Classification with Drug Consumption

Tao Shan

List of contribution

Tao Shan: Coding for Stepwise feature selection by Stepclass package, Xgboost, and Neural Network (also check each of the coding parts for whole modeling steps). Presenting Discussion & Conclusion in final presentation. Writing the report for Stepwise feature selection, Xgboost, Neural Network, Result, Discussion & Conclusion.

Janny Liu: Coding for Decision Tree in the module of feature selection, Random Forest in the module of model prediction. Presenting Random Forest, SVM, Boosting, and XGBoost in the module of Model Prediction in the final presentation. Writing the report for Decision Tree in the module of feature selection and Random Forest in the module of Model Prediction. Checking for all of the grammar and sentence structures in this final report.

Chen Dan: Coding for Random Forest in the module of feature selection, Boosting in the module of model prediction. Presenting feature selection section in final presentation. Writing the report for Random Forest in the module of feature selection and Boosting in the module of Model Prediction.

Ziwei Pan: Coding for data exploration, Support Vector Machine (SVM) in the module of model prediction in the final presentation. Presenting the part of data exploration, SVM in the final presentation. Writing the report of data exploration and SVM model prediction.

Zhengqi Sun: Coding for data cleaning, KNN in the module of model prediction. Presenting Introduction to the dataset, motivation and steps of data cleaning in the final presentation. Writing the report of description of data set and KNN algorithm in the introduction and model prediction.

Xinwei Lyu: Coding for Decision Tree in the module of model prediction. Presenting Neural Network, KNN, and Decision Tree in model prediction part in final presentation. Writing the report for Introduction part, and Decision Tree part in the module of Model Prediction.

Introduction

In this report, our group focuses on the drug consumption dataset, finding the relationship between the most influential variables and others through a series of measuring methods, and making predictions for people's drug consumption activity, which will be introduced in the following sections.

There are 49.6% of individuals aged 18 to 49 years old who have lifetime substance use disorders in the original dataset, while 21.8% have temporary negative situations (Brown et al., 1998). Also, the figures in this data are still increasing dramatically among college students. This circumstance motivates us to analyze the related data and find an effective way to analyze this situation. The drug use problem is quite important for all individuals. If we can seek out the result that causes the high potential of using drugs, we may be able to eliminate most of the negative influences from drug consumption. Specifically, the depression and anxiety rate are likely to decrease, some worse physical reactions like the feeling of disgust and insomnia problems are also possible to be relieved (Shoal & Giancola, 2001). Besides, the external factors, like the crime rates and alienation of the relation between family members will turn better as well (Caldeira et al., 2009).

Studies show that numerous people analyzed drug-related reports previously. For instance, a report from Ginoux and other researchers (Ginoux et al., 2019) state that they use a Predator-prey-based Model to predict marijuana consumption in the states of Colorado and Washington. This model is modified from a previous model called NERA (Ginoux et al., 2019). Ginoux and his peers set drug users as predators and non-drug users as prey. Meanwhile, they subdivide predators into three categories, which are E (experimental), R (recreational), and A (addict). Categories are dependent on the consumption frequency and individuals' self-control. Similarly, researchers use N (non) to represent prey. The data will be calculated and analyzed by genetic algorithms (Ginoux et al., 2019). After reviewing the associated research, our main task in this report is using classification to find out the most affectable variables, and make predictions for the frequency of using drugs.

Description of the Dataset

For the above purpose, we are choosing to use the dataset that relates to drug consumption's research on the UCI machine learning repository.

(url: https://archive.ics.uci.edu/ml/datasets/Drug+consumption+%28quantified%29)

Originally, the original dataset has 32 columns, which are the attributes in this dataset, and 1885 observations. The columns are age, gender, education, country, ethnicity, biological measurements (NEO-FFI-R score, BIS-11 score, ImpSS score), the interval time of the use of 18 legal and illegal drugs which includes alcohol, amphet, amyl, caff, cannabis, choc, coke, crack, ecstasy, heroin, ketamine, legalh, lsd,meth, mushrooms, nicotine, semer, and vsa.

