Predicting Terrorist Kidnapping Outcomes: What Are the Most Impactful Features?

David Li

December 15th, 2022

Introduction

The taking of hostages has long been a tactic used by terrorist organizations in order to achieve their political or ideological aims. In recent years, the frequency and severity of such incidents has increased significantly, making it all the more crucial to understand the factors that influence the outcome of a hostage situation. By gaining a better understanding of these variables, it may be possible to improve the chances of a successful resolution and the safe release of the hostages.

Hostage situations have the potential to have serious and long-lasting consequences for both the hostages and the broader society. For the hostages, the experience of being held against their will can be traumatic and can have lasting psychological effects. In some cases, hostages may be subjected to physical abuse or torture, which can further compound the psychological harm. In addition, hostage situations can have significant financial costs, as the terrorists may demand large sums of money or other resources in exchange for the release of the hostages. These events can also have significant political and social consequences, as they often generate intense media attention and public outrage. Hostage situations can also have wider security implications, as they may be used by terrorists to advance their ideological or political agendas or to draw attention to specific issues or grievances (Alexander, 2007)

In this paper, we seek to identify the key factors that determine the outcome of a terrorist hostage situation. By analyzing a dataset of past incidents, we hope to gain actionable insights into the variables that most influence the likelihood of a successful resolution. Our ultimate goal is to establish a threshold for determining the feasibility of extracting the hostages in a given situation. Through this analysis, we aim to provide valuable information to those tasked with responding to and resolving terrorist hostage situations. By highlighting the factors that are most important in determining the outcome, we hope to assist in the development of effective strategies for dealing with these complex and dangerous situations.

Literature Review

This research has revealed that all over the globe kidnapping incidents have risen over 275% in the last 10 years alone (Mohammed, 2007). These ransom kidnapping incidents are becoming more and more transnational meaning people are being kidnapped by citizens of other countries and held for ransom. The reason ransom is such a tricky part of hostage taking terrorism is the fact that ransom crimes have been treated as other crimes besides kidnapping. . *Express kidnapping*, where the victims are temporarily seized, taken to cash machines and forced to withdraw large sums of money, originally started in Columbia but is now most common in countries like Brazil and Mexico (UNODC, 2005; Harrigan, 2005). Double kidnapping is another unique trend of modern kidnapping where the abductors demand a second ransom after the first is paid and treat the first payment as a down payment (Canter, 2009). Another reason why ransom is such an interesting predictor is the business aspect and potential huge monetary gain that can come from it. So much so that there are hidden groups in well-known countries that have task force's devoted to nothing but kidnapping. Al-Qaeda number one leader in the arabian peninsula described terrorist kidnappings as " an easy spoil"(Shortland, 2019). Governments also may not directly take part in ransom but the people they employ sometimes use their position to enforce a "tax" on an unsuspecting victim in their jurisdiction. In Mexico this is a normal occurrence as local law enforcement and groups in tune with local law enforcement commit kidnappings and extort the friends of the victims out of the "tax" also known as ransom. Interestingly enough these types of kidnappings are seldom for the max amount of ransom they can get for an individual but a relatively small amount that really could just be seen as an expensive tax to exist in that environment they operate in(Shortland, 2019). The number of hostages in a terrorist kidnapping is also a interesting part of terrorism in how the outcome of an event like this plays out. In hostage situations the more hostages you have the more leverage you have so it would be interesting to see if that is the case with kidnappings as well. Another interesting parallel between hostage situations and kidnappings that should be explored is the danger hostages face if there is one or more. Do kidnapping victims have more wiggle room to hurt and demand higher ransoms or is it the opposite where the perpetrators are walking a tighter rope? It could also be that the more harm done to single/multiple hostages brings higher levels of government into play. When there is a large number of hostages, logistical success (i.e. the capture of some hostages) may be achieved even though some intended hostages escape(Gaibulloev K & Sandler T, 2009). At the Munich Olympics six hostages escaped from being taken but there were still multiple hostages that were taken during the event. Even though there were hostages that escaped it would still be considered a success because while everyone wasn't taken multiple people were still kidnapped. So while it may seem that out of 10 people attempted 6 got away the 4 captured would still be more than enough leverage for the kidnappers. Authorities may be more inhibited from using lethal force when many hostages are at risk, thereby making such targets more vulnerable at the abduction stage. This could potentially lead to different outcomes of these situations showing the need to study this correlation more.

