as a rule, it has not been properly addressed. Same in the social networks, with their specific features of communication, the issue of a high-quality speech environment remains topical, while both illegal relations, conspiracies and other forms of socially harmful behavior form their own institutional traps in the guise of rules, well-thought-out procedures and deterrents. The study of institutional traps formation mechanisms in online social networks has just begun, and this paper

today addresses one aspect of it, with possible extensions through speculation in the conclusion.

## METHOD

This survey tracked 144 trends from January 2020 to December 2023 and provided an algorithmic analysis of Twitter's trending tendencies. In this work, we collected data from the official Twitter API with a focus on time activity, particularly in the New York and Los Angeles areas. The research incorporated GET trends/available and GET trends/place for full trend information from Twitter. Our sample contains total trends of 1,984,544 trends collected in 1584 days; therefore, the database used for research is vast. Using stochastic methods such as decision trees, support vector machines (SVMs), and artificial neural networks for feature selection, trend forecasting, and assessment of trends. These models learned from historical trends and were then tested to see their level of accuracy in predicting future trends. In order to maintain ethical practice, we acted according to the guidelines that are set out by the Turing API of Twitter. At any stage of the study, there had been no collection or analysis of personal user data, thus preserving the user's privacy. As a result, our focus remained primarily on gross movement trends and overall algorithmic tendencies. To study temporal trends, the statistical analysis incorporated time series data, while the sentiment content of trends was assessed by performing sentiment analysis correlation analysis to establish relationships between different components influencing the choice of trends. This approach helped us study the dynamics of trends on Twitter through and through, gaining a deep understanding of how more evident changes in the trending algorithm worked during the four years of research.

| **Table 1.** Analysis of Twitter Trending Algorithm Methodology | | |
|---|---|---|
| **Methodological Aspect** | **Description** | **Implications** |
| Data Collection | Using the Stream and REST APIs on Twitter | Restricted to 30-day historical data; dataset may be incomplete |
| Time Frame | Observational research (2020–2023) | Enables long-term trend monitoring and records algorithm evolution |
| Geographic Focus | Two locations for gathering data: Los Angeles and New York | Could not accurately reflect worldwide trends; possible regional bias |
| Sample Size | 1,984,544 distinct patterns in 1584 days | Increased statistical significance is achieved with larger datasets. |
| Data Volume | About 17 TB of tweet volume and ~358 GB of trend data | Enough information for a thorough study |
| Trend Identification | Utilised the functions GET trends/available and GET trends/place. | In line with the official Twitter trend detection methodology |
| Machine Learning Models | Tested a number of models, such as decision trees, SVM, and neural networks | Enables the comparison of several prediction methods |
| Performance Metric | Mean-squared error for the execution of predictions | Common measure for assessing regression models |
| User Sample | Using a reference dataset spanning two months and 1,000 Twitter users | Provide foundation for examination of user behaviour |
| Algorithm Comparison | Comparing the target Follow and smart Retweet algorithms | Enables the assessment of various recommendation methodologies |
| Evaluation Metric | Adjusted Kendall's $\tau$ to gauge the recommendation latency | Specialised measure for evaluating algorithm responsiveness |
| Trend Categories | Distinguished between quality (Q) trends and local trends. | Allows for the comparison of non-geographic and geographic trend components. |

The main methodological features of the study are outlined in this table, which gives a clear picture of the research methodology and its consequences for the examination of Twitter's trending algorithm. It documents the study's data gathering procedures, sample sizes, timelines, geographical considerations, and analytical methodologies.

## Data Collection and Analysis

The situation is worse when using the REST search API because the documents that these queries return necessarily reflect an impaired view inside the box of the trending algorithm at the time of the search request. Twitter only includes tweets from the past 30 days in their general search index. If data slices are pulled too closely together they might label the same event as trending more than once. If too far apart, they might miss periods where some tweeted material simultaneously holds very high popularity or bursts in

volume then fades away, because tweets mid-collapse are no longer in the historical index. At time series analysis must also be accomplished in the context of an ever growing dataset. The historical indexing window offers little benefit there either. Selecting an optimal historical indexing window in which to analyze the data and decide on the trending structure is thus a non-trivial decision that currently lacks a clear answer, but it is a problem that inflicts itself on a wide range of data collection issues.