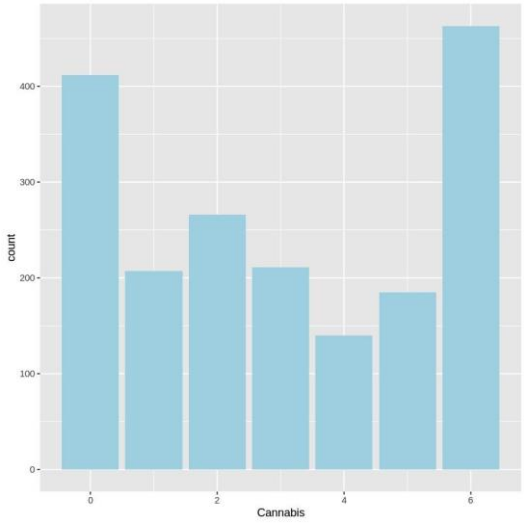
This dataset observes 1885 people's drug consumption before 2016 in different countries which are the UK, USA, Canada, Australia, Republic of Ireland, New Zealand and other countries. The ethnicities of these people include white, black, asian, mixed-white/asian, mixed-black/asian and mixed-white/black. White people are the majority of the sample population, meanwhile 1043 people are from the UK in this data, which means most of this dataset is collected from the British whites. The age of the sample population is from 25 years old to 65+ years old, that are divided into 6 classes, which means it is a categorical variable, same as the education, there are 9 classes of education levels in this data which are doctorate, professional certificate, master degree, university degree, some college or university, left school at 16/17/18 years and left school before 16 years. In this dataset, the biological measurement scores are numeric and other variables are categorical.

We worked on finding the best machine learning model to predict whether an individual is a cannabis user or not. Some personality traits are observed such as personality traits and we test the accuracy of the response variable. The independent variables about the use of drugs are divided into 7 classes, which are

CL0, CL1, CL2, CL3, CL4, CL5 and CL6, represents never used, used over a decade ago, used in Last year, used in Last month, used in last week, and used in last day.

Data Exploration

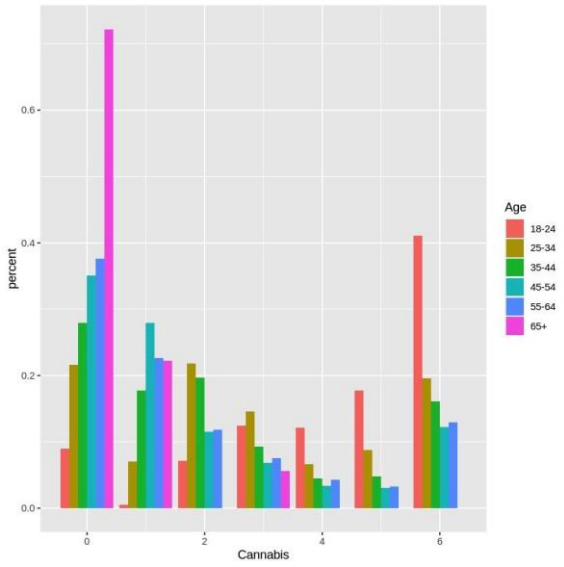
Data exploration is a useful approach to understand the characteristics of data by method of visualization and some introductions of useful key tools or concepts. In addition, it is also about efficiently extracting knowledge from data even if we do not know exactly what we are looking for (Idreos, 2015). Our group mainly uses some bar charts to show the features, since the target variable is Cannabis, then the graphs are highly related to the variable Cannabis.



To start with, we draw a count bar chart of the Cannabis. For x-axis, as stated before, number of x-axis means the level of 0 to 6, CL0 to CL6 that is discussed in the description of the dataset. It represents the time period of taking Cannabis last time. We can conclude that taking Cannabis last day has the largest proportion and also a relatively massive number of people never took Cannabis. Therefore, the distribution of taking Cannabis has the feature of polarization of CL0 and CL6.

