

Predicting Campus Crime Based on State Firearm Policy

Travis Smith, Muhammad Abusaqer

Math, Data, and Technology

Minot State University

Minot, ND, 58707

travis.smith.4@minotstateu.edu, muhammad.abusaqer@minotstateu.edu

Abstract

This paper explores the unique security challenges and threat assessment strategies that a university or college campus faces and methods they can use to protect themselves. The complexity of the environments in which campuses exist stem from many different characteristics, such as their open and accessible nature, diverse population, and a blend of public and private spaces. Delving into the campus environment, we emphasize the role in nurturing developing adults and their inherent difficulties in implementing strict access control measures.

State firearm policy plays a key role in determining how firearms are accessed and used by both crime victims and perpetrators. The RAND firearm policy database provided a list of firearm policies by various states that were labeled with an effect that was as either permissive or restrictive. The firearm policy's were tallied by state and an 'effectValue' was figured by adding 1 if the value was permissive and subtracting 1 if the value was restrictive. This value would label a state in the campus crime database which was retrieved from Campus Safety and Security of the U.S. Department of Education.

The sampled campus crime data provided the crime rates of various schools. The data on each school had various crime numbers over the years 2019, 2020, and 2021. We chose murder for the experiment and this is a planned expansion on the project in the future. The murders were tallied for each school over the three year period and averaged into an 'schoolMurderYearAverage' rate. The goal of the regression model which was trained was to predict this schoolMurderYearAverage for a university or college based on the firearm policy of the state that the campus is in.

Other considerations include an assessment of the U.S. Department of the Treasury's U.S. Secret Service and the adoption of their threat assessment methods in educational settings. This was not largely relevant to the regression experiment, because the regression model is predictive by nature and the work of preparation is different, but it is worth considering. The Exceptional Case Study Project and focuses on key elements of a comprehensive threat assessment program, such as identifying motives, target selection, and attack planning of a potential perpetrator.

The regression model was trained to predict the schoolMurderYearAverage based on the state of the university a campus is located on. This can be improved by expanding on the features of the model for the prediction to be able to have more factors. Some of these could include demographics of the University. You could also expand on the way that the effectValue of a state is figured based on the firearm policies. Such as including public opinion, or some sort of objective measurement for effect more dynamic than simply 'restrictive' or 'permissive'.

In conclusion, this provides a very basic attempt at predicting the crime on a campus based on the states firearm policy, but it is a very primitive model and should not be used for accurate predictions yet. The concepts used in this paper are experimental by nature and also face various ethical concerns about the use of AI in public policy decision making and prediction.

1. Background

Students at Minot State University and the rest of the state campuses under the North Dakota University System are not allowed to conceal or open carry on campuses in the state. This is in large effect because of public opinion on whether or not firearm policy correlates to campus crime.

Policy at many levels plays a part in this: campus policy set by campus administrators, state policy set by state legislatures and governors, and federal policy set by the federal government. The policy makers should make decisions based on data that informs them of their decisions and how they will affect crime.

This paper will research whether state firearm policies correlate to crime on campuses in that state. This will be done using an AI regression model. The data used to train this model will include crimes on campuses as well as firearm policies in the states of the included campuses. This research will focus on North Dakota, South Dakota, Minnesota, Iowa, and Wisconsin for prediction making, but data from other states were used in the research of the paper. Specifically, effectValues were calculated for each states firearm policies.

2. Literature Review

For literature review, I read various places. Specifically the datasets which were used including the RAND database and the Campus Safety and Security of the U.S. Department of Education. We also did research on other methods of threat assessment.

In addition to the data analysis we did of state firarm policies through the RAND Database, this study also considers established practes in threat assessment within educational environemtns. The U.S. Department of Education's guide on threat assessment in schools highlihgt the need for comprehensive strategies to identify and manage potential threats [8]. This approach underscores the imprtance of collaborative efforts involving various stakeholders, including school administration, law enforcement, and mental health proffesionals to effectively mitigate risks and enhance campus sfaety. One under involved demographic in the safety and protection of campuses is students themselves, and students should find various ways to get involved in the safety of their campuses through various means of self defense enablement.

