الإنتاج العلمي لمركز بحوث علوم الحاسب والمعلومات

عمادة البحث العلمي
Deanship of Academic Research

VISION
2030
المملكة العربية السعودية
KINGDOM OF SAUDI ARABIA

An Ensemble Learning Based Approach for Detecting and Tracking COVID19 Rumors

Authors	Sultan Noman Qasem, Mohammed Al-Sarem, and Faisal Saeed

Abstract: Rumors regarding epidemic diseases such as COVID 19, medicines and treatments, diagnostic methods and public emergencies can have harmful impacts on health and political, social and other aspects of people's lives, especially during emergency situations and health crises. With huge amounts of content being posted to social media every second during these situations, it becomes very difficult to detect fake news (rumors) that poses threats to the stability and sustainability of the healthcare sector. A rumor is defined as a statement for which truthfulness has not been verified. During COVID 19, people found difficulty in obtaining the most truthful news easily because of the huge amount of unverified information on social media. Several methods have been applied for detecting rumors and tracking their sources for COVID 19-related information.However, very few studies have been conducted for this purpose for the Arabic language, which has unique characteristics. Therefore, this paper proposes a comprehensive approach which includes two phases: detection and tracking. In the detection phase of the study carried out, several standalone and ensemble machine learning methods were applied on the Arcov-19 dataset. A new detection model was used which combined two models: The Genetic Algorithm Based Support Vector Machine (that works on users' and tweets' features) and the stacking ensemble method (that works on tweets' texts). In the tracking phase, several similarity-based techniques were used to obtain the top 1% of similar tweets to a target tweet/post, which helped to find the source of the rumors. The experiments showed interesting results in terms of accuracy, precision, recall and F1-Score for rumor detection (the accuracy reached 92.63%), and showed interesting findings in the tracking phase, in terms of ROUGE L precision, recall and F1-Score for similarity techniques.