

A COMPREHENSIVE STUDY AND ANALYSIS OF TWITTER TRENDING ANALYSIS USING ITS RANKING MODEL

¹Dr. S. Sujiya, ²Mr. R. Praveen Kumar, ³Mr. M. Praveen Raj.

¹Assistant Professor, Department of Computer Applications, Dr. SNS Rajalakshmi College of Arts & Science, Coimbatore.

²PG Student, II MCA, Department of Computer Applications, Dr. SNS Rajalakshmi College of Arts & Science, Coimbatore.

³PG Student, II MCA, Department of Computer Applications, Dr. SNS Rajalakshmi College of Arts & Science, Coimbatore.

Abstract: - Social media (SM) facts affords a massive document of humanity's ordinary thoughts, emotions, and movements at a resolution formerly unimaginable. Because person conduct on SM is a reflection of activities within the actual global, researchers have realized they could use SM that allows you to forecast, making predictions approximately the destiny. The gain of SM information is its relative ease of acquisition, massive quantity, and potential to seize socially relevant facts, which may be hard to collect from other statistics sources. User's contribution in social media is very essential and it's miles taken into consideration as a precious useful resource. With the rapid boom of social media, Twitter has come to be one of the most broadly adopted systems for people to post short and instantaneous messages. Because of such extensive adoption of Twitter, activities like breaking news and launch of famous movies can without difficulty seize humans' attention and unfold swiftly on Twitter. Therefore, the recognition and significance of an event may be approximately gauged through the volume of tweets overlaying the occasion. In this paper, examines the effect of the twitter and mainly focuses on the manner the ranking model used to prioritize the person facts. The proposed approach detects the trending topics of the real-time Twitter trends along with ranking the top terms and hashtags. The paper further discusses the motivation for trend prediction over the social media.

Keywords: - Social Media, Twitter, Ranking Model, User Data, Posts, Trend Analysis, microblogging.

1. INTRODUCTION

Due to first-rate achievement of the Online Social Network (OSN), a huge quantity of people at the moment are making use of OSN services with a view to benefit energetic collaboration, participation and interplay inside their groups with other users. Twitter, the most important microblogging on-line provider, has won big attention within the beyond few years [1]. Users share and speak everything in this social community. Microblogging is a content-oriented concept wherein humans can interact with others both regarded and unknown. Twitter that is a success micro running a blog social community, has received good sized popularity in recent years.

Twitter, created in 2006, is a web social networking and microblogging carrier which is global popular with greater than 255 million monthly lively users who post 500 million tweets,

handling 1.6 billion search queries in keeping with day [2]. Since this observes focuses on Twitter, standard facts and vital principles to recognize the have a look at is given in this section. Twitter permits its users to put up and read text messages that are one hundred forty-man or woman lengthy, and these messages are called tweets. Tweets are the primary atomic building block of all matters Twitter and are also recognized extra generically as repute updates. Tweets can be preferred by way of a person. Every person in Twitter has a unique username. Users may additionally sub- scribe to different users' tweets and this act is called as following wherein unsubscribing a subscribed consumer is called unfollowing.



Fig 1.1: - Sample Twitter Screen

Twitter is focused on the idea of following that is considered in this look at extensively. Subscribers of a consumer are referred to as fans wherein users subscribed by means of a user are referred to as pals. Friends' posts appear in reverse chronological order on the customers' domestic timeline. More than one Friend will bring about a mix of tweets scrolling down the page, which gets complicated for the user because the wide variety of tweets increases. Retweeting is any other crucial idea that is the act of forwarding a tweet which permits customers to percentage it with their personal fans. There are two different concepts in Twitter that are activated thru symbols [3].

Users can institution posts collectively with the aid of keywords by means of use of hashtags in which phrases or terms prefixed with "#" signal. Similarly, users can refer or reply to other customers the usage of "@" signal observed with the aid of a username, which is known as mentioning [8]. Tweets are publicly visible to all of us by means of default, but users can lock their account and restrict message shipping to just their fans. In this examine, most effective publicly seen tweets are blanketed, content material - username relationship is stored exclusive and personal privateness isn't always violated.