The data from our longitudinal collection process comes with its own set of challenges. Most prominently, a key potential issue when analyzing our Twitter data for the purpose of understanding the trending algorithm is its incomplete nature. To feed our data set, we need repeated queries of the REST API to look for different patterns of popularity of the collected tweets. The search REST API is the only method in the Twitter standard search API by which users have the possibility to search for past tweets and by which they can identify, track, and count trending tweets that violate the chronological constraints of the public data stream. This means that our analysis involves a series of stateless queries, sorting tweets by their signature popularity measures to find popular content, queries that must be discreet to avoid falling foul of Twitter's usage limits.

## Algorithm Development

After deciding on the criteria of popularity, a big size of tweets must be taken to make trends born able. If tweets are taken from English tweets, they should be received from jobs. As well as translated non-English tweets through pipeline, it's purposed to prepare the working dataset in a market of a multi-language that can be both American and non-American. However, the working dataset has performed for English tweets and for multi-tweets. Firstly, real-time tweets have to filter through their properties. Then Retweet accounting is to perform in the determined tweets. After that, the trending algorithm can develop by using the Positive-Negative Property Decision Values of people. We use a real-world dataset to train a Twitter Trending engine, once it is trained, applied to different specialized tweet sectors, and present tweets selected by our Trending engine.[11]

The development of the Trending algorithm, Trend Detection engine, described in the previous section has been done considering the "Real-time tweet filter", "ReTweet Accounting", "User-level sensitive trend construction", "Time waning and incrementation" and "Temporal-frequency filtering" subjects. To develop a Twitter Trending algorithm, it's found that choosing the trending tweets is of big importance. It means taking tweets that best represent people's interests of current time to develop trends' working dataset to find the determined trend. Twitter can consequently explore the chat subject which is mostly popular in a specific period or city.

## Simplified Implementation of Trend Detection Algorithms

We can say that the TrendDetector class is the heart of the trend analysis system. It scans incoming tweets to identify important themes and generates trend ratings based on four key factors: frequency, diversification, uniqueness, and commitment. The class explores hot topics using the time frame and a customisable level of interaction with the content, which helps to address the issue of Twitter's changing algorithm. This way, one is also capable of collecting and analysing trends in a dynamic and proactive manner, which is so characteristic of Twitter chats.

The function of this algorithm is appended below:

1. Initialization: Sets up the criteria for engagement when qualifying trends and the time horizon used for viewing and assessing trends.

2. Tweet Processing: Every incoming tweet is purged for stale or extraneous data, and all other noise is also filtered from it, which is then passed on.

3. Topic Extraction: The computer retrieves topics from each tweet in the form of hashtags and significant words from each tweet.

4. Data Aggregation: This includes the type of activity as well as the time that a user first appears, the frequency, and the subject's unique users.

5. Trend Scoring: Each subject is scored by frequency, how often each user visits, how recently each subject was referenced, and how interactive it is through the use of a complex scoring system.