The location of a kidnapping can possibly be a huge indicator of not only the outcome of a kidnapping but the harm done to an individual during the event. When looking at the location of the terrorist kidnapping the target type also plays a huge factor according to multiple the limited amount of sources found on this subject. There was not a lot of information that was available as far as this predictor is concerned and that is why it is so important that it is studied in depth. When looking at the research done by Forbes listing the most dangerous countries for Americans to visit, the data showed South Asian, Middle eastern and sub Saharan African countries as the most dangerous(Bloom, 2022). This conclusion comes from the US State Department and has marked the countries as level 4. Since level 4 is the highest level these countries have been marked this threat level because "In many high-risk areas, we cannot help you. This may be because of a lack of a functioning government, the ineffectiveness or policies of local authorities, armed conflict or poor governance."(Bloom, 2022). Unable to find any articles talking about the specific target type was very alarming. If it is unknown why a specific individual is targeted in a kidnapping then how will the most at risk individuals be able to guard themselves against this traumatic event. The literature that is on this subject only seems to cover Americans or other well off countries which make sense considering that country is better off than most of the world. By January 2022, at least 51 US citizens were held hostages by non-state actors such as terrorist organizations, criminal groups, pirates, or unknown captors, which includes illegal or unjust detentions sponsored by a foreign government(Ballesteros, 2022). Looking at the most at-risk countries listed on the Forbes website it is shown that the level 4 countries almost all have economic issues or are very close to war-stricken areas. This would indicate that the reason Americans are allegedly targeted more often than others is the reality there will be some type of monetary gain for the attackers. Another country that has problems that up until recently were worse or just as bad as Americans was France. As late as a few years ago, France had the highest number of nationals held hostages overseas, after the United States, according to terrorism watchdog intel center. Most of the French were captured in African countries by mafias or extremist groups such as Boko Haram, linked to Al Qaeda, for ransom and financing of operations(Ballesteros, 2022). Supporting the claim that these kidnappings are indeed driven by financial need. This however does not cover all of the other possible reasons as to why this is such a recurring problem for certain countries. This could be an elaborate plan to try to bring attention to an economic problem that the attackers from these countries try to bring to light. It could be a call to action or it could be a warning for people to stop visiting these high-risk countries as the locals do not want outside intruders in their land. There is simply not enough literature and information out there that doesn't just cover the financial aspect of the crime to make a real decision. This is why further research into this predictor is needed in order to accurately depict the reasoning as to why certain target nationalities are picked for kidnappings and why they take place in certain countries over others. It would also be beneficial to see the age of the kidnapped victims and see what data can be gathered about how they are handled versus others that are younger or older than them. This could be a very valuable predictor in the outcomes of these events that are not readily available.

There is little to no usable literature on weapon type and the duration of the kidnapping which is understandable considering all the seemingly more important factors of a crime of this nature. However, these aspects of a kidnapping could potentially be huge considering if the victims are taken for more than a week there is a bigger injury risk than if they were kept for a couple days. Also, if the victims are kidnaped by people with guns rather than brute force is there more of a risk of an outcome that involves death than others. Would a gun make the attackers bolder or make rescue teams more wary and privy to lethal measures? There is simply not enough literature that covers this to make an educated guess but there absolutely should be research done to look into these details. However if there is more research done to look into these topics more in depth and in detail it could potentially make a huge impact in saving lives in kidnapping situations. That's why learning more about different kidnapping outcomes could lead to beneficial conclusions.

In summary, our problem statement is to determine the variables in a dataset of terrorist hostage situations that most influence the outcome, and more attention to these variables can lead to a reduction in hostage deaths.

Data

Data for this project was acquired from the Global Terrorism Database (GTD), an open-source database on global terrorist events occuring from 1970 to 2020. This dataset has over 200,000 observations, with variables dealing with perpetrators, victims, date, location, and weapons. The GTD claims to be the largest and most comprehensive database publicly available globally. All information within this database is based on reports from open media sources, which are only added once deemed credible by the team of researchers behind this massive undertaking. These sources are obtained from previous academic research and databasing done on terrorism, as well as concurrently updated archives available on the internet.

A particularity comes with utilizing this database for analysis; that it does not clearly define what terrorism is. The GTD has set some guidelines as to what events can be included as terrorism, these are:

1. The violent act was aimed at attaining a political, economic, religious, or social goal;
2. The violent act included evidence of an intention to coerce, intimidate, or convey some other message to a larger audience other than the immediate victims; and
3. The violent act was outside the precepts of international Humanitarian Law.

As can be seen, these are very broad in scale.

The Global Terrorism Database (GTD) offers a wide range of flexibility for research purposes. While this may be seen as a potential limitation for some forms of analysis, it also allows for a wide range of topics to be studied using the dataset. For this particular analysis, no specific inclusion criteria were applied beyond those already present in the GTD. This means that all data included in the database is relevant and within the scope of the research. Additionally, the GTD includes an extra column that indicates whether an incident has doubts surrounding its classification as a terrorist event. While this was not used in this analysis, it is worth noting that further research could potentially benefit from the removal of such incidents from the dataset. However, it is important to bear in mind that a database of this size, compiled from open-source media sources, may contain a small amount of factual errors or missing information. This is an inevitable issue with datasets of this nature, and researchers must take this into account when interpreting the results of their analysis.