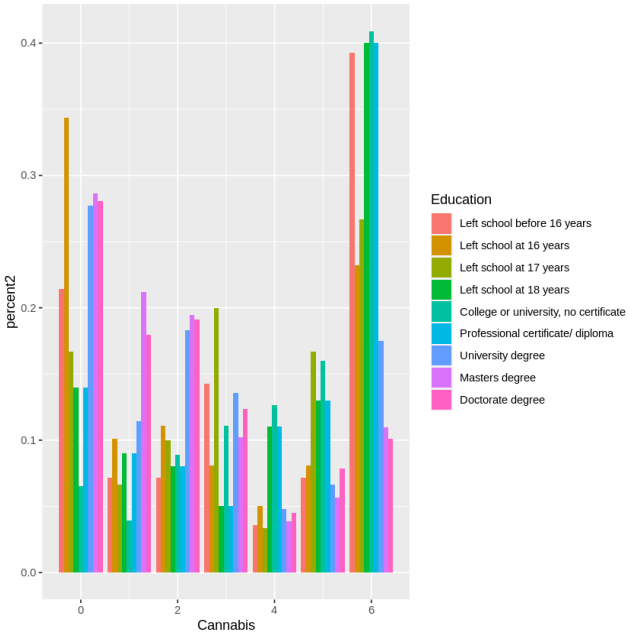
Next step, we include another categorical variable to divide up the data, and use a graph with grouped bars. Since the population in each level of variable is not the same scale, it is meaningless to count the number of people in groups directly. Calculating the percentage of each group shown in the plot will display the distribution of the Cannabis divided by another variable. Also, based on these plots, we

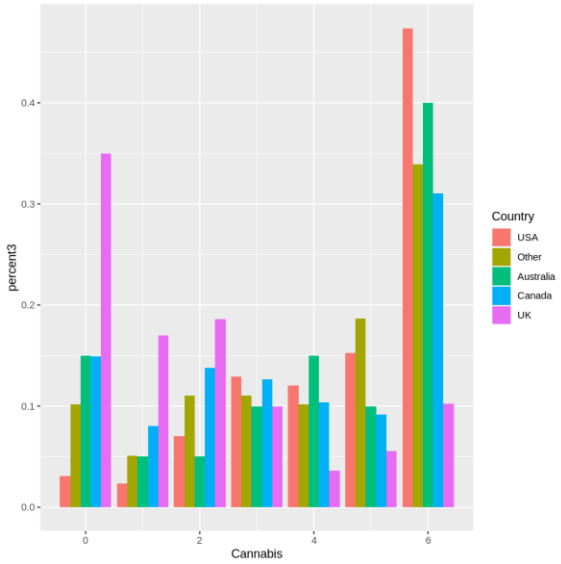
compare the same level of variables in different levels of Cannabis and contrast different levels of variables in the same level of Cannabis.



Based on the Age-Cannabis plot, we can find that young adults from 18 to 24 years old take Cannabis the most frequently whereas the majority of senior people who are more than 65 years old never took Cannabis before. With the category of Cannabis, from CL1 to CL4, we conclude similar percentages of each level of age.

Based on the Education-Cannabis plot, if we focus on the Cannabis level of CL6, we can notice that level of education such as leaving school before 16 years, or people who enrolled in college or university but failed to get a certificate has the largest proportion of people. If we focus on the Cannabis level of CL0, we can find higher levels of education has a relatively high proportion, which both follow an intuitive cognition of people.





Moreover, based on Country-Cannabis plot, as the aspect of country, we can find that people in USA have a more severe problem of taking Cannabis because the population is increasing from CL0 to CL6; and interestingly, people in UK are less likely to take Cannabis because the largest proportion lies in CL0.

Feature Selection

Feature selection is a process of getting an optimal subset from the original feature space, which is based on given standards to determine those features are important and useful (Kuhn, 2019). It is useful to improve the speed and improve the classification accuracy.