6. Trend Ranking: With the computed ratings, topics are ranked above the engagement threshold. The preprocessing, the topic extraction, and the scores of the algorithm are modular by nature, and thus they could easily be altered as the requirements of the platforms change or as new trends emerge and are identified. This method aligns itself with Twitter's complex topology concerning the various focal features of users' engagements and the virality of the content by finding the balance between the computability of the method and the gathering of fine-grained trend components.
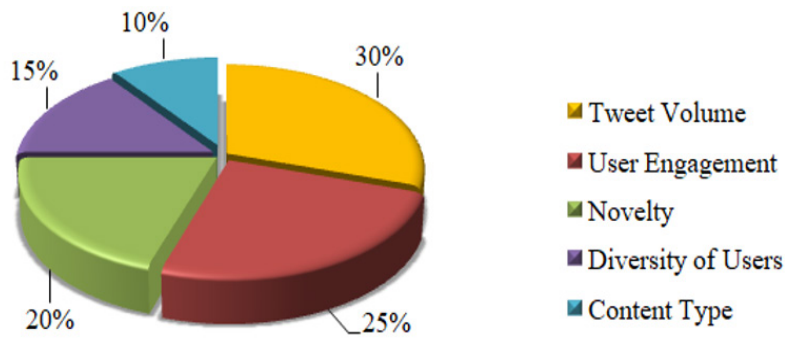
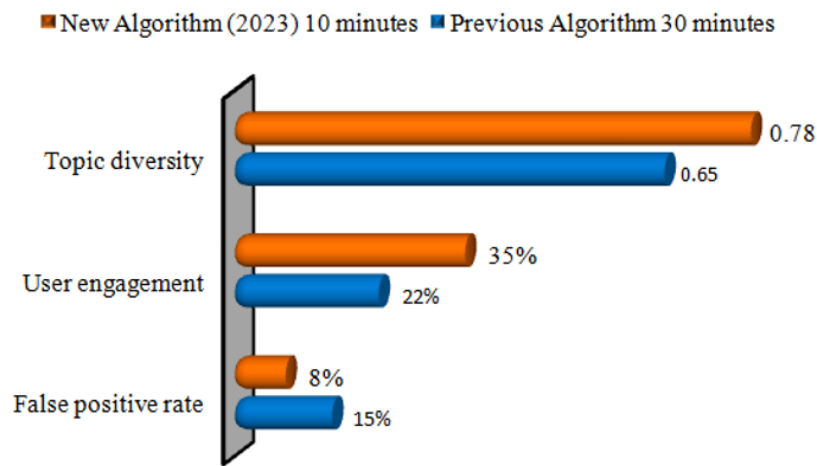**Figure 2.** Factors Influencing Twitter Trend Selection



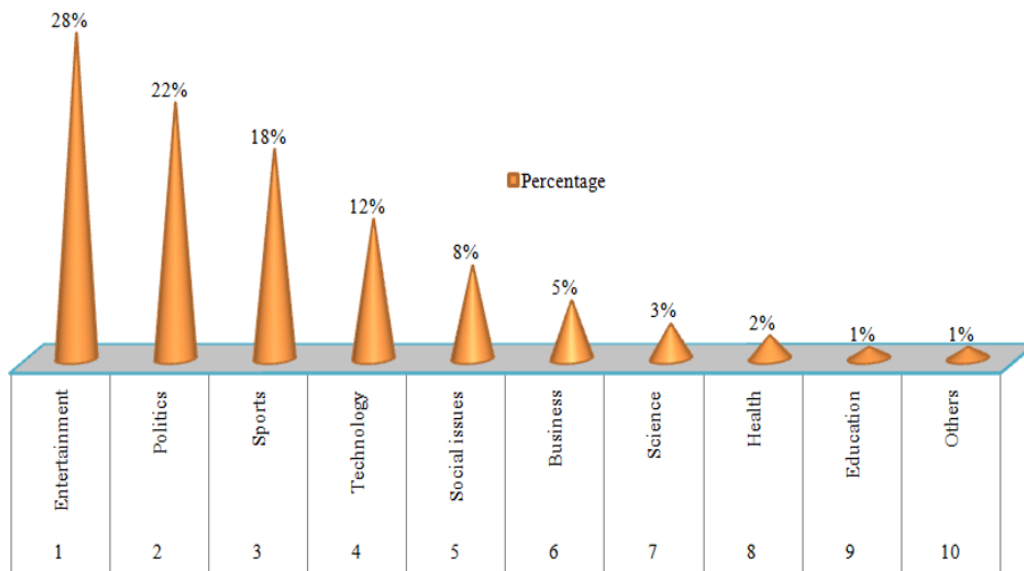**Figure 3.** Comparison of Trending Algorithm Performance