2. LITERATURE REVIEW

Micro-blogging is one of the maximum popular online communication paradigms nowadays, and brief informal messages are shared with the aid of customers to their social networks. Twitter, being the most popular and a success micro-blogging provider, attracts masses of hundreds of thousands of customers international [4]. Tweets posted by using Twitter users are combined with junk mail, nonsense, and self-promotion information; therefore, how to distinguish brilliant data from noisy statistics is the foremost task to make use of Twitter data.

Khan, H. U., et al., (2021), addressed the prediction process becomes challenging as the multiplicity of factors affect the popular elements in the social media content [5]. The proposed work aims to work on Twitter trend analysis and proposes a trend detection process over streams of tweets. The proposed approach detects the trending topics of the real-time Twitter trends along with ranking the top terms and hashtags. The paper further discusses the motivation for trend prediction over the social media; In addition to exploratory data analysis, the research paper explores the Term Frequency-Inverse Document Frequency (Tf-IDF), Combined Component Approach (CCA) and Biterm Topic Model (BTM) approaches for finding the topics and terms within given topics.

Madbouly, M. M., et al., (2020), purposed a new model is proposed to intensify sentiment analysis process on Twitter [6]. In the achievement of such, social network analysis is done using UCINET tool followed by artificial neural networks for ranking users. For sentiment classification, a hybrid approach is presented, where lexicon-based technique is combined with a fuzzy classification technique to handle language vagueness as well as for an inclusive analysis of tweets into seven classes; for the purpose of enhancing final results. The proposed model is practiced on data collected from Twitter. Results show a significant enhancement in tweets polarity scores represent more realistic sentiments.

Alsmadi, I., & Al-Abdullah, M. (2020), focuses in evaluating models to quantify the volume and strengths of interactions between users in OSNs [7]. The authors proposed a reputation rank in OSNs based on a tree graph in which users represent the tree nodes and edges represent their friends' connections or their generated activities. In each OSN such as Facebook, Twitter, LinkedIn, etc. each user will be given a single value that represents the user created activities and friends' interactions with those activities. The model focuses on volumes and statistics of interactions, rather than the content. We also extended the use of cliques' models in OSNs to be normalized or weighted based on the volumes of interactions among clique members.

Tidke, B., Mehta, R., & Dhanani, J. (2020), aims to propose an ensemble learning approach based on multimodal majority voting influence (MMMVI) to identify and weighted multimodal ensemble average influence (WMMEAI) to rank top-k influential nodes in Twitter network data set of well-known three influential nodes, that is, academic institution, scholar and journal [8]. The empirical analysis has been accomplished to learn practicability and efficiency of the proposed approaches when compared with state-of-the-art approaches. The experimental result shows that the ensemble approach using surface learning models (SLMs) can lead to better identification and ranking of influential nodes with low computational complexity.

Setu, S. J., et al., (2020), find the users who are very active in the social networks and consider them as influential users. It addresses on Temporal Twitter Ranking (TTRank) to rank the influential users in Twitter [9]. The authors applied Twitter-LDA topic modeling method to find the users' topical interests. The time interval is an important factor as users' topical interest can change over time i.e. users' have different degree of topical interests at different time interval. The proposed approach also considers the impact of "Follower Influence" and "Retweet Influence". The top influential users have been detected across different time intervals based on all the above mentioned factors and classified as "Highly Influential" and "Potential" users. Experiment results on a real Twitter dataset demonstrate the efficacy of the proposed system.