As for the variables included in this database, the GTD includes a large number of useful variables dealing with location, date, perpetrators, victims, weaponry, and of the event itself. As discussed in the literature review section of this report, the particular predictor variables that are of note for this analysis are: Whether ransom was demanded, number of hostages, region of event, weapon type, target/victim type, and duration of the event. As for the target variable, the GTD includes a column that denotes the outcome of a terrorist kidnapping event. For an observation to be included in this column, its type of attack must be a hostage-taking incident, which includes kidnappings, successful hijackings, and barricade incidents. The large majority of the events included in this are kidnappings, though hijackings and barricade incidents should not be discounted as valuable data points.

Methodology

The main goal of this analysis is to predict a categorical target variable, therefore, decision trees and random forest models were created for this purpose. Although there certainly are other options, these models seemed to be the highest performing and easiest to implement. As for the variable used in the models, the predictors (as stated in the data section) are: Whether a ransom was demanded, number of hostages, region of event, weapon type, target/victim type, and duration of the event. The target variable used for the models present in this analysis is the outcome of a kidnapping. No data collection was required for any of the variables as they are all included to an acceptable extent in the GTD. To start the process of data cleaning, the GTD as a whole was imported and filtered to include only rows that had values in the target variable column. This would be the main dataset used for all further cleaning and modeling.

For this analysis, all unknown and null values were removed, mostly because the database was large enough to allow this. Another reason for this was that the distributions already generally favored a category or value heavily, and taking the average and redistributing would only add to this disproportionate distribution. Another option would have been to distribute these unknowns and null rows according to the distribution; this could affect a model's predictive power and was deemed unnecessary. After all data cleaning and dropping all null values, the dataset remaining had a size of 2800 observations, prior to any partitioning.

There were only two numerical variables present in this analysis. Number of hostages was a relatively simple predictor to work with, as it did not include many null values. This predictor has a distribution heavily skewed right, towards the lower end of the spectrum, with 1 hostage being the dominant value with ~50% of the total amount of values. This predictor along with the other large numerical predictor, duration of kidnapping, were both binned into more manageable categories and then one-hot encoded into dummy variables for use in the model. The bins for the number of hostages predictor still has a skewed distribution towards the lower values, since 1 remains the most dominant value. However, the duration of the kidnapping predictor was put into relatively balanced bins of around 800-1000 values each. Bins for the number of hostages were set at 1 being low, 2-5 being moderate, and 5+ being high. These reflect the nature of kidnappings and hostage situations, as 1 is generally the most common, 2-5 is an event with a multitude of hostages, and 5+ can be categorized as a mass kidnapping. The bins for the duration of the kidnapping were set at 1 day or less being low, 2-7 days being moderate, and 7+ being high. These reflect how people generally think of time, with a day being a short amount of time, a week being a somewhat larger amount, and anything above a week being a substantial amount of time.

Duration ended up being the most difficult predictor to work with, as many observations had a null value for this column, around 30% of the whole dataset. It was also separated into two columns, one for hours of kidnapping, another for days of kidnapping. The value was coded as such that as time increased past 24 hours, it was put into the days of kidnapping column as a whole integer value stating the number of days. This means that to utilize this variable, these two columns had to be merged. To do this, the days of the kidnapping column were multiplied by 24 so that the units of measurement were the same, and then both columns were merged together to form the final predictor used in modeling. One drawback of the way this was coded into the dataset is that there is no hour incrementation for anything past 24 hours, which leads to some breakpoints in the distribution at 24, 48, 72, and 96 especially. Regardless, this was not of importance since binning resolved this issue.

Looking at the categorical predictors, whether a ransom was demanded was included as a binary variable, although some data points were coded as -99 for “unknown”. This predictor had a small amount of null values, though it did not account for a significant portion of available data. The other categorical variables were not as simple as this, with all of them needing to be one-hot encoded into dummy variables for ease of use by the model. Region, target/victim type, and weapon type were all made into dummy variables, which resulted in quite a bloated list of predictors (around 50). Unknown values were present in these predictors as well, with weapon type being the largest of the categoricals at around 1400 unknowns, these were dropped. No real data manipulation needed to be done other than the encoding for these variables.

This report's analysis is focused on which predictors are most important in the prediction of each different outcome of the target variable. The target variable did need some cleaning, as many of the categories present were not suitable for analysis. The categories of the target (in order) are:

1. Attempted Rescue
2. Hostage(s) released by perpetrators
3. Hostage(s) escaped (not during rescue attempt)
4. Hostage(s) killed (not during rescue attempt)
5. Successful Rescue
6. Combination
7. Unknown

As can be seen, many of these are inconclusive in nature. Category 7 is simply "Unknown" and was dropped completely from the dataset as it did not provide any real precise value. Category 6, "Combination", also does not provide any real conclusion and was not used in the final analysis, though these data points were not removed. The same can be said for category 1, as it does not state an actual conclusion; Were the hostages killed during the rescue attempt? Did the hostages escape? Was the rescue unsuccessful? This category simply does not answer any of these questions as it is coded, therefore it was also omitted from analysis. Finally, category 3, “Hostage(s) escaped” does provide a conclusion, however, it was an extremely small fragment of the overall data, and did not present significant results after resampling, thus, it was not included in the final analysis. Therefore, modeling was done with the 3 main categories of the target variable: Category 2 (Hostages released), category 4 (Hostages killed), and category 5 (successful rescue). These three categories were ~88% of the total target variable after dropping category 7. The categories also are the most prominent for the creation of suggestions, as category 2 is generally the result of negotiation, category 4 means authorities failed to save victims, and category 5 means authorities succeeded in rescuing victims. The target variable was one-hot encoded into dummy variables for use in the separate models corresponding to each respective category. This was done because this analysis requires an understanding of each respective category separately.

