Using Multiclass Classification Algorithms to Model Marijuana Legality

Ellemieke Van Kints
University of Georgia
Athens, Georgia
ejv88036@uga.edu

Noah Sivinski
University of Georgia
Athens, Georgia
ncs34188@uga.edu

Lilly Sharples
University of Georgia
Athens, Georgia
lillysharples@uga.edu

***Abstract*—The objective of this research is to determine if there is a correlation between the legalization of marijuana and a multitude of socioeconomic factors, such as happiness index, crime rate, GDP, etc., both domestically and internationally, in order to predict the legality of marijuana in a given state or country. Over the last several years, an increasing number of states and countries have voted to legalize marijuana medicinally as well as recreationally. We are curious to see how the decriminalization or legalization of drugs, or lack thereof, impacts the people and economy in that area. Our focus was on specifically determining legality using socioeconomic factors found in data sets from 2020 and 2021 for states, and 2021 for countries.**

To predict the legality of marijuana in a given state or country, we trained multiple machine learning models - Gaussian Naive Bayes, Multinomial Naive Bayes, K-Nearest Neighbor, and Decision Tree algorithms. Accuracy scores ranged from 0.400 for Multinomial Naive Bayes trained on country data, to 0.633 for Decision Tree algorithms trained on state data. The Decision Tree models resulted in the highest overall accuracy, and classifiers trained on country data outperforms those trained on state data. Our highest accuracy was 0.633, therefore we concluded that our machine learning models were not able to accurately predict the legal status of marijuana.

***Index Terms*—marijuana, marijuana legalization, machine learning, multiclass classification**

I. Introduction

Marijuana is the most commonly used illicit drug in the United States, with 3.1 million individuals reporting daily use in the last year and 8.1 million individuals reporting using marijuana most days in the last month in 2013 [17]. Domestically, the legalization of marijuana has been a controversial topic for decades. However, the trend toward legalizing marijuana has been driven largely by a shift in public opinion in favor of legalization. In 1996, California became the first state to legalize medicinal marijuana, sparking a trend that led to a multitude of states by 2016. Colorado and Washington became the first states to legalize marijuana recreationally in 2012 [7]. Meanwhile, internationally, the commercial sale of recreational marijuana is legalized in only two countries, Canada and Uruguay. A policy of limited enforcement has also been adopted in a select number of countries. For example, in the Netherlands, the sale of commercial marijuana is tolerated only at licensed coffee shops [26]. Finally, a select number of countries have legalized the medicinal use of marijuana.

The objective of our research project was to predict the legality of marijuana using a variety of socioeconomic factors. By determining the accuracy of our models, we will be able to determine how the legal status of marijuana (fully legal, mixed/medicinal, fully illegal) affects, or does not affect, a given state or country. In order to properly assess the impacts of marijuana in a variety of communities and cultures, we gathered data sets for US states as well as a multitude of countries around the world.

After collecting our data sets, we first merged the associated marijuana data with the socioeconomic data using `pandas`. After organizing and compiling our data, we normalized it using the Box-Cox Normalization algorithm from the `scipy` library. These steps were essential for the Gaussian Naive Bayes and K-Nearest Neighbor Models to train our data. The Shapiro-Wilks test was then used in order to ensure the normality condition was met. After normalizing our data, we split the `x` and `y` data into random train and test groups using the `sklearn` library. We then had to use the Synthetic Minority Oversampling Technique (SMOTE), to balance our `y` sets. After preprocessing our data, we began training our models using the Gaussian Naive Bayes, Multinomial Naive Bayes, K-Nearest Neighbor and Decision Tree classifiers from the `sklearn` library.

II. Related Work

While there is not much research in regards to predicting the legality of marijuana using socioeconomic factors, there is substantial research into the socioeconomic impacts of marijuana on a community, from a variety of viewpoints and on a variety of factors.

Marijuana and the legalization of marijuana affects not only governmental legislation, regulation, and the economy, but also has a direct impact on the individuals within those communities where marijuana has been legalized or is present. For instance, an article researching marijuana's impact on aggression and delinquent behavior in high school students found that cannabis use was a key factor in discriminating youths' membership to distinct patterns of conduct problems [9]. Cannabis use in adolescence and early adulthood is associated with poor social outcomes, including unemployment, lower income, and lower levels of life and relationship satisfaction [6].

Marijuana has a profound impact on the overall health of a community. In areas where medical marijuana was legalized, the prevalence of serious mental illnesses—like schizophrenia and bipolar disorder—were significantly higher following legalization compared to the period before legalization [24]. With recent changes in its legal status, the impact of marijuana on driving ability is increasingly relevant, as marijuana is the most common illicit drug reported in motor vehicle accidents [6].