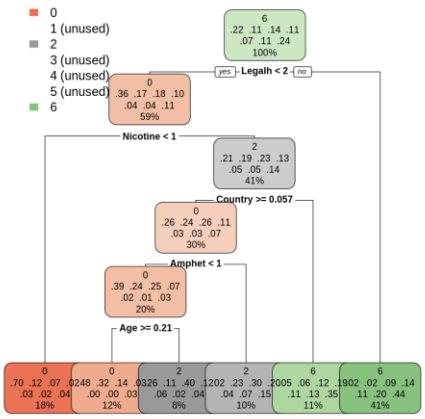
Our group uses stepwise feature selection to choose some of the important variables. Specifically, we use the Stepclass function in the klar package rather than the step() function that was taught in class, because Stepclass function is designed for classification purposes, but step function is based on linear regression. Stepclass can use models, iteratively add/drop features by stepwise feature selection selection based on model cross validation accuracy score with 10 folds. Our group uses the default setting of the model, selects accuracy score as the evaluation metrics (cross validation accuracy score with 10 folds). Finally, we selected the features by stepwise feature selection: Cannabis ~ Age + Gender + Education + Country + Ethnicity + Escore + Oscore + Ascore + SS + Alcohol + Amphet + Amyl + Benzos + Choc + Ecstasy + Heroin + Ketamine + Legalh + LSD + Meth + Mushrooms + Nicotine.

Generally, a decision tree makes plenty of decisions based on a set of features shown in the data. We build a model between the factor of Cannabis and all the other variables from the train dataset via rpart function.

```
[ ]  model1 <- rpart(Cannabis ~ . , data_train, method = "class")
```

We have the graph below,

```
[ ]  rpart.plot(model1, tweak = 1.2)
```



In this graph, yes is to the left and no points to the right. Each branch is a decision for splitting the data into a new classification. From Fig. 01, we can observe that the decision tree splits the data into only 3 of the 6 available classification, which are 0, 2, 6 respectively.

Fig. 01

```
rpart.plot(plot1, type = 4, extra = 101, tweak = 1.2)
```

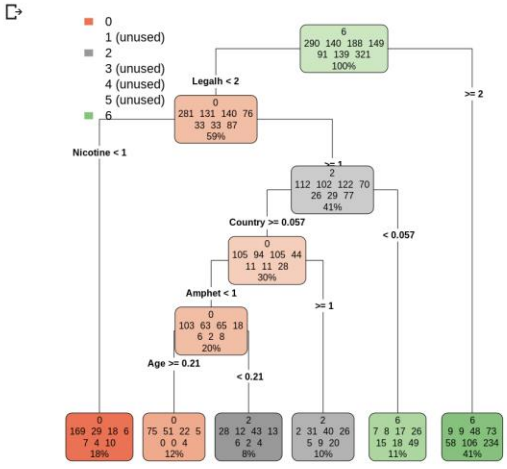


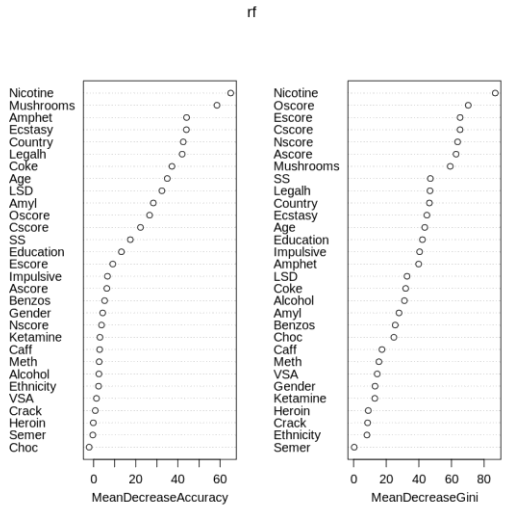
Fig. 02 has more details. It shows the furthest branches where this prediction makes a large amount of errors. We test its prediction and accuracy later in this report in the module of model prediction.