As for the actual modeling, decision trees (CART, C5.0) were originally made alongside a random forest model, however, after looking at accuracy values and understanding that a large list va of predictors is not ideal for regular decision trees, it was decided that random forest models would be created for each category. Train/test splitting was made after data cleaning was completed. The baseline accuracy was determined, then a basic model with all predictors and no parameter tuning was conducted. An accuracy value was determined with k-fold cross-validation with k=10. These values are shown in Table 1:

Model	Baseline Accuracy	Initial Model Accuracy
Category 2: Hostage(s) released	50%	65%
Category 4:	69%	76%

Hostage(s) killed		
Category 5: Successful rescue	92%	92%

As stated before in this section, since the list of predictors was quite heavily bloated, this meant that every model that was run needed a cut in predictors based on relevance, therefore, feature importance was determined, and dummy variables that were deemed unimportant at less than 3% model importance were removed. After this, the models were run again with the shortened list of predictors, and, as expected, a slight decrease in accuracy was observed. This is simply due to cutting predictors below 3%, as these do still provide some importance, just a small amount, resulting in a small decrease in accuracy.

Parameter tuning was conducted on the models at this point with number of estimators, max depth, and max leaf nodes being the main targets of this tuning, along with some testing changing the train/test split percentage and k value for k-fold cross-validation. The number of estimators, max depth, and max leaf nodes were included in parameter tuning simply to restrict the forest model in its scope to determine whether that would result in a better accuracy value and lower complexity. After determining the best values for these values and running the models with these parameters, there does seem to be a slight improvement, though not by much. These parameters do reduce the variance that can be seen in k-fold scores as well as overall accuracy score differences based on random state. Therefore, having the perfect values and combinations could lead to the model performing at its best state regardless of random state. The accuracy differences between inclusion and exclusion of these parameters ends up being somewhat relevant, however, as these models are not particularly accurate even with parameter tuning, finding the perfect combination of parameters was deemed unnecessary. As for the train/test split tuning, it can be seen that model performance increases in the training cross-validation scores but decreases in the test cross-validation scores and final evaluation scores as the training set percentage increases, this is true for the opposite as well. However, as training set percentages lower past a certain point, the model becomes less and less accurate overall. The value of 75-25 seemed to be the best at providing a good balance of training fit and evaluation score. There were only detrimental effects to changing the k value for k-fold cross-validation from 10, therefore that was the value used. Accuracy values after parameter tuning with k-fold cross-validation on the training set as well as the models prediction accuracy on the test set are provided in the results section.

Another possibility that was explored at this point of the analysis was the resampling of the training dataset. Downsampling or upsampling were both considered, as these processes would create a target with even values across the categories, perhaps providing a better training result for the model, resulting in more accurate predictions of unseen data. This should, in theory, provide even better results for categories that have low values in the target, allowing for these categories to have a better training process. Resampling also results in categories no longer being favored over another, as they have the same exact value count. This certainly would be an advisable path for this analysis if the target variable was not encoded into dummy variables and separated in the models. Regardless, an approach that included downsampling of the training set to the lowest value present, in this case category 5 with 162 observations, was tested. Downsampling was chosen as this would keep all values of the target variable even without needing to add any artificial data points. The same process of data cleaning, parameter tuning, and evaluation was used on these models. The result was an overall decrease in performance, except for category 4, which had a somewhat similar result to the model with no downsampling.

The decrease in performance was especially noticeable when running a prediction on the unseen test data. It was therefore determined that these models should not be utilized, and no resampling was done. This generally means that the models with greater number of observations should have a better accuracy value in comparison to the baseline accuracy, since they have a greater amount of training data to create patterns on. Discussion of the accuracy values and the feature importance for each of the final models will be done in the results and conclusions section of this report.

Results

This section will discuss the results obtained from the evaluation methods used for the purposes of analysis. Final model accuracy determined with k-fold cross-validation on the training data as well as the models' predictions of unseen test data provided by the train/test split is shown in Table 2:

	Baseline Accuracy	**Final Accuracy Value**	**Prediction on Unseen Data**
Model 1: Hostage(s) released	50%	65.8%	64.1%
Model 2: Hostage(s) killed	69%	76.6%	76.3%
Model 3: Successful rescue	92%	92.3%	90.7%