The legalization of marijuana not only directly affects a multitude of socioeconomic factors, but indirectly affects them as well. Following medical marijuana legalization in several U.S. states, there were significant increases in cardiac mortality rates, but there were concurrent reductions in the rates of opioid prescribing, particularly in areas where cannabis dispensaries were legal. However, in these states, there was a concurrent increase in tobacco sales [24]. The legalization of marijuana also affects the frequency with which people use/abuse other drugs, leading to a slew of additional negative socioeconomic impacts. With this information in mind, it is easy to see why the legalization of marijuana has been a long process, that is still going on worldwide.

Domestically and internationally, the legalization of marijuana is a heavily debated topic. This is due to the numerous pros and cons of the drug. While marijuana is predominantly known for its negative effects, as detailed above, it can also have positive effects, ranging from physical health to mental health to boosting economic growth.

Economically, marijuana has a ton of upside, from estimated savings from reduced spending on the criminal justice costs of marijuana law enforcement and revenue losses from shifts in law enforcement policies, to projected revenues from additional taxes and streams of income [26]. For example, a study on the effect of marijuana dispensary openings on housing prices, determined that the introduction of a new dispensary within a half-mile radius of a new home increases home prices by approximately 7.7 percent on average [8]. Although there isn't much research into how exactly the legalization of marijuana will impact an economy long-term, as it hasn't been legal anywhere recreationally for more than a decade, the short-term benefits are clear and have been proven.

The positive impact of marijuana on the health and mental health of an individual should not be overlooked. A recent meta-analysis of nine studies concluded the causal association of marijuana use with lower BMI and obesity rates. In addition, several studies have shown a lower prevalence of obesity and BMI among young marijuana users [25]. In terms of mental health, an article detailing marijuana's effect on suicide rates found that California's 1996 legalization resulted in statistically significant ($p<.05$) reductions in suicides and gun suicides, but only a non-significant reduction in non-gun suicides ($p>=.488$). Since the effect for non-gun suicides was indistinguishable from chance, we infer that the overall causal effect was realized through gun suicides. The mechanism could not be determined, however. Participation in the medical marijuana program legally disqualifies participants from purchasing guns. But since most suicides involve guns, it is possible that the effect on total suicide is driven by gun suicide alone [6].

As the aforementioned articles have shown, marijuana can have a significant impact on the socioeconomic status of a state or country. We intend to use this research along with our own to determine if we can accurately predict the legality of marijuana in a given state or country using a set of socioeconomic factors.

III. Software System

A. Software Overview

The goal of our software is to draw correlations and accurately predict the legality of marijuana in a state or country given a discrete set of socioeconomic factors. There are numerous variables which contribute to the socioeconomic health of a state or nation. We identified the variables most relevant to marijuana legalization and gathered data from 50 states and 195 countries. Our software compiles and preprocesses marijuana legality data and socioeconomic data into compatible forms. Then, the data is trained using the Gaussian Naive Bayes, Multinomial Naive Bayes, K-Nearest Neighbour, and Decision Tree algorithms. The trained models are fed unseen data and attempt to perform multi-class classification.

The programming language `Python` [5], and frameworks `SciKit-Learn` [2], `Pandas` [3], `SciPy` [4] were used to create our software system.

B. Data Compilation

The data we gathered is organized into two main categories: data by state and data by country. Each data set is independent and treated as its own classification problem. The state data set contains 2020 and 2021 data, whereas the country data set contains only 2021 data. We include state data from both 2020 and 2021 to ensure the data set is large enough to split into training and testing sets.

After the data is read in, we organize it into its associated *State Data* and *Country Data* sets. We utilize the `merge()` method from `pandas`, a Python Data Analysis Library [2], to merge the marijuana data with the socioeconomic data.

C. Data Preprocessing

In order to make our data sets compatible with the machine learning models we plan on training, the categorical legality data must be encoded to numerical form. In addition, our data frame contained numerical values in `string` form, requiring us to apply the `to_ numeric()` method from `pandas` to the entirety of our data.

Our software classifies marijuana legality into three groups. For both state and country data, our classes are *Fully Illegal*, *Mixed*, and *Fully Legal*, with *Mixed* indicating marijuana is legal for medicinal use or it is decriminalized but not fully legalized. The classes are encoded to 0, 1, and 2, respectively, using a dictionary and the `replace()` method.