Fig. 02

Also, through the code below, we find the top several important variables using rpart function in the decision tree of feature selection.

```
model_A <- rpart(factor(Cannabis) ~ . , data = train, method = "class")
varImp(model_A)
```

Random forest classifier realizes its classification by recursively following the principle of the majority vote before reaching a determined depth (García-Dominguez et al., 2020). We set the number of trees to be 1000, which means that the randomForest function builds 1000 decision trees, and applies the random forest algorithm to these trees. varImpPlot function returns the variable importance plots as the following:



The left dot chart shows if a variable is removed, the model accuracy will decrease by certain values. For instance, the accuracy is expected to decrease by 50 after removing Nicotine. The right plot represents if a variable is removed, the Gini score will decrease by certain numbers. The higher the value of mean decrease accuracy or mean decrease Gini score, the more important the variable is. Thus, we can choose the variable with higher scores on both graphs.

In the last step of the feature selection, we combine the predictors that are in stepwise feature selection, the result of the rpart algorithm, and the result of random forest algorithm. The final model is: Cannabis ~ Age + Gender + Education + Country + Ethnicity + Escore + Oscore + Ascore + SS + Alcohol + Amphet + Amyl + Benzos + Choc + Ecstasy + Heroin + Ketamine + Legalh + LSD + Meth + Mushrooms + Nicotine + Alcohol:Legalh + Caff:Legalh + Choc:Legalh + Country:Alcohol + Country:Caff + Country:Choc + Age + Nscore + Cscore + Coke + Nicotine. Some feature like Alcohol:Legalh is the feature interaction that created in rpart.

Model Prediction

According to the above discovery, we build machine learning models, based on 75% training data and 25% testing data. We choose to use Stratified folds (by R splitTools library) because each fold contains roughly the same proportions of all types of class labels. The performance of the following models are measured by the accuracy, where accuracy is defined by Accuracy = number of correct predictions / total number of predictions.

Our group has both results for predicting 7 classes and 5 classes. We are discussing the models for 5 levels for the following part, and will show both predictions' results in the conclusion. For 5-classes prediction we decide to put CL0, CL1 and CL2 into one category that shows this person who never used this drug or not used this drug for at least one decade, since 7 classes may result in low accuracy of our model predictions. Further reason for discussing 7 classes and 5 classes prediction will be in the result part.

Neural Network:

Our group uses a simple neural network, feed-forward neural network with a single hidden layer by caret and kknn package. A feedforward neural network is an artificial neural network in which connections between the nodes do not form a cycle. We choose to use the neural network because the neural network uses the hidden layer to make the prediction more accurate. For our classification purpose, neural networks are all inclusive useful approximators in that neural organizations can inexact any capacity with subjective precision. (Zhang, 2000)

By hyperparameter pruning on decay and size with 10 folds cross validation (decay is parameter for weight decay, size is number of units in the hidden layer), we get our final model from the caret package: size = 3, decay = 0.1. The accuracy is 0.636 by calculating from the accuracy table.

Support Vector Machine (SVM):

In addition, we also use the Support Vector Machine(SVM) method for model prediction. SVM is a supervised learning method for classification and regression prediction in order to maximize the accuracy of a model and also automatically avoid overfitting the data. However, since example data is often not linearly separable, SVM introduces the notion of a "kernel induced feature space", casting the data into a higher dimensional space where the data is separable (Boswell, 2002), and we want the hyperplane to have the largest geometric distance to the closet data points.

In programming, we library "e1071" package using the svm function to apply the training and testing data to the model, and the accuracy for the testing data is 0.651.

Boosting:

In gradient boosting, learners are successively trained via the weak classification algorithms, generating a sequence of weak classifiers. In each iteration, the learning rate is modified, and the residual is updated. The final prediction is produced by combining all the predictions.

Compared with other boosting models, the gradient boosted model generates a procedure with higher accuracy and efficiency (Hastie et al., 2009). Caret package is used to find the tuning parameters, and employ 10 folds cross-validation for future performance estimation. We randomly select interaction.depth = c(2,3,4,5,6), n.trees = 600, shrinkage = c(0.1, 0.05, 0.01). The best tuning parameters for data without street address is interaction.depth = 4, n.trees = 500, shrinkage = 0.01. The testing accuracy is 0.6255.