As can be seen, the first two models, which had lower baseline accuracy, do provide some increase over this baseline, while the third model shows little to no improvement over simply guessing the largest category of the target. All models had greater accuracy when evaluating on the training set than when predicting on unseen data, which is expected. This decrease in accuracy does seem to show some slight overfitting on the training set, though this is slight. The model dealing with category 5 seems to be underfitting the data quite heavily as it simply does not and cannot predict the outcome correctly. This is most likely due to choice done with resampling, as upsampling this data would certainly increase the model's overall accuracy. Further evaluation methods were utilized for all three models, as show in Table 3, 4, 5:

Table 3: Evaluation metrics of model on hostages released (category 2)

	Precision	**Recall**	**F1-Score**
0	0.68	0.55	0.61
1	0.62	0.73	0.67
Weighted Average	0.65	0.64	0.64

Table 4: Evaluation metrics of model on hostages killed (category 4)

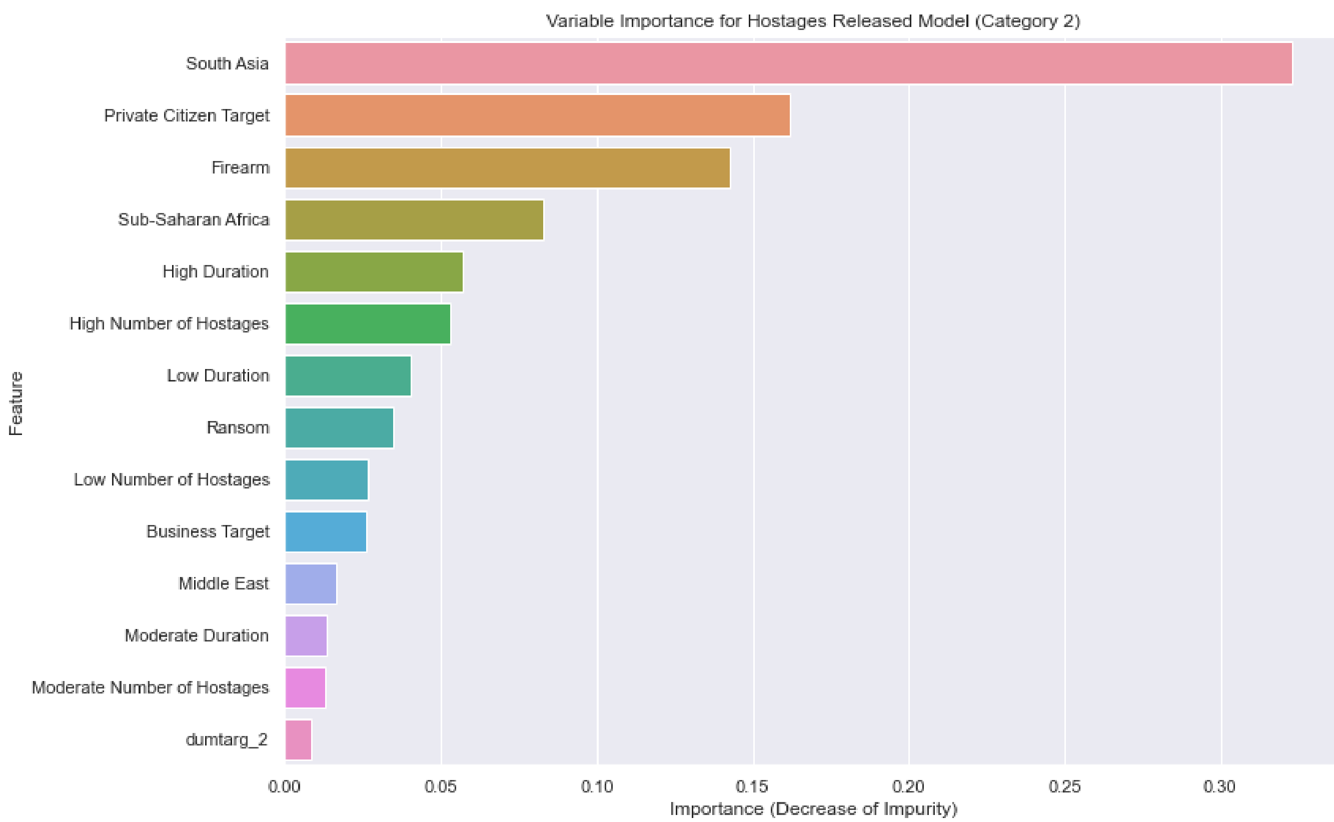
	Precision	**Recall**	**F1-Score**
0	0.78	0.91	0.84
1	0.67	0.42	0.68
Weighted Average	0.75	0.76	0.74

Table 5: Evaluation metrics of model on successful rescue (category 5)