The target variables for state and country are represented as marijuana legality data sets. Similarly, the socioeconomic columns from the state and country data sets are compiled into a new data set representing our independent variable.

```python
encode = {'LegalStatus': {'FullyIllegal':0, 'Mixed':1, 'FullyLegal':2},
          'Medicinal': {'Yes':1, 'No':0},
          'Decriminalized': {'Yes':1, 'No':0}}
data = data.replace(encode)
```

Fig. 1. Our code which transforms the categorical legality data to numerical form.

The independent and dependent variables are set to `x` and `y` respectively. From here, we create two functions called `normalizeXS(col)` and `normalizeXC(col)` which normalize the specified column in a data set using the Box-cox Normalization algorithm from the `scipy` library. The `normalizeXS(col)` function is used to normalize the state socioeconomic data, while the `normalizeXC(col)` function is used to normalize the country socioeconomic data.

This standardization technique is necessary for the Gaussian Naive Bayes and K Nearest Neighbor Models, as they assume all features are centered around zero and have variance in the same order. To ensure the normality condition is met, our functions perform the Shapiro-Wilks test for normality on each column in the associated data set. This is done using the `stats` module in the `scipy` library, and "tests the null hypothesis that the data was drawn from a normal distribution" [11]. The test statistic and p-value from the Shapiro-Wilks test before and after the BoxCox normalization is applied are printed when the `normalizeXS(col)` and `normalizeXS(col)` functions are called. A p-value greater than or equal to 0.05 indicates that we can assume the socioeconomic factor follows a normal distribution. Our output at this stage is shown in *Figure 2* and *3*.

Once the data is normalized, our software splits the `x` and `y` data into random train and test groups using the `sklearn` library. However, the `y` sets are imbalanced. In order to develop accurate models, we must balance them using the Synthetic Minority Oversampling Technique, or SMOTE for short. SMOTE works to balance the distribution of a data set by synthesizing new examples from the minority class. This technique randomly selects a member of the minority class, finds a few of the nearest neighbours, and a "synthetic example is created at a randomly selected point between the two" [1]. The distributions of marijuana legality before and after oversampling can be seen in *Figures 4-7*. We utilized the `collections` and `matplotlib` libraries to create a `distribution(y)` function which takes in a data set and prints the class labels, their sample size, distributions, and the corresponding bar plot.

D. Model Training

We begin model training using the Gaussian Naive Bayes, Multinomial Naive Bayes, K-Nearest Neighbour, and Decision Tree classifiers from the `sklearn` library. The goal is to train and test multiple different models on the same data and determine which performs best. Our software fits each classifier with the over sampled `x` and `y` training data. Then, the trained model is fed the testing data and returns the predicted class. To evaluate the performance of our

```
Happiness_Rank
SHAPRIO-WILKS TEST SCORE (statistic, p-value) BEFORE applying Boxcox Normalization:
(0.9518184661865234, 0.002170937368646264)
SHAPIRO-WILKS TEST SCORE (statistic, p-value) after applying Boxcox Normalization:
(0.9532076120376587, 0.002654424635693431)

Score
SHAPRIO-WILKS TEST SCORE (statistic, p-value) BEFORE applying Boxcox Normalization:
(0.9843503832817078, 0.35494503378868103)
SHAPIRO-WILKS TEST SCORE (statistic, p-value) after applying Boxcox Normalization:
(0.9842210412025452, 0.34821030497550964)

Emotional_Physical_Wellbeing
SHAPRIO-WILKS TEST SCORE (statistic, p-value) BEFORE applying Boxcox Normalization:
(0.9481477737426758, 0.0012885953765362501)
SHAPIRO-WILKS TEST SCORE (statistic, p-value) after applying Boxcox Normalization:
(0.9512153267860413, 0.0019907099194824696)

Work_environment
SHAPRIO-WILKS TEST SCORE (statistic, p-value) BEFORE applying Boxcox Normalization:
(0.9542649984359741, 0.003097675507888198)
SHAPIRO-WILKS TEST SCORE (statistic, p-value) after applying Boxcox Normalization:
(0.956375777721405, 0.004231169354170561)

community_environment
SHAPRIO-WILKS TEST SCORE (statistic, p-value) BEFORE applying Boxcox Normalization:
(0.9563535451889038, 0.004217214882373381)
SHAPIRO-WILKS TEST SCORE (statistic, p-value) after applying Boxcox Normalization:
(0.9591416120529175, 0.006412710528820753)

homicideRate
SHAPRIO-WILKS TEST SCORE (statistic, p-value) BEFORE applying Boxcox Normalization:
(0.9406560063362122, 0.0004635351651813835)
SHAPIRO-WILKS TEST SCORE (statistic, p-value) after applying Boxcox Normalization:
(0.9867695569992065, 0.49934715032577515)

firearmDeaths
SHAPRIO-WILKS TEST SCORE (statistic, p-value) BEFORE applying Boxcox Normalization:
(0.8216538429260254, 4.770627004546668e-09)
SHAPIRO-WILKS TEST SCORE (statistic, p-value) after applying Boxcox Normalization:
(0.979752123358655, 0.17356784641742706)

Work_environment
SHAPRIO-WILKS TEST SCORE (statistic, p-value) BEFORE applying Boxcox Normalization:
(0.9542649984359741, 0.003097675507888198)
SHAPIRO-WILKS TEST SCORE (statistic, p-value) after applying Boxcox Normalization:
(0.956375777721405, 0.004231169354170561)

Work_environment
SHAPRIO-WILKS TEST SCORE (statistic, p-value) BEFORE applying Boxcox Normalization:
(0.9542649984359741, 0.003097675507888198)
SHAPIRO-WILKS TEST SCORE (statistic, p-value) after applying Boxcox Normalization:
(0.956375777721405, 0.004231169354170561)

Work_environment
SHAPRIO-WILKS TEST SCORE (statistic, p-value) BEFORE applying Boxcox Normalization:
(0.9542649984359741, 0.003097675507888198)
SHAPIRO-WILKS TEST SCORE (statistic, p-value) after applying Boxcox Normalization:
(0.956375777721405, 0.004231169354170561)
```