3. Methodology

A variety of methods including software utilization, programming, data correlation and normalizing,

A variety of software was used to do this project. An HP Envy laptop running Windows 11, was used to download and install 'PyCharm 2023.3.4', as well as the Python interpreter and the Jupyter Notebook packages to run the python libraries used for the AI regression model.

3.1 Software

A variety of software was used to do this project. An HP Envy laptop running Windows 11, was used to download and install ‘PyCharm 2023.3.4’, as well as the Python interpreter and the Jupyter Notebook packages to run the python libraries used for the AI regression model. Figure 1 shows a screencapture of Windows 11 opened to JupyterNotebook which was used to train the model.

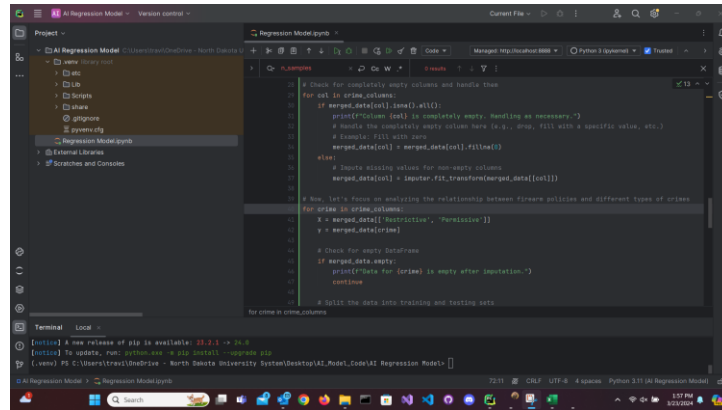


Figure 1: Screenshot of PyCharm with JupyterNotebook.

3.1.1 Programming

Figure 2 shows some of the packages that were included in the python jupyternotebook project. I then read the excel datasets from the hard disk.

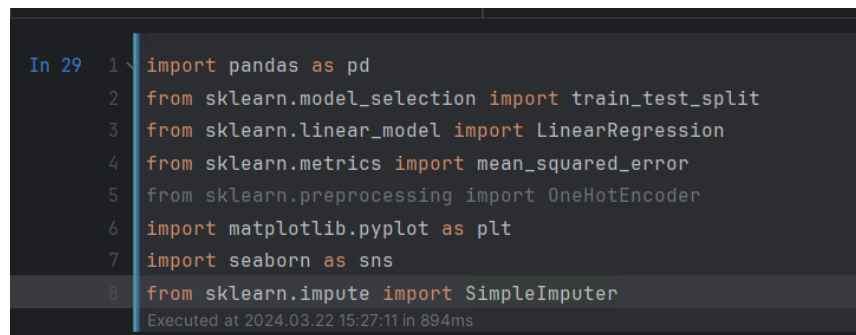


Figure 2: Imported Python packages in jupyter notebook project.

I read from the hard disk the different excel sheets that I needed for the project and printed them out to ensure that they had been loaded in correctly. I utilized ChatGPT 4 [5] for this part of the project, which is great for generating code. Figure 3 shows these packages in the Jupyter Notebook project in pycharm.

```

# Path to the excel file
3 firearm_path = r"C:\Users\travi\OneDrive - North Dakota University
  System\Documents\Academic\Research\Campus Violence Prediction\FirearmPolicySample.xlsx"
4
5 campuscrime_path = r"C:\Users\travi\OneDrive - North Dakota University
  System\Documents\Academic\Research\Campus Violence Prediction\UniversityCrimeSample.xlsx"
6
7 # Read data from the 'Firearm Policy Database' sheet
8 firearm_policy = pd.read_excel(firearm_path, sheet_name='FirearmPolicySample', engine='openpyxl')
9 campuscrime_policy = pd.read_excel(campuscrime_path, sheet_name='UniversityCrimeSample',
  engine='openpyxl')
10
11 # Read data from the 'Institutional-Characteristics22' sheet
12
13 # Display the first few rows from each sheet to verify they've loaded correctly
14 print("Firearm Policy Database:")
15 print(firearm_policy.head())
16
17 #print("\nInstitutional Characteristics 2022:")
18 #print(institutional_chars.head())
19
20 print("\nOn Campus Crime 2019 - 2021:")
21 #print(campus_crime.head())

```

Figure 3: Reading of excel sheets and print statements for testing.