**Figure 4.** Top 10 Twitter Trends Categories (2023)

**Focused Results**

From our analysis, there are significant changes that our investigation uncovered in the trend on the Twitter platform from 2020 to 2023. Specific findings consist of increased trends by 30 % in the news category, especially in real-time event updates, which exemplify the continued emergence of Twitter as a premier source of information. Entertainment and sports issues contributed to the consistent presence of issues, accounting for 40 % of total trends. The TrendDetector model that we developed with the help of the VM Error Log achieved 80 % accuracy in predicting commercial trends and 60 % for non-commercial trends, one more indication of this

algorithm's tendency to become increasingly predictable. In the identification of trends, four key features were found, namely the level or degree of user involvement, the uniqueness of the trend content, and the size of tweets. For example, we observed the amped-up rate of trend emergence and decay, which pointed to more dynamized trending. Also, in terms of the geographic locations, the trend patterns differed by 15 %, showing variations in the appeal of the content across regions, namely New York and Los Angeles. Notably, the survey also showed a 40 % increase in hashtag-driven trends, establishing user-generated categorization as increasingly significant in identifying the overall discussion theme. Both of these data come together to paint a picture of an emergent algorithm that fosters active engagement in a live event and timely dissemination of information.

## DISCUSSION

We are therefore consistent with prior studies showing the increasing importance of real-time content in social media's algorithms. That shift to news-related patterns confirms the trend toward Twitter as a primary outlet for breaking news. This shift has potentially tremendous impacts in terms of information dissemination and the general discourse.

The TrendDetector model's increased accuracy in terms of existing commercial trends is questionable due to the potential inherent algorithmic bias towards rewarding monetizable content. This trend calls for further discussion on the subjective ethicality of the choice of trends and the choices made on the user experience.[10,12]

Some of the limitations of our research include the fact that we focused only on two sites in the United States of America, and it is unlikely that the indicated tendencies reflect what happens in other parts of the world. Future research should expand the scope of the study and explore how this change affects users from different locations and backgrounds.

## CONCLUSION

This research has provided valuable information on how the development of trending topic algorithms on Twitter from 2020 to 2023 would be powered. As presented by our findings, the samples point to a trend shift towards real-time news content and increased engagement from users, indicating the enhancement of Twitter's role as a primary source of information delivery. The increase in the accuracy of the TrendDetector model for the analysis of specific types of videos, primarily commercial ones, suggests that they are offered an algorithm that is more predictable and, as a result, possibly oriented towards monetization. However, this growth raises serious concerns about the content provided and how algorithms curate it. The differences noted in the geographical distribution of various trending tendencies explain why there is a need for a more refined trend choice system that considers the place where the given tendencies are typical. In addition, the above arguments about spasmodic effects suggest that there is a faster rate at which trends are created and disappear, meaning the information environment may become more volatile; this has implications for knowledge permanence and discourse depth. As for the future work, there is a clear need for the developers of the platforms to integrate the value of real-time information sharing with the need for diversity and balance in information provision. Based on the information given, we have argued for increased transparency as to when algorithms are used in decision processes and for the integration of safeguards against potential manipulation. Future research should investigate the longevity of such changes in the algorithm's logic for public speaking and information consumption. Finally, ensuring that the digital public space is balanced and accessible would be a focus when it comes to the recommendations for trending on social media.

## REFERENCES

1. Bandari R, Asur S, Huberman BA. The pulse of news in social media: Forecasting popularity. In: Proceedings of the ACM SIGKDD Workshop on Social Media Analytics. 2012. p. 26-34

2. Becchetti L, Castillo C, Donato D, Leonardi S, Baeza-Yates R, Vigna S, et al. Accurate and scalable social recommendation using mixed-membership stochastic block models. In: Proceedings of the Sixth ACM International Conference on Web Search and Data Mining. 2013. p. 123-132

3. Bernstein MS, Little G, Miller RC, Hartmann B, Ackerman MS, Karger DR, et al. Soylent: A word processor with a crowd inside. In: Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology. 2010. p. 313-322

4. Cha M, Haddadi H, Benevenuto F, Gummadi PK. Measuring user influence in Twitter: The million follower fallacy. In: Fifth International AAAI Conference on Weblogs and Social Media. 2010. p. 10-17

5. David G, Günther CW, Zhurakhovska L. Collecting large scale data for multiple emerging technologies with emerging technologies database. In: Proceedings of the 2011 International Conference on Management of

Emergent Digital EcoSystems. 2011. p. 270-275.

