

Empowering influence discovery: Utilizing machine learning for social media influencer identification

D.D.W.N. Devyanjalee✉, I.U. Hewapathirana

Software Engineering Teaching Unit, University of Kelaniya, Dalugama, Kelaniya, Sri Lanka

✉nethminid1999@gmail.com, +94773910188

In today's dynamic digital landscape, influencer marketing has become a cornerstone of marketing strategies, leveraging social media platforms to engage with audiences. Accurately identifying influencers within social media platforms poses a formidable challenge. Traditional machine learning approaches relying solely on metrics such as network analysis and user profile data, often fall short in capturing the dynamics of influencer impact and resonance with audiences. To address this gap, this study aimed to enhance influencer identification accuracy by leveraging both user profile and engagement metrics alongside text analysis. The methodology adopts a sequential explanatory design, combining quantitative analysis of user profile metrics with qualitative analysis of text-related factors. Data collection from social media platforms, particularly X, comprises user profile and social data. The quantitative phase employs established algorithms like the PageRank algorithm to identify top influencers based on user profile data, while machine learning models, logistic regression, decision trees, and random forest are trained using user profile data to discern influential user profiles. The qualitative phase involves text analysis techniques, including keyword matching and lemmatization, to extract valuable insights from tweets. Machine learning models are then trained using both user profile and social data alongside text analysis data to discern influential user profiles. The models are then compared to assess the impact of incorporating engagement metrics with text analysis. Findings from this study indicate that while user profile metrics alone exhibit high accuracy in influencer identification, with the random forest model achieving an F1 score of 0.90, the incorporation of engagement metrics introduce complexities affecting model performance, resulting in an F1 score of 0.70. The random forest model emerges as the most robust performer, maintaining high accuracy despite these challenges. This research contributes to advancing influencer identification strategies within digital marketing, offering insights into the effectiveness of integrating both user profile and engagement metrics with text analysis for capturing the true essence of influencer influence and resonance with audiences. The findings underscore the challenges of leveraging engagement metrics for influencer identification and highlight the need for further refinement of methodologies to empower marketers in navigating the complexities of the ever-evolving digital landscape.

Keywords: Influencer identification; social media analytics; machine learning models; text analysis techniques