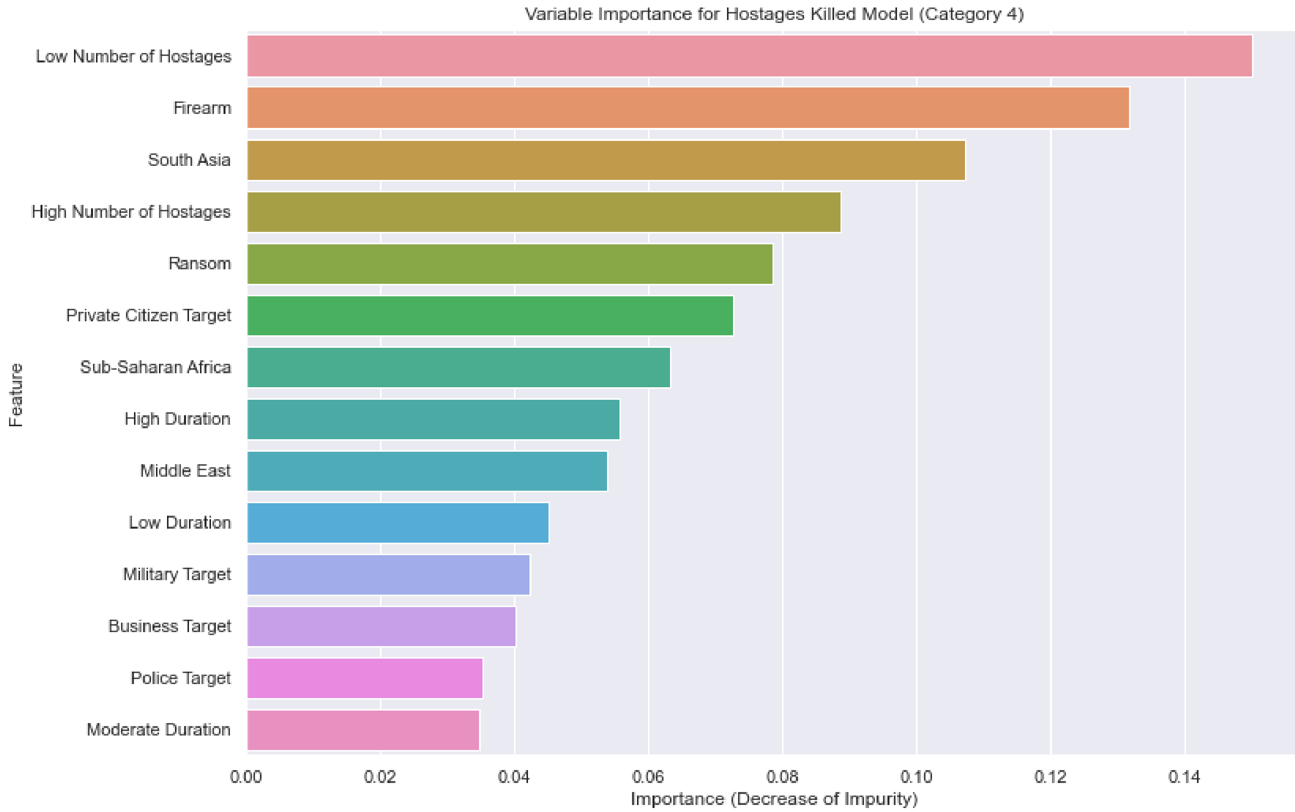
	Precision	**Recall**	**F1-Score**
0	0.91	1.00	0.95
1	0.00	0.00	0.00
Weighted Average	0.82	0.91	0.86

As can be seen, both the first and second model have generally good precision scores, with slight peculiarities with their recall scores. Table 3 shows that model 1 has much better recall on observations with released hostages as compared to those without this outcome. This is quite interesting, meaning the model does well in finding all cases of released hostages but miscategorizes other observations while doing this. Meanwhile, Table 4 shows the complete opposite of this, with model 2 having much better recall on observations that are not of the target outcome. Regardless, the scores show a general reflection of the overall model accuracy.