Fig. 2. The code output from the the Shapiro-Wilks test, both before and after Boxcox Normalization, for each state socioeconomic factor.

```
lifeExp
shapiro score (statistic, p-value) before applying boxcox normalization
(0.9500476121902466, 9.447557386010885e-05)
shapiro score (statistic, p-value) after applying boxcox normalization
(0.9697828888893127, 0.004647860769182444)

homicideRate
shapiro score (statistic, p-value) before applying boxcox normalization
(0.5204280614852905, 5.297808514042343e-19)
shapiro score (statistic, p-value) after applying boxcox normalization
(0.9901911020278931, 0.47353529930114746)

happinessScore
shapiro score (statistic, p-value) before applying boxcox normalization
(0.9876304268836975, 0.277960866689682)
shapiro score (statistic, p-value) after applying boxcox normalization
(0.9878937602043152, 0.2944507598876953)

Generosity
shapiro score (statistic, p-value) before applying boxcox normalization
(0.945961594581604, 4.605889625963755e-05)
shapiro score (statistic, p-value) after applying boxcox normalization
(0.992624044418335, 0.717618465423584)

gdpPerCapita
shapiro score (statistic, p-value) before applying boxcox normalization
(0.7026705741882324, 4.53306283568905e-15)
shapiro score (statistic, p-value) after applying boxcox normalization
(0.9751510620117188, 0.015372253954410553)

Freedom_life_choices
shapiro score (statistic, p-value) before applying boxcox normalization
(0.9440122246742249, 3.2994532375596464e-05)
shapiro score (statistic, p-value) after applying boxcox normalization
(0.9799484014511108, 0.04675779864192009)
```

Fig. 3. The code output from the the Shapiro-Wilks test, both before and after Boxcox Normalization, for each country socioeconomic factor.

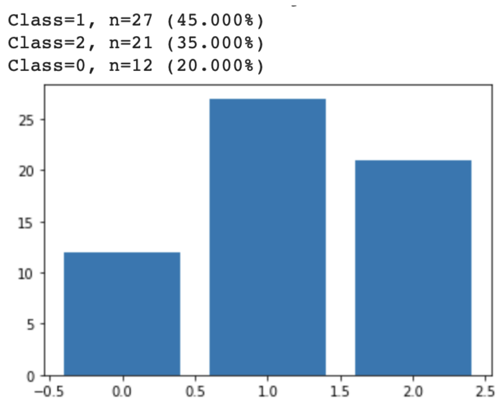

Fig. 4. The distributions of marijuana legality by state before oversampling.

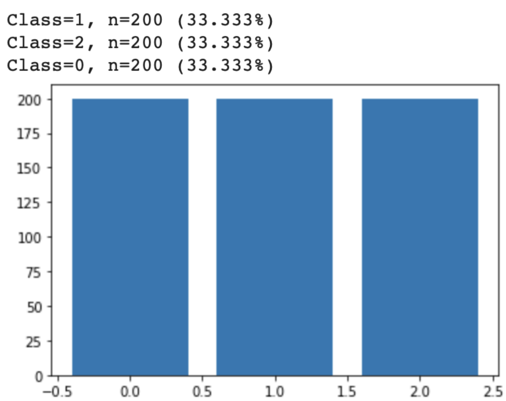

Fig. 5. The distributions of marijuana legality by state after oversampling.

