Using Machine Learning To Build A Predictive Model Of No Recidivism- A Pilot Analysis Using Decision Tree

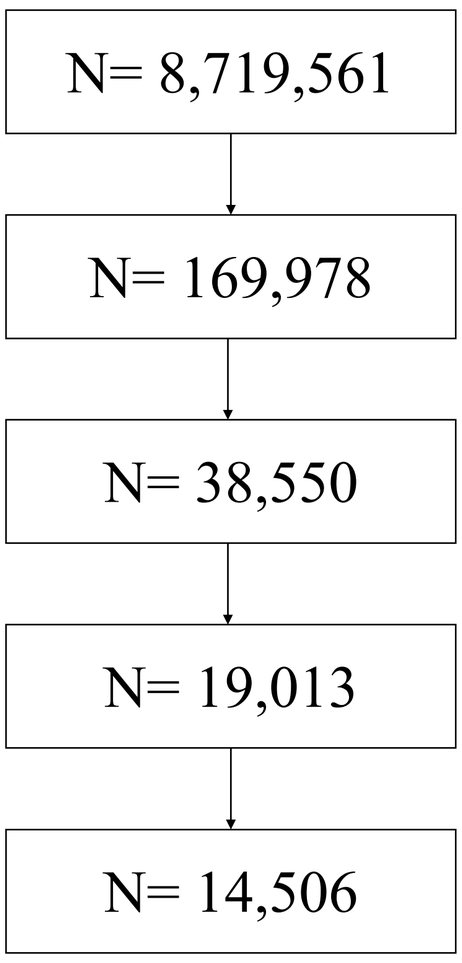
En Chao[1,2], Yi-Chien Ku[1,3], Hung-Ping Chung[4], Chia-Yu Hsu[3], Yuan-Hao Cheng[3],
Freya In-Chu Hsu[3], Yi-Chia Tsai[3], Tony Szu-Hsien Lee[1,5]

[1]CTBC Center for Addiction Prevention and Policy Research, National Taiwan Normal University, Taiwan
[2]Tri-Service General Hospital Songshan Branch, Taiwan
[3]Crime Prevention Research Center, Academy for the Judiciary, Taiwan
[4]Institutum Iurisprudentiae, Academia Sinica, Taiwan
[5]Department of Health Promotion and Health Education, National Taiwan Normal University, Taiwan

Introduction

According to national statistics on drug abuse released by Taiwan's Ministry of Health and Welfare, heroin was the most commonly reported form of drug abuse in 2020, with heroin users comprising more than 90% of Schedule One drugs among the top 13 reported drugs. In addition to prison sentences, drug offenders were asked to either participate in institutional rehabilitation, mandatory treatments in correctional facilities, treatment in a hospital or community interventions.

Deferred Prosecution with Condition to Complete the Addiction Treatment (DPCCAT) is currently Taiwan's judicial priority for dealing with drug users, but there are few scientific empirical studies. This study is to use relevant big data of those who complete DPCCAT to establish a predictive model about no drug use recidivism within one year.