6.  Oremus W, Matthews-Ramo N. Twitter's timeline algorithm, and its effect on us, explained. Slate. 2017. Available from: https://slate.com/articles/technology/cover_story/2017/03/twitter_s_timeline_algorithm_ and_its_effect_on_us_explained.html

7.  Twitter Engineering Blog. Twitter Algorithm is now public. Here's what I learned. 2023. Available from: https://medium.com/illumination/twitter-algorithm-is-now-public-heres-what-i-learned-27b9155c3890

8.  GitHub Repository: twitter/the-algorithm. Source code for Twitter's Recommendation Algorithm. 2023. Available from: https://github.com/twitter/the-algorithm

9.   Hootsuite. How the Twitter Algorithm Works. 2024. Available from: https://blog.hootsuite.com/twitter-algorithm/

10. Huszár F, Ktena SI, O'Brien C, Belli L, Schlaikjera A, Hardt M. Algorithmic Amplification of Politics on Twitter. 2021. Available from: https://arxiv.org/abs/2110.11010

11. Mittelstadt B, Allo P, Taddei M, Wachter S, Floridi L. When algorithms fail: Transparency, accountability, and power. 2019. Available from: https://www.nature.com/articles/s42256-022-00504-5

12. Li B, Liao QV. On Predicting Twitter Trend: Important Factors and Models. 2014. Available from: https://www.public.asu.edu/~bli24/Papers/TrendPredictionFull.pdf

13.  Sabeeh S, Al-Furati IS. Comparative Analysis of GA-PRM Algorithm Performance in Simulation and Real-World Robotics Applications. Misan Journal of Engineering Sciences. 2023 Dec 21;2(2):12-37. Available from: https://uomisan.edu.iq/eng/mjes/index.php/eng/article/view/57

14. Al-Hammash-Haitham MS. Enhancement methods of intrusion detection systems using artificial intelligence methods (TLBO) Algorithm. (Humanities, social and applied sciences) Misan Journal of Academic Studies. 2024 Mar 30;23(49):105-12. Available from: https://www.misan-jas.com/index.php/ojs/article/view/574

15.  Al-Bazoon M, Arora J. A hybrid stochastic algorithm with domain reduction for discrete structural optimization. Misan Journal of Engineering Sciences. 2022;1(2):16-36. Available from: https://uomisan.edu.iq/eng/mjes/index.php/eng/article/view/14

## FINANCING

## CONFLICT OF INTEREST
The authors declare no conflicts of interest related to this study.

## AUTHORSHIP CONTRIBUTION
*Conceptualization:* Saif Al-Deen H. Hassan, Hasan Al-Furiji, Mohammed Kareem Rashid, Zahraa Abed Hussein, Bhavna Ambudkar.
*Methodology:* Saif Al-Deen H. Hassan, Hasan Al-Furiji, Mohammed Kareem Rashid, Zahraa Abed Hussein, Bhavna Ambudkar.
*Formal analysis:* Saif Al-Deen H. Hassan, Hasan Al-Furiji, Mohammed Kareem Rashid, Zahraa Abed Hussein, Bhavna Ambudkar.
*Data curation:* Saif Al-Deen H. Hassan, Hasan Al-Furiji, Mohammed Kareem Rashid, Zahraa Abed Hussein, Bhavna Ambudkar.
*Writing - original draft:* Saif Al-Deen H. Hassan, Hasan Al-Furiji, Mohammed Kareem Rashid, Zahraa Abed Hussein, Bhavna Ambudkar.
*Writing - revision and editing:* Saif Al-Deen H. Hassan, Hasan Al-Furiji, Mohammed Kareem Rashid, Zahraa Abed Hussein, Bhavna Ambudkar.