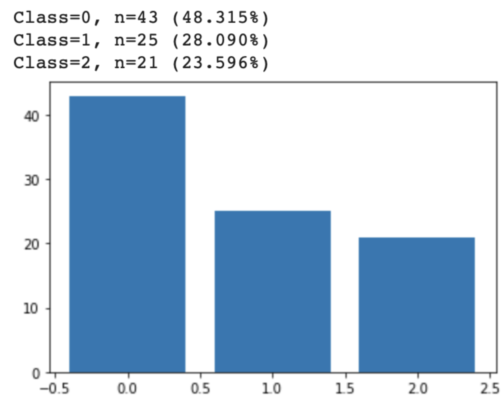

Fig. 6. The distributions of marijuana legality by country before oversampling.

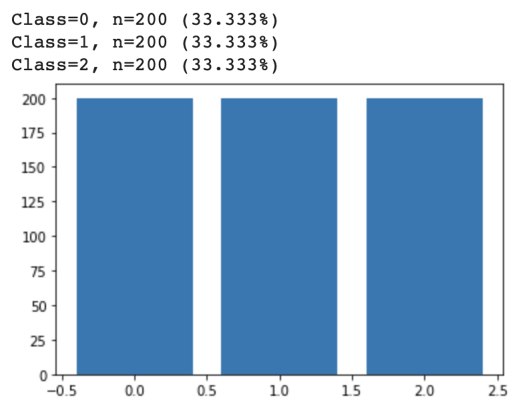

Fig. 7. The distributions of marijuana legality by country after oversampling.

classification models, we created a `plot(classifier, x, y, title, y_pred)` function, which utilizes the `plot_confusion_matrix` module to print a confusion matrix and the `metrics` module to calculate the accuracy for a given model. Both modules are imported from the `sklearn` library. This process is repeated for each algorithm on both the state and country data.

IV. Results

The models we trained produced accuracy scores ranging from 0.400 for Multinomial Naive Bayes trained on country data, to 0.633 for Decision Tree trained on state data. This was below our expectations. Our results show that classifiers trained on country data outperform those trained on state data. The accuracy scores for corresponding state and country models were averaged together. The final scores are 0.551 for Gaussian Naive Bayes, 0.439 for Multinomial Naive Bayes, 0.596 for K-Nearest Neighbors, and 0.600 for Decision Tree. Our data shows that the Decision Tree classifier performed best overall (0.600 accuracy), while Multinomial Naive Bayes performed worst (0.439 accuracy).

Below (*Figures 10-17*) are the plotted confusion matrices and accuracy score for each trained model. The models trained on the state data are shown in *Figures 10-13*, and the models trained on the country data are shown in *Figures 14-17*.

The feature importance scores are included with the Decision Tree output to help us identify which socioeconomic factors were most useful in predicting marijuana legality. Our data in *Figure 13* shows that unemployment rate was the most useful feature in the state Decision Tree model. Likewise, our data in *Figure 17* shows that life expectancy was the most useful feature in the country Decision Tree model. These feature importances were calculated using the `sklearn.tree feature_importances_ method`, where "importance of a feature is computed as the (normalized) total reduction of the criterion brought by that feature" [2].

Additionally, we can graph the whole Decision Tree as seen in *Figure 8* and close up in *Figure 9*. By definition, a Decision Tree is "a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features" [2]. Close up, we can see that the decision tree goes through each value, (starting value is [200,200,200] for our over sampled set of 200 values for each legal status category), decides how important the current feature is (gini importance), and assigns it a class. In *Figure 8*, orange and white refer to class 0 (marijuana is illegal), green refers to class 1 (mixed legal status), and purple refers to class 2 (marijuana is legal).

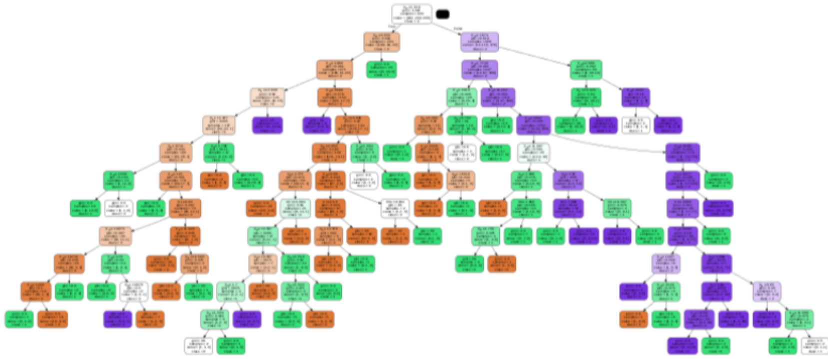

Fig. 8. The Decision Tree from the state data.