I have a value called 'statesEffectValue' which I calculate by looking at various policies in the state, and their 'effect' value which can either be 'Permissive' or 'Restrictive', and I add one for permissive and I subtract one for restrictive. The setup of this is shown in Figure 4.

```

In 230 1 # Initialize the dictionary for state enabled values
2 statesEffectValue = {'AK': 0, 'AL': 0, 'AR': 0, 'AZ': 0, 'CA': 0, 'CO': 0, 'CT': 0, 'DC': 0, 'DE':
  0, 'FL': 0, 'GA': 0, 'HI': 0, 'ID': 0, 'IL': 0, 'IN': 0, 'KS': 0, 'KY': 0, 'LA': 0, 'MA': 0, 'MD':
  0, 'ME': 0, 'MI': 0, 'MO': 0, 'MS': 0, 'MT': 0, 'NC': 0, 'NE': 0, 'NH': 0, 'NJ': 0, 'NM': 0,
  'NV': 0, 'NY': 0, 'OH': 0, 'OK': 0, 'OR': 0, 'PA': 0, 'RI': 0, 'SC': 0, 'TN': 0, 'TX': 0, 'UT': 0,
  'VA': 0, 'VT': 0, 'WA': 0, 'WV': 0, 'WY': 0, 'WI': 0, 'IA': 0, 'MN': 0, 'ND': 0, 'SD': 0}
3
4 # Iterate over the rows of the firearm policy DataFrame
5 for index, policy in firearm_policy.iterrows():
6     if policy['Effect'] == "Permissive":
7         statesEffectValue[policy['State']] += 1
8     elif policy['Effect'] == "Restrictive":
9         statesEffectValue[policy['State']] -= 1
10
11 print(statesEffectValue)
12 #print(campus_crime)
  Executed at 2024.04.03 13:57:55 in 329ms

```

```

{ 'AK': -2, 'AL': -3, 'AR': -4, 'AZ': -11, 'CA': -63, 'CO': -9, 'CT': -34, 'DC': -32, 'DE':
  -22, 'FL': -22, 'GA': -2, 'HI': -41, 'ID': -1, 'IL': -41, 'IN': -6, 'KS': 0, 'KY': 0, 'LA':
  -2, 'MA': -42, 'MD': -27, 'ME': -3, 'MI': -12, 'MO': -3, 'MS': -1, 'MT': 0, 'NC': -15, 'NE':
  -13, 'NH': -4, 'NJ': -40, 'NM': -9, 'NV': -18, 'NY': -34, 'OH': -11, 'OK': -7, 'OR': -14,
  'PA': -13, 'RI': -25, 'SC': -4, 'TN': -7, 'TX': -4, 'UT': -7, 'VA': -16, 'VT': -8, 'WA': -23,
  'WV': -6, 'WY': -2, 'WI': -13, 'IA': -8, 'MN': -7, 'ND': -4, 'SD': 1}

```

Figure 4: Setup of statesEffectValue variable.

I created a value column in the campus crime dataset called 'schoolMurderYearAverage' which was calculated by totaling the number of murders over the three years 2019, 2020, and 2021 and dividing it by 3 to get the average. This would become the predictive target of the model. This is shown in Figure 5.

```

In 231 1  Years = 3
2
3  # Adding a new column for the average murders
4  campus_crime['schoolMurderYearAverage'] = (campus_crime["MURD19"] + campus_crime["MURD20"] +
5  campus_crime["MURD21"]) / Years
6
7  # Displaying statistical summary and the first few rows of the DataFrame
8  print(campus_crime)

```

Executed at 2024.04.03 13:58:40 in 29ms

14	Main Campus	FORT WORTH	TX	76129
15	Main Campus	WEST LAFAYETTE	IN	479071076
16	Brite Divinity School	Fort Worth	TX	76129
17	Main Campus	Roseburg	OR	974700226
18	Main Campus	HOUSTON	TX	77004

	men_total	women_total	Total	MURD19	MURD20	MURD21	\
0	8474	37981	46455	0	0	0	
1	1807	4358	6165	0	0	0	
2	6	142	148	0	0	0	
3	1297	2242	3539	0	0	0	

Figure 5: Setup of schoolMurderYearAverage variable.

