

Abstract No: PO-12

**A study on factors associated with child sexual abuse and recognizing the severity:
Special reference to Galle district**

L. H. K. Dilshan^{1*}, N. Withanage² and N. V. Chandrasekara¹

¹Department of Statistics and Computer Science, Faculty of Science, University of Kelaniya, Sri Lanka

²Department of Statistics, Faculty of Science, University of Sri Jayewardenepura, Sri Lanka
kavindudilshan19971029@email.com*

Child Sexual Abuse (CSA) has been a universal and social crisis with serious life-long consequences. One in four girls and one in six boys worldwide have experienced some form of sexual abuse in their childhood. According to Police statistics, CSA cases have been increasing rapidly in recent years in Sri Lanka. Galle is among the four districts where the reported child abuse cases are high, and the reported CSA complaints are rising drastically. Further, no previous study has been carried out in the Southern part of the island regarding the crisis of CSA. Therefore, the main objective of this study is to determine the key risk factors affecting the CSA cases in Galle Police Division and to develop suitable statistical and machine learning models to recognize the severity of CSA. All the 225 CSA cases reported to the Police Child and Women Bureau of Galle Police Division during the 2017 – 2020 period were considered for this study. The severity of CSA can be categorized into not fatal, child sexual exploitation, and fatal categories. Out of the twenty-one risk factors, which were found from the literature and knowledge of domain experts, sixteen factors showed a significant relationship with the severity of CSA at 10% significance level according to the chi-square test of association. These significant risk factors were area, child's age, gender, whether mother lives with child, reason, the willingness of child, frequency of abuses, place of incident, relationship to the perpetrator, perpetrator's age, education level of the perpetrator, perpetrator's job, marriage status, whether the perpetrator has children, the number of children he has, and drug addiction of perpetrator. The Ordinal Logistic Regression (OLR) model was trained using a backward selection method with different data selection criteria. Next, the machine learning techniques: Decision Tree (DT), Support Vector Machine (SVM), and Probabilistic Neural Network (PNN) were employed to predict the severity of CSA. The random over-sampling technique was used to overcome the class imbalance problem that persists in the dataset. The bagging technique was implemented to preserve the robustness of the models and to improve their performance. The adequacy of the OLR model with the oversampling technique was examined and it was selected as the best model after considering the proportional odds assumption and analysis of deviance. The model classified the severity of CSA with 68.85% accuracy and area, gender, reason, frequency of abuses, place, perpetrator's job, and whether the perpetrator has children can be identified as the significant predictors for CSA. The DT, SVM and PNN models classified the severity of CSA with an accuracy of 82.15%, 77.68% and 81.25%, respectively for the bagging technique. The PNN model performed better than the other fitted models with higher accuracy. The results obtained from this study can be used to get precautions and to arrange awareness sessions for parents and adults to reduce CSA in Galle Police Division. Similarly, the scope of the study can be extended to the whole island to reduce CSA and to make a better place for children.

Keywords: Bagging technique, Child sexual abuse, Machine learning techniques, Ordinal logistic regression, Oversampling