

## **Social Media and Online News Analytics for Identifying Crime Patterns in Crime Prediction**

S.P.C.W Sandagiri, B.T.G.S Kumara and K. Banujan

*Department of Computing & Information Systems, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka*

Social media provides opportunities for users to share their thoughts freely. Every year they generate a large volume of data. In the context of social media, they may include hidden details, which may convey significant events. Crime prediction with the help of Social media provides new dimensions in researches. This research aims to collect data from Twitter posts and validate them using online news to avoid false data. First and foremost, we selected the top crimes happening in the world after an extreme literature review. We used Twitter API and News API to fetch data from Twitter and News blogs. We used two filters to collect data. In the first filter, we fetch Twitter posts and News posts for a specific time duration. These data are fetched by using keywords that relate to crime. In the second filter, eliminate noisy Twitter posts from the collected dataset. We have collected many noisy posts in both sources, i.e. Twitter and News. With the help of collected datasets, we will compare each tweet and news datum and give ratings for comparison data. We can build a crime prediction model with integrating data. The result shows that 68% of collected Twitter posts are excluded after using the second filter. Future development can divide into two main parts. To get more accuracy, we can integrate other factors that affect crime prediction such as weather, human behavior analysis data and we can improve the second filter using the SVM algorithm. Secondly, we can integrate other Social media platforms to fetch data.

*Keywords:* Crime prediction, Social media, Twitter, News