XGboost:

XGBoost is an implementation of gradient boosted decision trees designed for speed and performance. Tree boosting is a highly effective and widely used machine learning method. Other than normal boosting methods, XGboost incorporates a novel tree learning calculation for dealing with sparse data. XGboost is good for our classification problem because XGboost is better than R's GBM. The reason is that GBM only expands one branch of a tree by greedy approach, but XGboost can learn the full tree (Chen & Carlos 2016).

For XGB boosting, we choose the following parameters to get the optimal choice: learning rate, max depth, number of estimators, gamma(Minimum loss reduction required to make a further partition on a leaf node of the tree), subsample (Subsample ratio of the training instances.), and other parameters. By hyperparameter pruning with 10 folds cross validation on these parameters, the final model is: learning rate = 0.01, max depth = 4, number of estimators = 1000, gamma = 0.85, subsample = 6. The accuracy is 0.653 by calculating from the accuracy table.

KNN:

KNN is a supervised machine learning algorithm that can be used to solve classification and regression problems. We use the KNN method to test the accuracy of prediction of the data. One advantage of KNN is that it is a non-parametric model which means it does not make any assumptions about the data set, making the algorithm simpler and effective due to handling realistic data (Srinivas Gurrala 1970). However, one of the disadvantages of KNN is that it does not predict the result with high accuracy. If the dataset is large because of the huge cost of calculating the distance between the new point and each existing point, it degrades the performance of the algorithm (Srinivas Gurrala 1970).

Our first step is choosing the distance calculating function. We normally use Euclidean Metric to calculate the distance between each point, then we use the package caret and create testing data and training data to find out the best K value is 17 by the trainControl and train function in caret. Lastly, normalizing the data with the min-max standardization and use the knn function to see the result table, after calculating we can get the accuracy is 0.64.

Random Forest:

Random forest is a random subset getting the predictor variables. The generalization error of a forest of tree classifiers depends on the strength of the discrete trees in the forest and the correlation between them (Greenwell, 2020). It is a highly effective algorithm based on pattern combination thoughts for

classification. The random forest algorithm can aggregate the predictions, varying depth of multiple decision trees (Kuhn, 2019).

After setting a portion of our data aside for testing, we initialize an instance of the randomForest class. The model will automatically attempt to classify each of the samples in the Out-Of-Bag dataset and show a confusion matrix with the results. We use our model to predict whether the people in our testing set are Cannabis users.

Since this is a classification problem, we use a confusion matrix to evaluate the performance of our model. Specifically, we use randomForest function to observe the relationship between factor(Cannabis) and the rest of the variables we selected in the feature selection from our training data set before obtaining OOB estimate of error rate and confusion matrix, where OOB dataset means the portion of samples are left out during the construction of each decision tree in the forests (Greenwell, 2020). The model automatically evaluates its own performance by running each of the samples in the OOB dataset through the forest.

Also, we use the trainControl function and expand.grid function within the caret function to choose the tuning parameter -mtry, and then specify the grid for mtry. The testing accuracy is 0.636 and the selected key parameter “mtry” = 6.

Decision Tree

Furthermore, we use the C50 algorithm to perform a decision tree to see the most related variables with Cannabis. C50 firstly grows a large tree that overfits the drug consumption training data, and then, it will remove some branches without influencing the final classification. We choose this method because of several benefits. Firstly, this method is a highly automatic learning process, which can handle numeric or nominal features, as well as missing data. Also, it can ignore some unimportant features. Secondly, This way can not only deal with large datasets, but small ones. It is a quite efficient and easy-understanding

method for every individual (Patel & Prajapati, 2018). In our dataset, we try to draw a tree diagram, but we find it is quite hard to visualize the tree due to the large number of observations, which is also a weakness of decision trees. Thus, we classificate the training data, then use it to evaluate the performance of testing data. As a result, the accuracy of the testing data, which is approximately 0.5596.