Devika, R., & Subramaniaswamy, V. (2021), proposed a relatively effective method called Semantic graph-based Keyword Extraction Method (SKEM) from Twitter using ranking methods is proposed [10]. In the proposed model, the exhaustive preprocessing is carried out, and a semantic graph-based model has been constructed. The numeric graph metrics are then used to weigh the nodes of the semantic graph. Page Rank algorithm is applied to arrange the nodes, and the top ten nodes that are found to be very relevant and effectively represent the most influential node. Combining both semantic as well as numeric graph metrics, it greatly enhances the quality of keywords extracted. The extensive preprocessing enhances the quality of the input and minimizes the noise in the input. The keywords extracted by the proposed model have been more relevant and meaningful. The performance of the proposed SKEM model is validated with real-time tweets of Twitter API. The experimental results are confirmed that the proposed method is achieving high performance in terms of precision, recall, and F-measure.

3. TWITTER DATA ANALYSIS AND ITS PROCESS

The proliferation of online social networks has been one of the maximum splendid Internet activities on this decade. Because of the excessive penetration of Internet-enabled devices along with non-public computer systems, smart phones, and capsules, online social networks have grow to be without problems reachable structures for customers to speak and percentage data [11].

Many popular online social networks together with Twitter, Facebook, LinkedIn, and Google Plus had been growing swiftly. On every type of social community statistics, a huge variety of analysis duties may be finished to expose treasured records.

□ **Content-based totally evaluation** studies heterogeneous and unstructured content generated by way of social network customers inclusive of blogs, photos, films, and tags. Some famous evaluation practices consist of opinion mining, fashion detection, sentiment evaluation, collaborative advice, and so forth. This kind of analysis has numerous applications in business, politics, and consumer media studies.

□ **Linkage-based analysis** studies linkages among social community customers. It is used to reveal the structural properties and evolution patterns of social networks, decide essential vertices and social affect, hit upon groups, predict unobserved or destiny hyperlinks, and so forth. This type of analysis is vital for various application fields consisting of social psychology, viral advertising and marketing, and terrorism protection.

The essential goal is to reduce Twitter customers' effort to access to the tweet sporting the statistics of interest. To this purpose, a tweet recommendation device below a user interest model generated thru named entities is provided [12, 13]. The device particularly involves six phases; facts amassing, information base creation, information preprocessing, named entity reputation, consumer hobby version generation based totally on named entities and in the end recommendation. General information at the stages is as follows:

□ **Data Gathering** is the system of collecting a Twitter user's facts; including consumer's pals' posts as well as consumer's very own posts. In this segment, person-buddy courting is also extracted and buddies' relative ranking is generated as an output.

□ **Knowledge Base Construction** is the procedure of generating a graph-based totally expertise base of Turkish Wikipedia article titles and their links to each other, as a way to

validate named entity candidates generated as an output of Named Entity Recognition section. Keeping this understanding base up to date is also blanketed in this segment. Although different levels iteratively follow each other and one's output is the opposites enter, this section is independent and performed in parallel.

□ **Data Preprocessing** consists of doing away with pointless elements of tweet texts together with mentions, hashtags, smileys, vocatives, links and so forth. Since casual writing fashion is commonly followed in tweets, this segment is likewise responsible from normalizing the tweet text such as putting off unnecessarily repeated characters, slang words, correcting asciification associated issues.

□ **Named Entity Recognition** is the subsequent segment of statistics preprocessing phase. In this phase, tweet segmentation on preprocessed tweets is performed by way of international context and segments as candidate named entities are generated. Then, these candidates are demonstrated as named entities or unnoticed by way of utilization of previously constructed understanding base of Turkish Wikipedia article titles.

□ **User Interest Model Generation** phase is a should. In this section, using named entities extracted from user's and user's buddies' tweets and consumer-friend relationships, a user hobby version is generated. In other words, a Twitter person is represented through weighted named entities.

□ **Tweet Recommendation** is the last segment, where kinds of advice programs implemented by using evaluating candidate tweets with the generated consumer interest model. Tweet category that's the project of determining whether a candidate tweet is thrilling for the user or now not, and tweet rating which aims to type tweets from the most recommendable to the least recommendable are performed on this segment.

