Decision Tree Induction of Emotional Violence against Women

Esra Kahya Özyirmidokuz[1,a,*] and Yasemin Kaya[2,b]

[1] Kayseri Vocational College, Erciyes University, Kayseri, Turkey

[2]Social Science Vocational College, Erciyes University, Kayseri, Turkey

[a]esrakahya@erciyes.edu.tr, [b]ykaya@erciyes.edu.tr

*Corresponding author

Abstract. Knowledge Discovery and Data Mining (KDD) is an interdisciplinary area focusing upon methodologies for extracting useful and hidden knowledge from data. KDD is the process of analyzing data from different perspectives and summarizing it into new information. It uses various types of algorithms including statistical and machine learning. KDD starts with understanding data. Then data preprocessing techniques are used because quality decisions must be based on quality data. Data mining modeling phase is used to extract more meaningful knowledge from a given data set. Decision trees (DTs) are one of the most powerful and popular approaches in data mining. Their effective structure helps us to make the best decisions on the basis of existing information. DTs also predict future and possible decisions.

Analyzing violence against women data is a very important subject. There are still a lot of women subjected to violence in the 21st century. Violence against women is not only a physical, emotional violence negatively affects the health. Emotional violence acts by a partner included: being insulted or made to feel bad about oneself; being humiliated in front of others; being intimidated or scared on purpose; or being threatened directly, or through a threat to someone the respondent cares about. Although violence against women, especially by men, is important issue, presently there are too few original researches on these topic. Women who experience emotional violence are at increased risk of injury and death, as well as a range of physical, emotional, and social problems. Women play important roles in healthy individuals in the community to grow it healthy, violence against women is also a problem for the society. This research aims to extract meaningful knowledge from emotional violence against women data set. The data covers 12795 women in the age group of 15-59 in Turkey. The data is collected from Turkey Statistical Institute (TURKSTAT)'s Web site. As a result of the data obtained from the analysis can be submitted suggestions for women exposed to emotional violence.

The paper explores the use of different preprocessing and data mining techniques (anomaly detection analysis, data reduction, data cleaning, DT and etc.). In order to support the prevention of violence against women, a DT model is generated. Modeling is performed with C4.5 which is one of the popular DT modeling algorithms, an extension of the earlier well known algorithm ID3 to find useful patterns.

The cross-validation method involves partitioning the examples randomly into n folds. Ten fold normal cross-validation accuracy evaluation is used.

Consequently, a DT is generated. The resulting DT model can be used for the future decisions. The relationships of the input variables with target variables are found. The results show that the proposed method has substantially good performance.