The next thing that I did was I merged the data. The data is merged on the EffectValue and State with the schoolMurderYearAverage into a single ‘merged-Data’ dataframe which is used for training the model in the next step. Figure 6 shows the merging code.

```

In 232 1  # Convert the statesEffectValue dictionary to a DataFrame
2  states_effect_df = pd.DataFrame(list(statesEffectValue.items()), columns=['State', 'EffectValue'])
3
4  MurderYearAverage = pd.DataFrame(list(campus_crime.items()), columns=['State',
5  'schoolMurderYearAverage'])
6
7  # Merge the campus crime dataset with the states effect values
8  merged_data = pd.merge(campus_crime, states_effect_df, on='State')

```

Executed at 2024.04.03 13:58:55 in 28ms

Figure 6: Merging of data into a single dataframe.

The next step was to train and fit the regression model. A test size of 0.2 was chosen for validation. The EffectValue is chosen as the feature, which can be used to label a campus based on it’s state firearm policy. The target is the schoolMurderYearAverage, hoping to be predicted. The mean square error printed out is 0.1149789836132866 . Figure 7 shows this process.

```
In 236 1 # Define the feature and the target variable
2 X = merged_data[['EffectValue']] # Feature
3 y = merged_data['schoolMurderYearAverage'] # Target
4
5 # Split the data into training and testing sets
6 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
7
8 # Initialize and fit the Linear Regression model
9 model = LinearRegression()
10 model.fit(X_train, y_train)
11
12 # Predict on the test data and evaluate
13 y_pred = model.predict(X_test)
14
15 mse = mean_squared_error(y_test, y_pred)
16 print(f'Mean Squared Error: {mse}')
    Executed at 2024.04.03 14:11:48 in 67ms

Mean Squared Error: 0.1149789836132866
```

Figure 7: Setup of statesEffectValue variable.

Finally, we set up some testing procedures and some programming to allow testing of the various states based on their calculated effectValue. The code for this is shown in Figure 8. The output is shown in Figure 9. We expanded the total set beyond just the midwestern states due to a lack of murders in the campuses in the original states. All values were initialized to 0 and then set after the model tests for that state.

```
In 241 1 # Function to predict the murder rate for a given state
2 def predict_murder_rate(state):
3     effect_value = get_effect_value(state)
4     if effect_value is None:
5         print("State not found in the mapping.")
6         return None
7     # Create a DataFrame with the effect value
8     input_data = pd.DataFrame({'EffectValue': [effect_value]})
9     # Predict using the trained model
10    predicted_murder_rate = model.predict(input_data)
11    return predicted_murder_rate[0]
12
13 # Example usage
14 statePredictedMurders = {'AK': 0, 'AL': 0, 'AR': 0, 'AZ': 0, 'CA': 0, 'CO': 0, 'CT': 0, 'DC': 0,
15    'DE': 0, 'FL': 0, 'GA': 0, 'HI': 0, 'ID': 0, 'IL': 0, 'IN': 0, 'KS': 0, 'KY': 0, 'LA': 0, 'MA': 0,
16    'MD': 0, 'ME': 0, 'MI': 0, 'MO': 0, 'MS': 0, 'MT': 0, 'NC': 0, 'NE': 0, 'NH': 0, 'NJ': 0, 'NM':
17    0, 'NV': 0, 'NY': 0, 'OH': 0, 'OK': 0, 'OR': 0, 'PA': 0, 'RI': 0, 'SC': 0, 'TN': 0, 'TX': 0, 'UT':
18    0, 'VA': 0, 'VT': 0, 'WA': 0, 'WV': 0, 'WY': 0, 'WI': 0, 'IA': 0, 'MN': 0, 'ND': 0, 'SD': 0}
19
20 for state in statePredictedMurders:
21     statePredictedMurders[state] = predict_murder_rate(state)
22
23 print(statePredictedMurders)
```