Results

	Accuracy for 5 levels	Accuracy 7 levels
Neural network	0.636	0.508
KNN	0.640	0.457
Decision Tree	0.557	0.367
Random Forest	0.636	0.499
SVM	0.651	0.487
Boosting	0.632	0.501
XGboost	0.653	0.510

Fig.1. Prediction result table

Our results for the above discussion are in this table. We find that XGboost and SVM have the best accuracy for five levels. XGBoost works well because it has parallel tree building, and regulation through L1 and L2 metrics to avoid overfitting. SVM works well since SVM does well on classifications, and works well in high dimension space. We find the huge difference between five level and seven level's prediction, then we compare the accuracy table between five levels and seven levels by using boosting as an example as below,

Boosting : (5 levels : 0.6255319)

	X2	X3	X4	X5	X6
X2	205	18	6	9	20
X3	5	7	2	3	7
X4	1	0	2	1	3
X5	2	3	4	3	7
X6	8	25	21	31	77

(7 levels: 0.5010616)

	X0	X1	X2	X3	X4	X5	X6
X0	85	25	14	4	2	2	5
X1	10	17	5	2	0	1	3
X2	6	7	35	10	3	4	7
X3	0	0	5	4	3	2	5
X4	0	0	0	0	1	0	0
X5	0	0	0	1	1	1	2
X6	2	2	8	32	25	37	93

Fig.2. Accuracy table

According to the accuracy table of Boosting, we find that the accuracy to predict label 2 is higher than other labels. This is because label 2 has more records than other labels. We merge the data at the beginning, and merge label 0,1,2 to label 2. By merging some levels, the models are still the same models, but it can help our group to solve problems with not enough data in one of the levels.

Conclusion and Discussion

For the real world application, our group can use our dataset to solve different problems. When we want to know who had Cannabis before within any time, we can make our prediction to 2 levels. The first level is CL0, people never have drugs, the second level is merging CL1 to CL6, people had drugs before. Our group's result has 5 levels, which has more specific details for people who have Cannabis in the past week, month, year. We can use the model to discover different levels of people who had Cannabis within one year, predicted by other relevant and irrelevant data.

From our results above, we get the best result by use data exploration to get the patterns of data by statistics and data visualization, feature selection to support important features by tree methods and stepwise feature selection, and using different machine learning models and apply hyper pruning parameters based on 10-folds's cross validation score. The best model is XGboost and the accuracy is 0.653. The accuracy will be higher if we use less labels and more data.

Our dataset has some limitations, since there is not enough data for some levels. This is because our prediction has 5 levels, but there are only about 1800 data points (and we split the train/test dataset). Though we satisfy the project's requirements, we may need more data points to get a more accurate score since it is enough for 2-level's prediction but not for 5-levels.

There are some methods that we can still try for the further improvements:

1. Generate more data by the researcher
2. Stacking Classifier: Model stacking is an efficient ensemble method in which the predictions, generated by using various machine learning algorithms, are used as inputs in a second-layer learning algorithm. This second-layer algorithm is trained to optimally combine the model predictions to form a new set of predictions. Stacking (Wolpert, 1992) is frequently used to gain proficiency with a joining strategy besides the ensemble of classifiers. Normally stacking is better than voting (Todorovski & Dˇzeroski, 2002).
3. Advanced Deep learning model: Our previous deep learning model only has a single dense layer. But there is a much more advanced setting for the deep learning models. We can put more layers inside the deep learning models, such as max pooling layer, convolutional layer. As an example of the convolutional layer, CNN is intended to consequently and adaptively learn spatial chains of importance of features through backpropagation by utilizing numerous structure blocks, for example, convolution layers, pooling layers, and fully connected layers (Yamashita, 2018).
4. Dimensional Reduction

   Our group may also use dimensional reduction methods, such as principal component analysis. Principal component analysis (PCA) is a procedure to decrease the dimensionality of such datasets, expand interpretability and yet limit information loss. It does so by making new uncorrelated factors that maximize variance (Jolliffe & Cadima, 2016). By using PCA, our group can use 50% - 70% of data to explain 90% of explained variance ratio for the original dataset. It can reduce the computational cost for our model, but may have information loss.