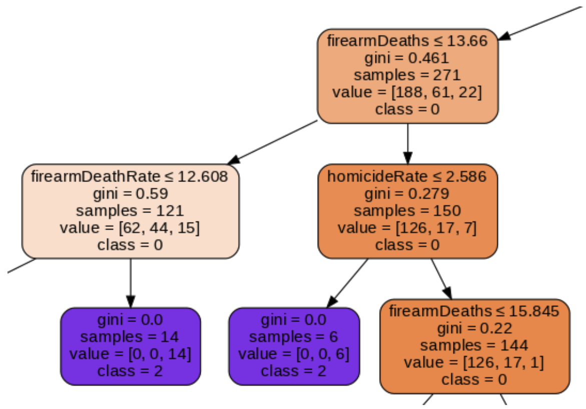

Fig. 9. Close up view of Decision Tree from the state data.

V. Conclusions

These models were created to determine if it is possible to predict the legality of marijuana in a given state or country given a set of socioeconomic factors. As our highest accuracy was 0.63, overall these models were not significantly accurate in predicting whether our selected socioeconomic factors have an impact on predicting the legal status of marijuana in states and countries. This may be due to an overall lack of correlation between socioeconomic factors and legal status, as well as the size of our data set. With only 50 US states and 195 countries with usable data, it is difficult to create an optimal sized data set of thousands of entries. Oversampling to 200 entries per legal status (600 total for our training sets) and the addition of 2020 state data were steps taken to combat the lack of data, but this was still not enough to conclude that machine learning models were able to accurately predict the legal status of marijuana.

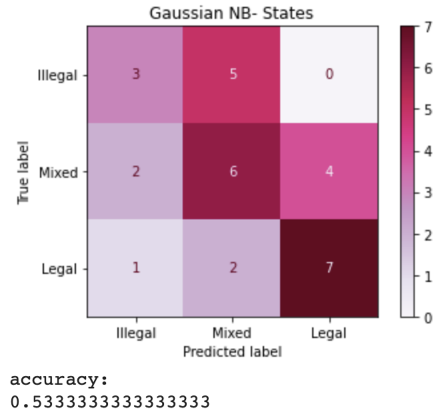

Fig. 10. The confusion matrix and accuracy for the Gaussian Naive Bayes model trained on state data.

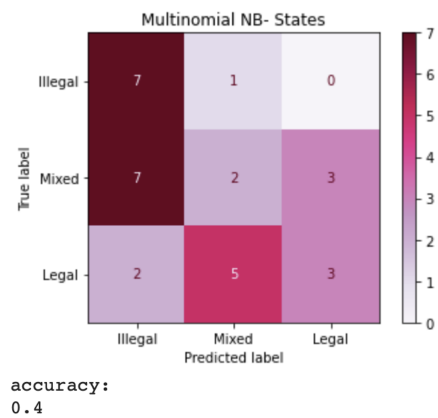

Fig. 11. The confusion matrix and accuracy for the Multinomial Naive Bayes model trained on state data.

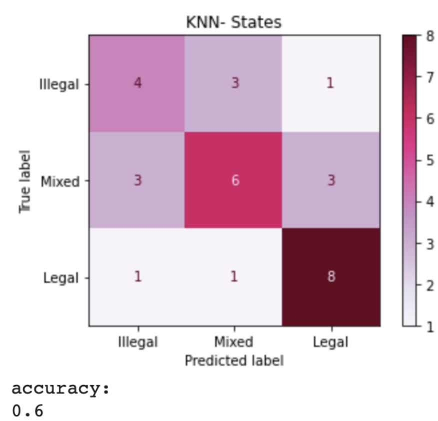

Fig. 12. The confusion matrix and accuracy for the K-Nearest Neighbor model trained on state data.

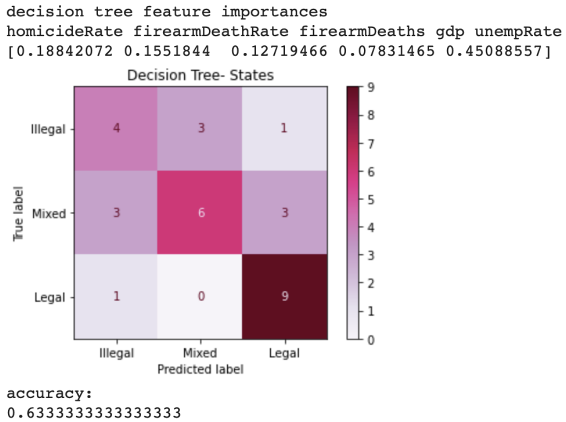

Fig. 13. The confusion matrix and accuracy for the Decision Tree model trained on state data.