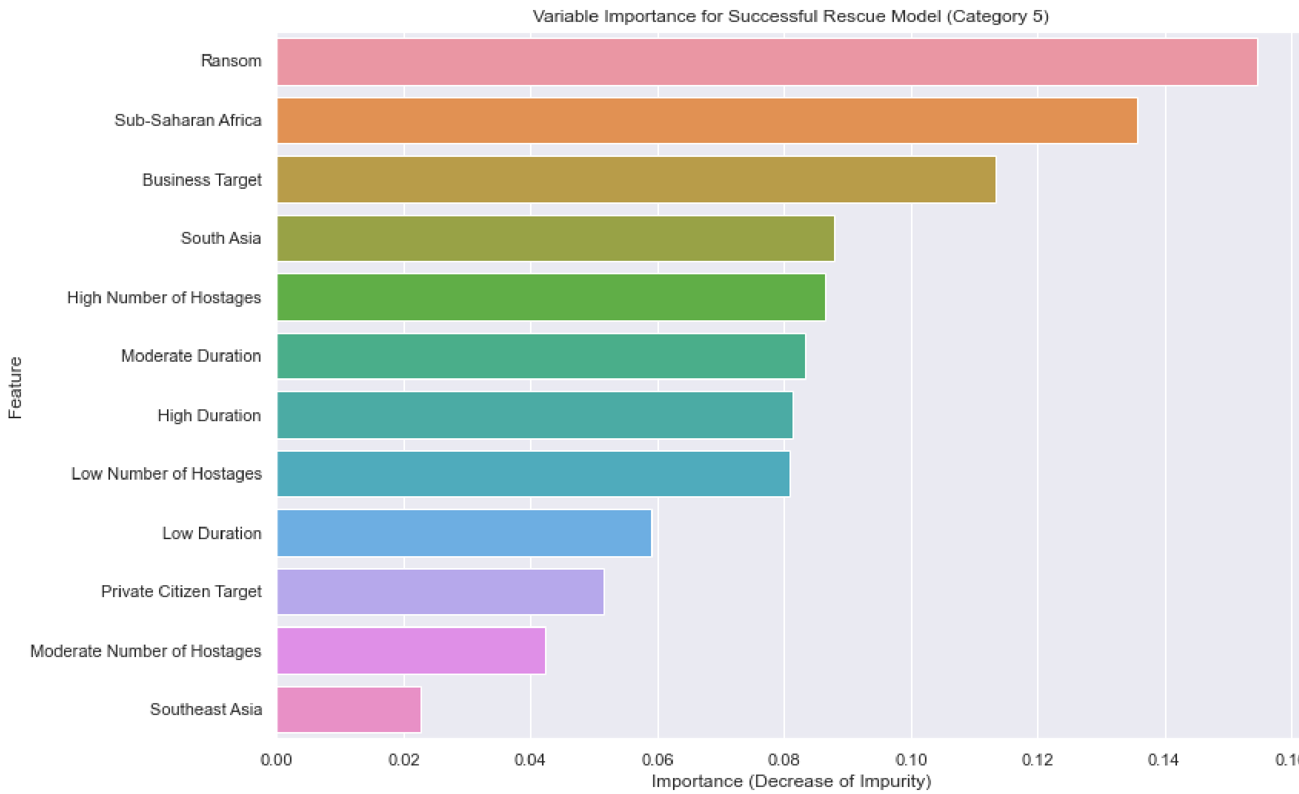
Meanwhile Table 5 shows that model 3 is not predicting its target outcome correctly whatsoever, with no evaluation metrics showing any form of positive predictive value for this category. As the target category is small, this is somewhat understandable and simply means this model has a large amount of underfitting on the data and cannot predict the desired outcome at all. Regardless of this, feature importance was determined for each model, this is so conclusions can be made, no matter how inconclusive the models are, this is shown in Graphs 1, 2, 3:



As can be seen from this graph, the South Asian dummy variable coming from the region predictor is by far the greatest contributor in this model's predictive power. This is followed by having a target that is a private citizen, and the use of firearms as a weapon type. These results show that South Asian countries have more kidnappings ending in released hostages, with private citizens and firearm use being other important factors in determining this outcome.



For model 2, having a low number of hostages seems to be the greatest contributing factor to this model's predictive power, with firearm use, South Asia, and having a high number of hostages being other leading predictors. It does seem to be logical that a low number of hostages leads to death more often, as it is easier for terrorists to control and monitor these hostages. Meanwhile, having a high number of hostages means that there are simply more targets for harm for terrorists. Firearms are some of the deadliest weapons available publicly, therefore this being a strong predictor is not surprising. South Asia is seen again, much like model 1, though it is generally unclear as to why this predictor is so valuable, this could be a topic of interest for future research.



Model 5, regardless of its inaccuracies, still shows interesting differences in feature importance compared to the other models. Whether a ransom was demanded is the most important factor to this model, while this was an important factor to the other models as well, not to this extent. Sub-Saharan Africa is also seen, while South Asia is somewhat important though less than other models. These regions being in the top of feature importance is consistent with previous literature. The target type being a business is another important factor, this is quite an interesting proposition, meaning that perhaps successful rescues are more common for hostage situations involving businesses as targets.

Tools

The tools that were utilized in this analysis are: data cleaning and manipulation tools, modeling techniques, model evaluation methods, and coding tools. Many of these were previously mentioned and reviewed in previous sections of this report. Data cleaning and manipulation methods discussed and utilized at some point in this analysis are: Binning, one-hot encoding, dummy variables, and resampling. As stated, binning, one-hot encoding, and dummy variables were used in the final models of this analysis, while resampling was attempted but generally unsuccessful. Modeling techniques used and tried in this analysis are: decision tree modeling and random forest modeling. As decision trees did not accurately predict the target variable, they were not used in the final modeling phase. Random forests were used for all models and are generally quite successful, though model 3 is quite inaccurate. This inaccuracy is most likely not due to random forest being the model choice but due to decision-making surrounding data manipulation and predictor choice. Model evaluation methods include: k-fold cross-validation, train/test splitting, precision, recall, and F1-score. All of these were used to evaluate model performance and determine their viability in use outside of the dataset used. They showed that the model does not reflect performance adequate enough for use on prediction of unseen data of such importance as terrorist kidnapping incidents.

