

**Abstract No: SO-10**

**Comparison between the Machine Learning Algorithms to determine the suitable input features for personal theft, sexual assault, and house burglary victimization prediction**

W. S. V. Lakshan<sup>1\*</sup> and W. A. C. Weerakoon<sup>1</sup>

<sup>1</sup>Department of Statistics and Computer Science, Faculty of Science, University of Kelaniya, Sri Lanka  
wsvla191@kln.ac.lk\*

The crime-related predictions can be vastly supported by most of the available supervised machine learning models. The possibility of becoming a victim increases daily in each crime category. The main difficulty is to find how severe the impact is upon the victim after the crime. Here, the Random Forest, Support Vector Machine (SVM), K-Nearest Neighbour (KNN) algorithm, and Neural Network models were compared with the use of available features found from a secondary dataset to build a better prediction model, which has been implemented in four main phases over two aspects based on the possibility of becoming a victim and severity of the crime. The available features were used as the inputs for phase I and Principal Component Analysis and correlation tests were performed to identify the appropriate and essential feature combinations for the rest of the phases. The pre-processed datasets were used to implement and train the models. Moreover, the Random Forest model was proven to be the most efficient model with an accuracy of 85.33% in phase four when comparing the accuracy levels of the models over different phases, while the KNN and Neural Network models obtained an accuracy of over 70% and SVM obtained the least accuracy in the same phase. In phase one, the Random Forest algorithm was executed with a precision of 76%, while KNN and Neural Network model obtained around 70%. The final outputs obtained for phase four showed that factors such as age, year, gender, race, and relationship with the perpetrator will be the most suitable features to build an accurate machine learning model for victimization prediction. The mentality level of the offender and intention of doing it has the main impact on the severity level. Also, authorities need to keep track of the fact whether it is a repeat offence or not, the main offender or not and the contribution of the offender to support better information inputs for the prediction models. This study developed a victimization prediction model with reference to personal theft, sexual assault, and house burglary. This would be a step forward from previous research works of rule-based victimization possibility index prediction for small victim clusters. Further, new features were identified in the last phase, which can be used to develop models to predict criminal behaviour after sending them back to the society. This will greatly benefit the authorized bodies to monitor them and reduce the possibility of victimization.

**Keywords:** Correlation, Machine Learning, Neural Network, Severity, Victimization