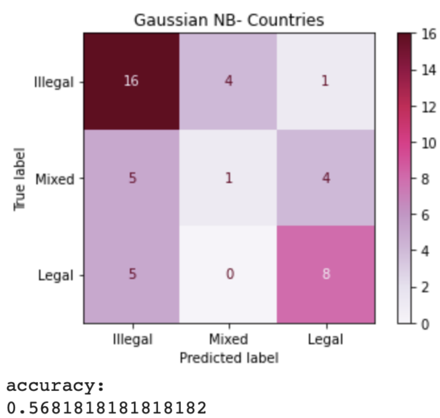

Fig. 14. The confusion matrix and accuracy for the Gaussian Naive Bayes model trained on country data.

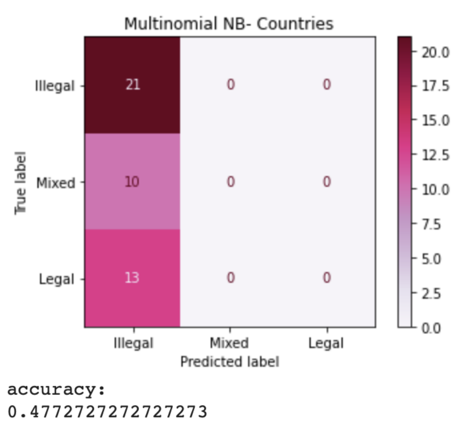

Fig. 15. The confusion matrix and accuracy for the Multinomial Naive Bayes model trained on country data.

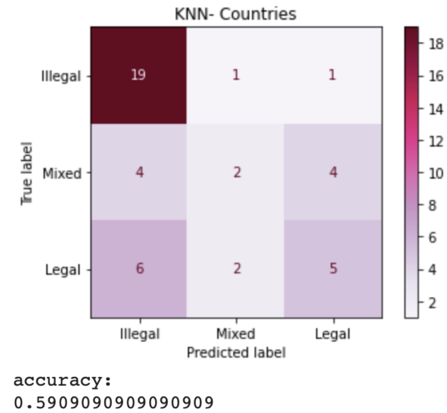

Fig. 16. The confusion matrix and accuracy for the K-Nearest Neighbors model trained on country data.

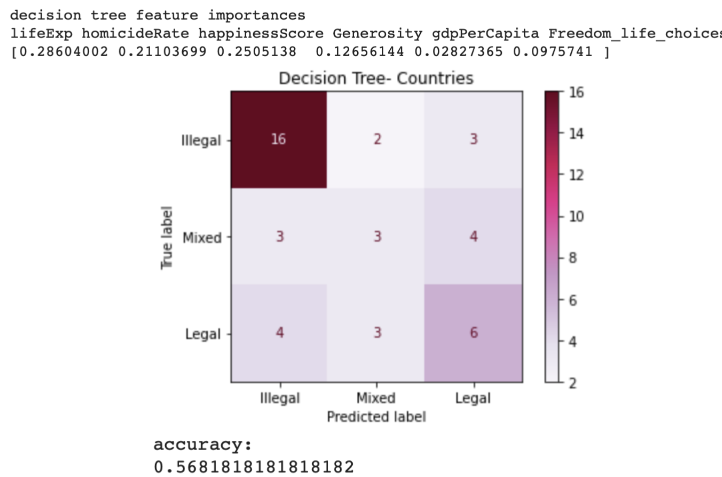

Fig. 17. The confusion matrix and accuracy for the Decision Tree model trained on country data.

References

[1] J. Brownlee. SMOTE for Imbalanced Classification with Python. Jan. 2021. url:https://machinelearningmastery.com/smote- oversampling- for- imbalanced- classification/.

[2] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: Journal of Machine Learning Research 12 (2011), pp. 2825–2830.

[3] The pandas development team, "pandas-dev/pandas: Pandas." Zenodo, Feb-2020.

[4] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, and P. van Mulbregt, "Scipy 1.0: Fundamental algorithms for scientific computing in python," Nature Methods, vol. 17, no. 3, pp. 352–352, 2020.

[5] G. Van Rossum and F. L. Drake, "Python," Python Reference Manual. Centrum voor Wiskunde en Informatica Amsterdam, 1995.

[6] B. J. Bartos, C. E. Kubrin, C. Newark, and R. McCleary, "Medical marijuana laws and suicide," Archives of Suicide Research, vol. 24, no. 2, pp. 204–217, 2019.