4. TWITTER RANKING METHODS

For each candidate tweet in the tweet set of a input query tweet Q_t , It find a hard and fast of features primarily based on its content material and the consumer who posted the tweet. Based on these functions, we use different ranking strategies to compute the score for hashtags. The tweet content related function includes tweet text, hashtag recognition and temporal information of the tweet [13]. And the consumer associated feature consists of mutual buddies, mutual fans and social signals like @mentions, favorites and co-occurrence of hashtags. And then combine all capabilities ratings to compute the rank for the candidate hashtag set to endorse the K maximum promising hashtag for a given question.

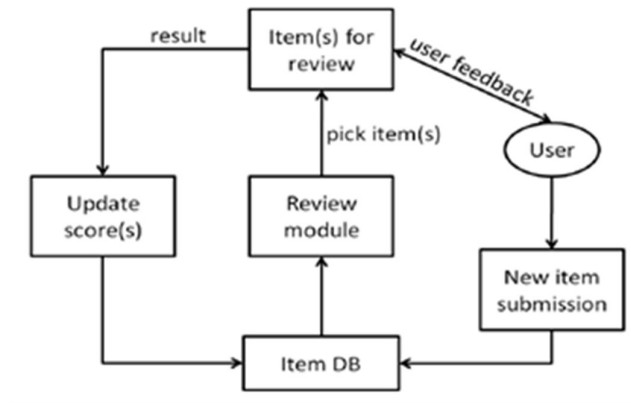


Fig 4.1: - Twitter Ranking Methods

Each candidate tweet within the set will have its own illustration of function scores, which are computed w.r.t to the features of an input query tweet [15, 16]. It describes the ranking methodologies in detail within the following sections.

□ **Similarity Measure:** - This ranking method is based totally at the similarity between the tweet text of input question tweet Q_i and the entries contained within the set of tweets T_j with hashtag belonging to the social graph of the consumer who published the question tweet Q_i . It assumes the entries T_j that contain comparable text contents could have more chance closer to Q_i . So, this ranking technique gives greater weights to the hashtag belonging to the entries of T_j that have extra phrase similarity to Q_i . In this example of trying to find the great matching entry for certain enter question tweet, cosine similarity is computed among the query tweet and each and every tweet inside the candidate tweet set.

□ **Recency Score:** - Temporal function of a hashtag is one of the maximum crucial functions in ranking hashtags. Based in this ranking technique, the hashtag that is temporally near sufficient to the question tweet is maximum in all likelihood to be utilized by a user. Especially inside the case of dynamic microblogging systems like twitter, the trending topics that keep evolving are based totally on how frequently a hashtag is lately used. In this ranking approach, hashtags which can be temporally in the direction of the query tweet receives better ranking.

□ **Social Trend Score:** - The social fashion rank computes the recognition of hashtags in the candidate hashtag set H_j of a specific question tweet Q_i and it varies for every Q_i . As the candidate hashtag set H_j is derived from the timeline of the person U_x who published the question tweet Q_i , it's miles intuitive that the hashtag with excessive frequency is popular in the user's U_x social community. This also can be taken into consideration as tailored trending subjects for a selected person. Social Trend rank is computed based at the "One character, One vote" technique. It is used to get the rely of the frequently used hashtag in H_j . Only few hashtag are discovered to have high frequency based on its common utilization. In this ranking method, the hashtag with high frequency is most probably a popular hashtag and will receive a better rating.

□ **User Related Feature Score:** - The attention rank captures the current conversation between two customers based totally on their @mentions and replies. This ranking technique ranks the pals based on the current conversations. If a particular person was @referred to these

days, it's miles much more likely that they share topic of common pursuits. This additionally way they could use comparable hashtags. Here the @mentions and replies acts as a social signal emitted between the users. This also enables to determine the tie-electricity between the users.

5. CONCLUSION

Twitter and other social networking web sites are more and more used to stay linked with the pals, tuned to their reputation updates, having every day chatters, conversations, facts sharing and information reporting. The popularity and uncured nature is susceptible to noise. With a large social network in Twitter, statistics overload is a daunting thing that limits user's social engagement and less energetic on-line users. But customers in Twitter invented the atypical conventions of using hashtag as a context or metadata tag for the tweet. Even though hashtags allows deriving the context for a tweet; they're not well adapted with the aid of Twitter customers. So a largest unmarried thing that could persuade customers to get engaged of their social network is to have a better recommendation device that recommends suitable hashtag for the tweets. This could inturn might boom the net social engagement. Twitter's software grows as one's network grows; and one's network grows the greater they interact with others. This in turn might additionally effect Twitter to hold its monthly active online users in its social community.