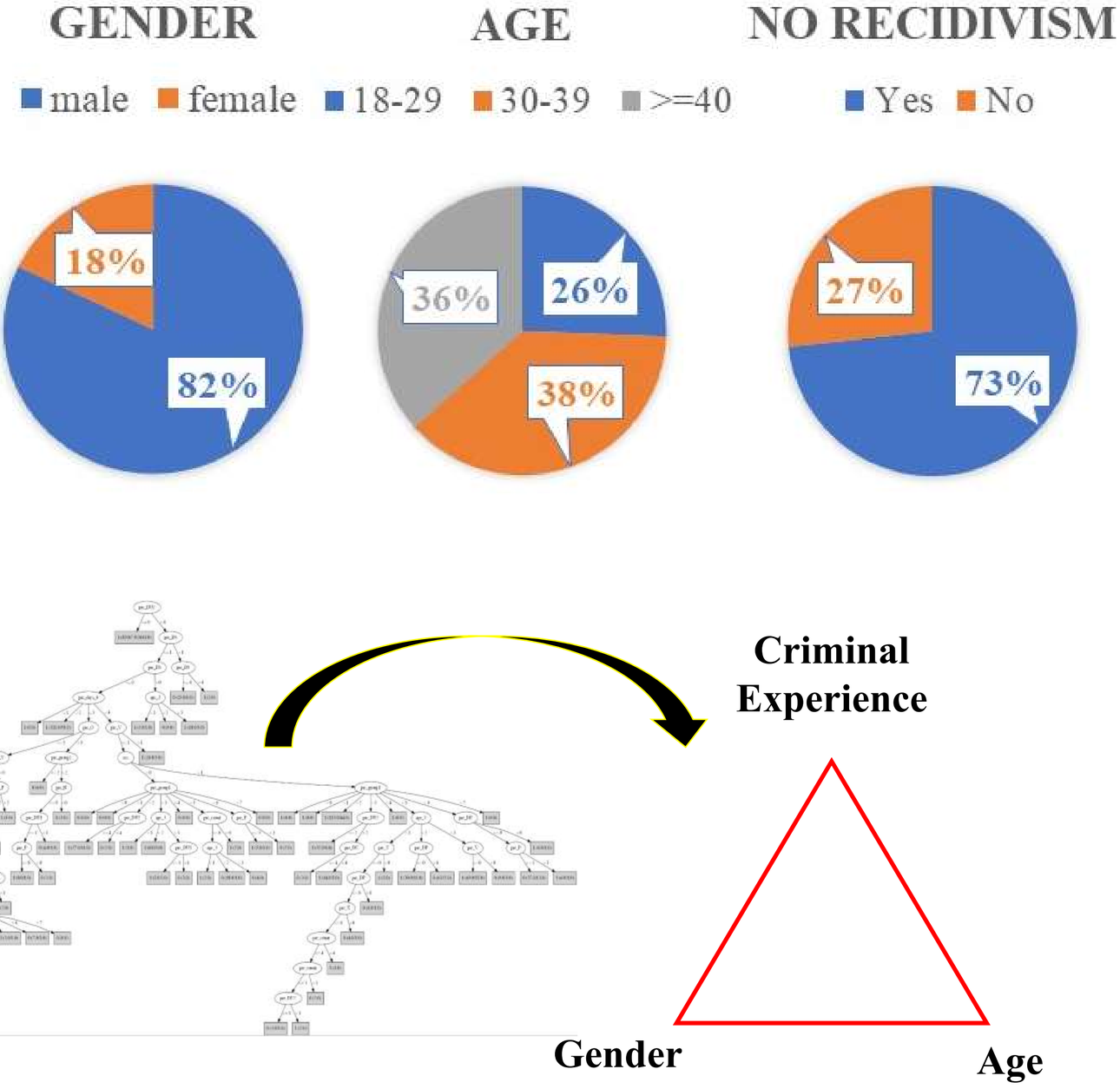


Materials and Methods

The population based data is from the investigation database of the Ministry of Justice of Taiwan, January 2008 to July 2020. We uses Weka 3.8's decision tree classifier for prediction. The ratio of training data to test data is 8:2. The input prediction model data included gender, age and criminal record for participating in the DPCCAT.

Results

A total of 14,506 people were included in the analysis, of which 11,912 (82%) were male, 2,594 (18%) were female, 3,715 (26%) were aged 18-29, and 5,482 (38%) were aged 30-39, and 5,309 people (37%) over the age of 40, 10,630 people (73%) who successfully have no drug use recidivism within 1 year after completing DPCCAT. After using the training data to build the model and substituting the test data, the prediction model performance recall rate = 0.99, precision rate = 0.74 and F1 measure = 0.85.



Conclusions

The model in this study can effectively predict the success of no drug use recidivism within 1 year for people who have completed DPCCAT. Using machine learning to analyze such data may be helpful in the future, and adding dynamic factors should make the model perform better.

Further Information

En Chao：johnny2cindy@gmail.com
Tony Szu-Hsien Lee：tonylee@ntnu.edu.tw