[7] D. Hammond, S. Goodman, E. Wadsworth, V. Rynard, C. Boudreau, and W. Hall, "Evaluating the impacts of cannabis legalization: The

International Cannabis Policy Study," International Journal of Drug Policy, vol. 77, Mar. 2020.
[8] J. Burkhardt and M. Flyr, "The effect of marijuana dispensary openings on housing prices," Contemporary Economic Policy, vol. 37, no. 3, pp. 462–475, Nov. 2018.
[9] J. R. Dugré, S. Potvin, L. Dellazizzo, and A. Dumais, "Aggression and delinquent behavior in a large representative sample of high school students: Cannabis use and victimization as key discriminating factors," Psychiatry Research, vol. 296, Dec. 2020.
[10] S. T. Wilkinson, S. Yarnell, R. Radhakrishnan, S. A. Ball, and D. C. D'Souza, "Marijuana legalization: Impact on physicians and public health," Annual Review of Medicine, vol. 67, no. 1, pp. 453–466, Jan. 2016.
[11] Marijuana laws by state 2021, 2021. [Online]. Available: https://worldpopulationreview.com/state-rankings/marijuana-laws-by-state. [Accessed: 02-Dec-2021].
[12] "Regional Data GDP and Personal Income," BEA. [Online]. Available: https://apps.bea.gov/itable/iTable.cfm?ReqID=70amp;step=1amp;acrdn=1. [Accessed: 02-Dec-2021].
[13] A. McCann, "Happiest States in America," WalletHub, 14-Sep-2021. [Online]. Available: https://wallethub.com/edu/happiest-states/6959. [Accessed: 02-Dec-2021].
[14] S. Waterfield, "If you live in this State, you're happier than any other American," Newsweek, 23-Sep-2020. [Online]. Available: https://www.newsweek.com/usa-states-ranked-happiness-hawaii-1533784. [Accessed: 03-Dec-2021].
[15] "State unemployment rate in the U.S. as of October 2021," Statista, 30-Nov-2021. [Online]. Available: https://www.statista.com/statistics/200017/state-unemployment-rate-in-the-us/. [Accessed: 03-Dec-2021].
[16] "Crime rate by state 2021," 2021. [Online]. Available: https://worldpopulationreview.com/state-rankings/crime-rate-by-state. [Accessed: 03-Dec-2021].
[17] J. Helliwell, R. Layard, J. D. Sachs, J.-E. De Neve, L. Aknin, S. Wang, and S. Paculor, Eds., "World happiness report 2021," World Happiness Report 2021, 2021. [Online]. Available: https://worldhappiness.report/ed/2021/appendices-and-data. [Accessed: 03-Dec-2021].
[18] "Violent crime rates by country 2021," 2021. [Online]. Available: https://worldpopulationreview.com/country-rankings/violent-crime-rates-by-country. [Accessed: 08-Dec-2021].
[19] "Legality of Cannabis," Wikipedia, 25-Nov-2021. [Online]. Available: https://en.wikipedia.org/wiki/Legality-of-cannabis. [Accessed: 03-Dec-2021].
[20] "GDP, current prices," International Monetary Fund - Homepage, 2021. [Online]. Available: https://www.imf.org/external/datamapper/NGDPD@WEO/OEMDC/ADVEC/WEOWORLD/AFG/ALB. [Accessed: 08-Dec-2021].
[21] "Unemployment 2021," countryeconomy.com, 2021. [Online]. Available: https://countryeconomy.com/unemployment. [Accessed: 03-Dec-2021].
[22] "Reported violent crime rate in the U.S. by State 2020," Statista, 29-Sep-2021. [Online]. Available: https://www.statista.com/statistics/200445/reported-violent-crime-rate-in-the-us-states/. [Accessed: 03-Dec-2021].
[23] "Impacts of marijuana legalization in Colorado," Jul-2021. [Online]. Available: https://cdpsdocs.state.co.us/ors/docs/reports/2021-SB13-283-Rpt.pdf. [Accessed: 06-Dec-2021].
[24] A. Bahji and C. Stephenson, "International Perspectives on the implications of cannabis legalization: A systematic review amp; thematic analysis," International Journal of Environmental Research and Public Health, vol. 16, no. 17, p. 3095, 2019.
[25] T. Parekh and F. Fahim, "Building risk prediction models for daily use of marijuana using Machine Learning Techniques," Drug and Alcohol Dependence, vol. 225, p. 108789, Aug. 2021.
[26] D. G. Evans, "The economic impacts of marijuana legalization," The Impact of Drug Policies on Our World, 2013. [Online]. Available: https://www.dfaf.org/wp-content/uploads/2018/11/Vol-7-Issue-4.pdf. [Accessed: 07-Dec-2021].