Figure 8: Setup code for the prediction of various states.

```

{ 'AK': 0.4292349383285141, 'AL': 0.4336560086046739, 'AR': 0.43807707888083364, 'AZ':
0.4690245708139518, 'CA': 0.6989202251742586, 'CO': 0.46018243026163236, 'CT': 0
.570709187165626, 'DC': 0.5618670466133064, 'DE': 0.517656343851709, 'FL': 0.517656343851709,
'GA': 0.4292349383285141, 'HI': 0.6016566790987442, 'ID': 0.4248138680523544, 'IL':
0.6016566790987442, 'IN': 0.4469192194331531, 'KS': 0.42039279777619465, 'KY': 0
.42039279777619465, 'LA': 0.4292349383285141, 'MA': 0.6060777493749039, 'MD': 0
.5397616952325077, 'ME': 0.4336560086046739, 'MI': 0.47344564109011156, 'MO': 0
.4336560086046739, 'MS': 0.4248138680523544, 'MT': 0.42039279777619465, 'NC': 0
.4867088519185908, 'NE': 0.47786671136627135, 'NH': 0.43807707888083364, 'NJ': 0
.5972356088225844, 'NM': 0.46018243026163236, 'NV': 0.49997206274707007, 'NY': 0
.570709187165626, 'OH': 0.4690245708139518, 'OK': 0.45134028970931284, 'OR': 0
.4822877816424311, 'PA': 0.47786671136627135, 'RI': 0.5309195546801883, 'SC': 0
.43807707888083364, 'TN': 0.45134028970931284, 'TX': 0.43807707888083364, 'UT': 0
.45134028970931284, 'VA': 0.49112992219475055, 'VT': 0.4557613599854726, 'WA': 0
.5220774141278688, 'WV': 0.4469192194331531, 'WY': 0.4292349383285141, 'WI': 0
.47786671136627135, 'IA': 0.4557613599854726, 'MN': 0.45134028970931284, 'ND': 0
.43807707888083364, 'SD': 0.4159717275000349}

```

Figure 8: Output for the prediction of various states.

These results will be discussed later in the Results / Discussion paragraph. But the various states showed various murder prediction for each state.

3.2 Data Collection

Data was collected from various sources. The data sources used to train the AI regression model was the RAND State Firearm Database [1], and data from the office of Campus Safety and Security under the U.S. Department of Education [2].

I also retrieved the 2007 – 2013 data from ‘Healthy Minds Publications’ data on student mental health [3], and some demographic data from the National Center for Education Statistics [4]. Neither of these datasets were used to train the model, as they fell outside the scope of the initial experiment.

3.3 Data Normalizing and Correlation

The RAND Firearm Policy Database included a column labeled ‘Effect’ which could either be ‘Permissive’ or ‘Restrictive’ based on the opinion of the data collector. After correlating the policy to a University by it’s state, the ‘EffectValue’ of that state is added plus one for permissive and minus one for restrictive. Not all of these values were set, so we reduced the table to a sample of only the set of rows with the effect value present. The program is also optimized to only count for the labeled rows.

Result / Discussion

The results of the experiment are as follows for each of the midwestern states, defined as “North Dakota”, “South Dakota”, “Minnesota”, “Wisconsin”, and “Iowa”. We also explore predictions for the other states.

1.1 Predicted Number of Murders

Data from campuses across the U.S. was used to average the yearly murder experienced by each campus. This was then used to train the model as a target for the regression model. The results were as follows:

