

Universidad de Lima
Facultad de Ingeniería y Arquitectura
Carrera de Ingeniería de Sistemas



COMPARISON OF CLASSIFIERS MODELS FOR PREDICTION OF INTIMATE PARTNER VIOLENCE

Tesis para optar el Título Profesional de Ingeniero de Sistemas

Ashly Mercedes Guerrero Muguerza

Código 20150626

Asesor

Juan Manuel Gutiérrez Cárdenas

Lima – Perú

Marzo de 2021

Comparison of Classifiers Models for Prediction of Intimate Partner Violence

Ashly Guerrero, Juan Gutiérrez Cárdenas(✉), Vilma Romero,
and Víctor H. Ayma

Universidad de Lima, Lima, Peru
hash.ash1505@gmail.com,
{jmgutier,vromero,vayma}@ulima.edu.pe

Abstract. Intimate partner violence (IPV) is a problem that has been studied by different researchers to determine the factors that influence its occurrence, as well as to predict it. In Peru, 68.2% of women have been victims of violence, of which 31.7% were victims of physical aggression, 64.2% of psychological aggression, and 6.6% of sexual aggression. Therefore, in order to predict psychological, physical and sexual intimate partner violence in Peru, the database of denouncements registered in 2016 of the “Ministerio de la Mujer y Poblaciones Vulnerables” was used. This database is comprised of 70510 complaints and 236 variables concerning the characteristics of the victim and the aggressor. First of all, we used Chi-squared feature selection technique to find the most influential variables. Next, we applied the SMOTE and random under sampling techniques to balance the dataset. Then, we processed the balanced dataset using cross validation with 10 folds on Multinomial Logistic Regression, Random Forest, Naive Bayes and Support Vector Machines classifiers to predict the type of partner violence and compare their results. The results indicate that the Multinomial Logistic Regression and Support Vector Machine classifiers performed better on different scenarios with different feature subsets, whereas the Naïve Bayes classifier showed inferior. Finally, we observed that the classifiers improve their performance as the number of features increased.

Keywords: Intimate partner violence _ Random forest _ Multinomial logistic regression _ Support Vector Machine _ Naïve Bayes _ SMOTE

Proceedings of the Future Technologies Conference (FTC) 2020. Advances in Intelligent Systems and Computing, vol 1289. Springer

© 2020 Springer Nature Switzerland AG. Part of Springer Nature.

https://doi.org/10.1007/978-3-030-63089-8_30