References

Brown, R., Leonard, T., Saunders, L., & Papasouliotis, O. (1998). The Prevalence and Detection of Substance Use Disorders among Inpatients Ages 18 to 49: An Opportunity for Prevention. Preventive Medicine, 27(1), 101–110. https://doi.org/10.1006/pmed.1997.0250

Caldeira, K., Kasperski, S., Sharma, E., Vincent, K., O'Grady, K., Wish, E., & Arria, A. (2009). College students rarely seek help despite serious substance use problems. Journal of Substance Abuse Treatment, 37(4), 368–378. https://doi.org/10.1016/j.jsat.2009.04.005

Ginoux, J., Naeck, R., Ruhomally, Y., Dauhoo, M., & Perc, M. (2019). Chaos in a predator–prey-based mathematical model for illicit drug consumption. Applied Mathematics and Computation, 347, 502–513. https://doi.org/10.1016/j.amc.2018.10.089

Overview of Data Exploration Techniques. (2015). SIGMOD '15: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, 277–281.

https://doi.org/10.1145/2723372.2731084

Shoal, G. D., & Giancola, P. R. (2001). Cognition, negative affectivity and substance use in adolescent boys with and without a family history of a substance use disorder. Journal of Studies on Alcohol, 62(5), 675–686. https://doi.org/10.15288/jsa.2001.62.675

Zhang, Peter. (2000). Neural Networks for Classification: A Survey. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on. 30. 451 - 462. 10.1109/5326.897072.

Chen, Tianqi, and Carlos Guestrin. (2016). XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Boswell, D. (n.d.). Introduction to Support Vector Machines. http://www.work.caltech.edu/~boswell/IntroToSVM.pdf.

Wolpert, D. (1992). Stacked generalization. Neural Networks, 5:2, 241–260

Todorovski, L., & Dˇzeroski, S. (2000). Combining multiple models with meta decision trees. In Proceedings of the Fourth European Conference on Principles of Data Mining and Knowledge Discovery (pp. 54–64). Berlin, Springer.

Yamashita, R., Nishio, M., Do, R.K.G. (2018). Convolutional neural networks: an overview and application in radiology. Insights Imaging, 611–629. /https://doi.org/10.1007/s13244-018-0639-9

Jolliffe Ian T. and Cadima Jorge (2016), Principal component analysis: a review and recent developmentsPhil. Trans. R. Soc. A.37420150202

Group, E. (1970). Understanding the concept of knn algorithm using r: Excelr. Retrieved April 20, 2021, from https://www.excelr.com/blog/data-science/machine-learning-supervised/understanding-the-concept-of-knn-algorithm-using-r

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning, second edition: data mining, inference, and prediction. Springer.

García-Dominguez, A., Galván-Tejada, C. E., Zanella-Calzada, L. A., Gamboa-Rosales, H.,

Galván-Tejada, J. I., Celaya-Padilla, J. M., … Magallanes-Quintanar, R. (2020, February 13). Feature Selection Using Genetic Algorithms for the Generation of a Recognition and Classification of Children Activities Model Using Environmental Sound. Mobile Information Systems. https://www.hindawi.com/journals/misy/2020/8617430/.

Patel, H. H., & Prajapati, P. (2018). Study and Analysis of Decision Tree Based Classification Algorithms. *International Journal of Computer Sciences and Engineering*, *6*(10), 74–78. https://doi.org/10.26438/ijcse/v6i10.7478

Comparative analysis of Filter-Wrapper approach for random Forest performance on multivariate data. (n.d.). Retrieved April 19, 2021, from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6965035

Greenwell, B. (2020, February 01). Hands-on machine learning with r. Retrieved April 20, 2021, from https://bradleyboehmke.github.io/HOML/random-forest.html

Kuhn, M. (2019, March 27). The caret package. Retrieved April 11, 2021, from https://topepo.github.io/caret/feature-selection-overview.html