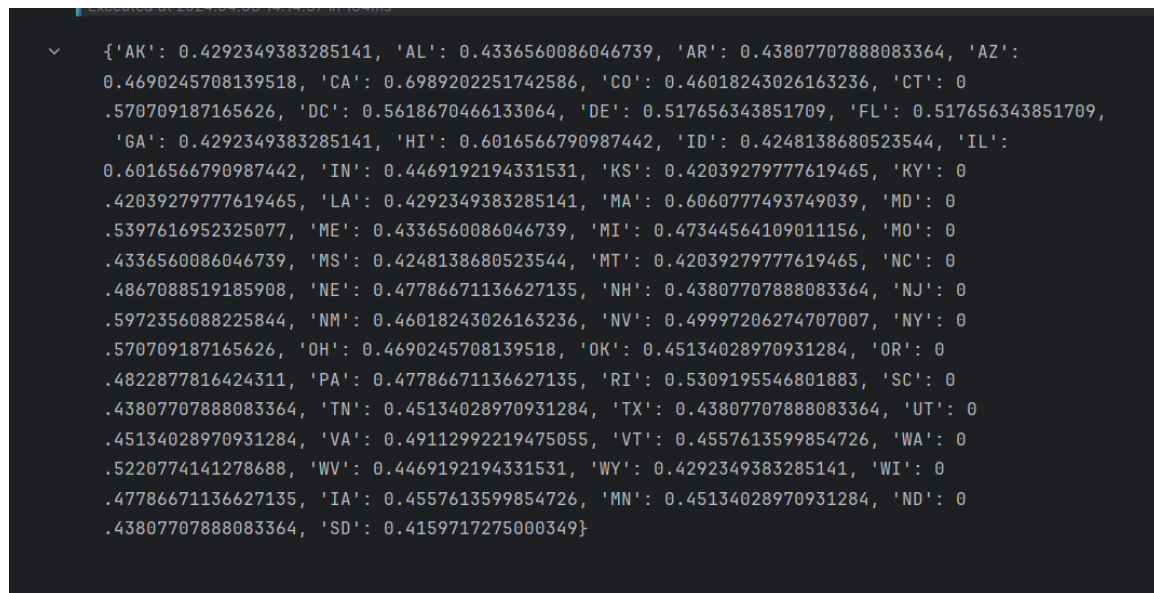


Figure 1: State Results

1.1.1 Midwestern State Results

The results for North Dakota, South Dakota, Wisconsin, and Minnesota are as follows:

State	Predicted Number of Average Yearly Murders
North Dakota	0
South Dakota	0.4
Wisconsin	0

Iowa	0.4
Minnesota	0.4

Table 1: Midwestern on campus murder predictions.

These predictions should not be taken literally, but serve as an example of what can be done with an AI regression model if you predict the on campus crime rate based on the state firearm policies of the state that campus falls into.

Conclusion

This exploratory research offers a preliminary but insightful glance into the potential correlation between state firearm policies and campus crime rates, focusing specifically on murder occurrences in this research. By tallying the effects of state-specific firearm legislation and employing an advanced predictive AI in the form of a linear regression model, this study has taken foundational steps towards understanding the broader implications of policy on university safety.

Our model, built upon data from the RAND State Firearm Law Database and crime reports from various universities, proposes a novel approach to predicting crime rates by accounting for legislative context. The finding suggest that there may be a relationship between the permissiveness or restrictivness of a state's firearm laws and the average number of murders on its college campuses. However, it is crucial to acknowledge the models limitations, giving the complexity of social phenomena and the high number of total factors influencing crime rates that extend beyond the scope of this study.

The potential societal benefits of using AI for policy impact prediction are profound, yet they come with ethical considerations, particularly concerning data privacy, model transparency, and the stigmatization of certain demographics. The results of the study should be interpreted as a framework rather than a definitive predictor, serving as a starting point for policymakers, educational institutions, and public safety officials to consider as part of a holistic approach to campus safety.

Future research is warranted to refine the predictive capabilities of the model by incorporating additional variables, such as economic and educational factors, which may influence crime rates. By advancing the methodologies and expanding the data sets, subsequent studies can build on this work to offer more comprehensive insights and contribute to the creation of a safer campus environment for everyone.

In conclusion, this study has laid the groundwork for understanding the interplay between firearm policies and campus crimes, it underscores a need for a cautious and measured approach to the use of AI in public policy and safety management. The promise of such

technology must be balanced against its limitations and the necessity of human oversight in its application and interpretation.

References

- [1] <https://www.rand.org/pubs/tools/TLA243-2-v2.html>
- [2] <https://ope.ed.gov/campussafety/#/datafile/list>
- [3] <https://nces.ed.gov/ipeds/datacenter/DataFiles.aspx?year=2022&surveyNumber=7&gotoReportId=7&sid=6e1b3982-9fe2-4217-be4c-4461d8b703bd&rtid=7>
- [4] <https://healthymindsnetwork.org/publications/>
- [5] ChatGPT 4 – OpenAI.com
- [6] <https://www.youtube.com/watch?v=i5E2hruuLaQ&t=979s>
- [7] <https://www.youtube.com/watch?v=fATVVQfFyU0&t=769s>