REFERENCES

- [1] Kwak, H., Lee, C., Park, H., & Moon, S. (2010, April). What is Twitter, a social network or a news media?. In Proceedings of the 19th international conference on World wide web (pp. 591-600).
- [2] Riquelme, F., & González-Cantergiani, P. (2016). Measuring user influence on Twitter: A survey. *Information processing & management*, 52(5), 949-975.
- [3] Alrubaian, M., Al-Qurishi, M., Al-Rakhami, M., Hassan, M. M., & Alamri, A. (2017). Reputation-based credibility analysis of Twitter social network users. *Concurrency and Computation: Practice and Experience*, 29(7), e3873.
- [4] Singh, D. K. S., Nithya, N., Rahunathan, L., Sanghavi, P., Vaghela, R. S., Manoharan, P., & Tunze, G. B. (2022). Social network analysis for precise friend suggestion for twitter by associating multiple networks using ml. *International Journal of Information Technology and Web Engineering (IJITWE)*, 17(1), 1-11.
- [5] Khan, H. U., Nasir, S., Nasim, K., Shabbir, D., & Mahmood, A. (2021). Twitter trends: a ranking algorithm analysis on real time data. *Expert Systems with Applications*, 164, 113990.
- [6] Madbouly, M. M., Darwish, S. M., & Essameldin, R. (2020). Modified fuzzy sentiment analysis approach based on user ranking suitable for online social networks. *IET software*, 14(3), 300-307.
- [7] Alsmadi, I., & Al-Abdullah, M. (2020). A model for reputation rank in online social networks and its applications. *International Journal of Social Network Mining*, 3(1), 77-98.
- [8] Tidke, B., Mehta, R., & Dhanani, J. (2020). Multimodal ensemble approach to identify and rank top-k influential nodes of scholarly literature using Twitter network. *Journal of Information Science*, 46(4), 437-458.

- [9] Setu, S. J., Islam, T., Bhuiyan, M. A. A., & Anwar, M. M. (2020). TTRank: A Temporal Model to Rank Online Twitter Users. *International Journal of Automation, Artificial Intelligence and Machine Learning*, 1(1), 42-53.
- [10] Devika, R., & Subramaniaswamy, V. (2021). A semantic graph-based keyword extraction model using ranking method on big social data. *Wireless Networks*, 27(8), 5447-5459.
- [11] Drakopoulos, G., Kanavos, A., & Tsakalidis, A. K. (2016, April). Evaluating Twitter Influence Ranking with System Theory. In *WEBIST (1)* (pp. 113-120).
- [12] De Silva, T. H. M. (2021). A Network Analysis Based Credibility Ranking Model to Combat Misinformation on Twitter (Doctoral dissertation).
- [13] Aggrawal, N., Ahluwalia, A., Khurana, P., & Arora, A. (2017). Brand analysis framework for online marketing: ranking web pages and analyzing popularity of brands on social media. *Social Network Analysis and Mining*, 7, 1-10.
- [14] Mittal, D., Suthar, P., Patil, M., Pranaya, P. G. S., Rana, D. P., & Tidke, B. (2020). Social network influencer rank recommender using diverse features from topical graph. *Procedia Computer Science*, 167, 1861-1871.
- [15] Bibi, F., Khan, H. U., Iqbal, T., Farooq, M., Mehmood, I., & Nam, Y. (2018). Ranking authors in an academic network using social network measures. *Applied Sciences*, 8(10), 1824.
- [16] Curiskis, S. A., Drake, B., Osborn, T. R., & Kennedy, P. J. (2020). An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit. *Information Processing & Management*, 57(2), 102034.