As for the coding tools used in this analysis, several Python libraries were used to code all of the required segments. These libraries include scikit-learn for all modeling classifiers and evaluation methods, Pandas for data cleaning needs, Seaborn for plotting of feature importance, and Matplotlib for visualization of predictors in the data exploration phase. All of these libraries are open source coding tools for use in data science projects such as this. The main coding tool all of this was done in is Jupyter Notebooks, a web-based, free, integrated development environment for Python coding. All of these tools were successful in completing what was required of them for the purposes of this project.

Ethical Implications

Because of the nature of the subject matter of this project. There are some ethical implications to take into account. Firstly how it may be interpreted by policymakers and decision makers may interpret the results gained from this project. For example applied force or the applied use of police force. If a police force were made to act based on the predictions from the models in this project. They would want to prevent a hostage situation if there were a low number of hostages. So the question would be, would this hypothetical police department resort to lethal force in a faster manner than they would have otherwise, out of fear that hostages may be at greater risk of being harmed, because of the prediction made by the models present in this project. This is both a serious and realistic concern as Eubanks states in “Automating Inequality” that people oftentimes will defer to a decision made by an algorithm over their own discretion no matter how experienced they might be. As algorithms are often assumed to be unbiased and accurate by the uninitiated even though that is not always the case (Eubanks, 2019).

To counteract this possible unethical effect a disclaimer and recommendation to policymakers and decision makers must be made when presenting the results of the models present in this project. A disclaimer stating that the results of this project are firstly inaccurate thus not a good predictor of hostage outcomes and that utilizing the results gained from this project in policy making or decision-making could lead to real harm and questionable ethics. And while presenting the results of the presentation a recommendation could be made. For example, because of the seriousness of this subject matter. As in life or death. We recommend that these results not be taken into consideration for any substantial decision-making or policymaking. Thus the ethical effects that come along with the results of this project should be minimal.

Conclusion

Overall, to answer our research question, “Can we predict what factors are most important in determining hostage outcomes?”. There cannot be any significant conclusions made. As the results from the three models each corresponding with a specific hostage outcome such as hostage(s) released, hostage(s) killed or successful rescue are either not accurate enough more than baseline to be used in any substantial decision making or in one case the final accuracy does not even surpass baseline accuracy. Thus the models in this project are quite useless when making actual predictions. Because of the poor accuracy results and the seriousness of the situations this model predicts it would be advisable to not use the results gained from this project in any substantial potential policy making or decision making by authorities during a hostage situation.

Despite the lackluster accuracy results there are still useful conclusions that can be gleaned from the feature importance performed on each hostage outcome. For example, despite the models inaccuracies. Every single model pointed to the variables of duration of hostage event and number of hostages taken as the two single most important predictors of a certain outcome relative to all other predictors used by a large margin. Another useful observation to glean from the models may be the fact that an important variable in determining whether hostages were released or killed is firearms. But hostages rescued does not have the firearms variable present on its feature importance chart. A third interesting observation is that hostages that are rescued are often rescued because of a ransom demanded. This could go on and on. So while the accuracy of the models were not fantastic there are still useful conclusions that can be ,not applied, but drawn that could be useful for future research on hostage situations and their respective outcomes.

In the future, to improve upon the models applied in this project. The possibility of using multiple datasets should have been explored further. As only the GTD was used in this project and even that was simplified and filtered down to a great degree to make it usable and fitting for the parameters of this project. Having more data to train models on may have improved the accuracy. Thus making the models potentially useful in making actual predictions. However as it stands, the models created are satisfactory as, despite the accuracy, there are useful conclusions to be observed that can be utilized in further research on this topic.

References

Alexander, David A, and Susan Klein (2009). "Kidnapping and Hostage-Taking: A Review of Effects, Coping and Resilience." Journal of the Royal Society of Medicine, U.S. National Library of Medicine, Jan. 2009, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2627800/.

Ballesteros, F. L. (2022, October 25). *US citizens are the prime targets for kidnappers abroad*. Itempnews. Retrieved December 14, 2022, from https://www.itempnews.org/2022/01/20/kidnappers-abroad/

Bloom, L. B. (2022, April 14). *35 countries where Americans are most likely to get kidnapped*. Forbes. Retrieved December 14, 2022, from https://www.forbes.com/sites/laurabegleybloom/2019/04/25/35-countries-where-americans-are-most-likely-to-get-kidnapped/?sh=40d91970bdf1

Canter, D. V. (2009). *The faces of terrorism multidisciplinary perspectives*. Wiley-Blackwell.

Eubanks, V. (2019). Automating inequality: How high-tech tools profile, police, and punish the poor. Picador.

Gaibulloev, K., & Sandler, T. (2009). Hostage taking: Determinants of terrorist logistical and negotiation success. *Journal of Peace Research*, *46*(6), 739–756. https://doi.org/10.1177/0022343309339249

Mohamed, M. K. (2007). Kidnap for ransom in South East Asia. *Asian Journal of Criminology*, *3*(1), 61–73. https://doi.org/10.1007/s11417-007-9040-1

Shortland, A. (2019). *Kidnap: Inside the ransom business*. Oxford University Press.