PREDICTING THE SUCCESS OF TERRORIST ATTACKS USING LOGISTIC REGRESSION ANALYSIS

Jaden Goodwin
STA3180

INTRODUCTION

- Terrorism has continued to be a threat to global and national security
- 232,000 deaths worldwide over the span of 2000 – 2020
  - Highlight risk outcomes for future attacks based on the current number of successful ones
- Research goal: Provide data algorithms to help:
  - Predict future threats
  - Investigate terrorism not as an academic topic but as an evolving global issue
- GTD: highlights many mass casualties based on variables such as country, city, exact location, summary, and success

LITERATURE REVIEW

- Found that there is a gap in the literature about predicting success with more recent global data (Pape, 2003)
  - Suicide attacks are identified as highly lethal due to attacker commitment and ability to bypass security (Pape, 2003)
- Qu (2024) used confirmatory factor analysis to identify harm level indicators, which combined subjective harm (per area) with objective metrics to score on a global basis
- Xu (2024) identified key regions where terrorism has occurred for visual purposes, finding that prominent regions included Middle and North Africa, South Asia, Central Asia, and Central America and the Caribbean (Xu, 2024, p. 1)

LITERATURE REVIEW

- Logistic Regression to Predict Likelihood of Future Incidents
  - Tarakji (2021) performed logistic regression to classify events
    - Random forests for capturing non-linear relationships
    - Neural networks to enhance prediction accuracy
- Logistic regression revealed statistically significant predictors:
  - Political instability, unemployment, and proximity to conflict zones
- Each factor increased the likelihood of an attack by 10–15% (Tarakji, 2021, p. 31)
  - Found an increase in various regions like the Middle East and South Asia (Tarakji, 2021, p. 56-57).

DATA

- The data used comes from the Global Terrorism Database 2000–2020 records. To clean the data, doubtful cases, cases with missing predictor values, outliers, and undefined categories are excluded
- Variables: The predictor variables used are “attacktype1”, “region”, “weaptype1”, “targtype1”, “nperps”, and “suicide”. They each had their own levels/responses. Since the response variable “success” is binary the two possible outcomes are 1=successful and 0=unsuccessful

START

NATIONAL CONSORTIUM FOR THE
STUDY OF TERRORISM AND RESPONSES TO TERRORISM

PREDICTING THE SUCCESS OF TERRORIST ATTACKS USING LOGISTIC REGRESSION ANALYSIS PART 2

Jaden Goodwin
STA3180

METHODS

METHODS: ANALYSIS

- Logistic Regression Model: the primary modeling method with 5% alpha level to fit binary nature of the response
- AIC Stepwise Selection: AIC helps determine the best variable combination for fitness and complexity
- K-fold Cross Validation: cross validation model with k=10 helps prevent overfitting
- Confusion Matrix: the confusion matrix helps determine the sensitivity and specificity of the prediction model
- ROC Curve and AUC: the ROC curve helps to visualize and analyze model performance across different classification thresholds and the AUC measures the overall performance of the binary classification model
- Youden's J-Statistic: this method attempts to balance the sensitivity and specificity of the model and is determining the optimal classification threshold

RESULTS

LOGISTIC REGRESSION

- Binary response: success
- Model reduction: stepwise selection
- Chosen model: 5 predictors: suicide, attacktype1, region, weaptype1, and targtype1
- Unstable estimates for predictor weaptype1: large standard errors and coefficients

weaptype1Chemical		439.747
weaptype1Radiological		520.865
weaptype1Firearms		439.746
weaptype1Explosives		439.747
weaptype1Fake Weapons		439.748
weaptype1Incendiary		439.747
weaptype1Melee		439.746
weaptype1Vehicle		439.746
weaptype1Sabotage Equipt		510.354
weaptype1Other		439.747

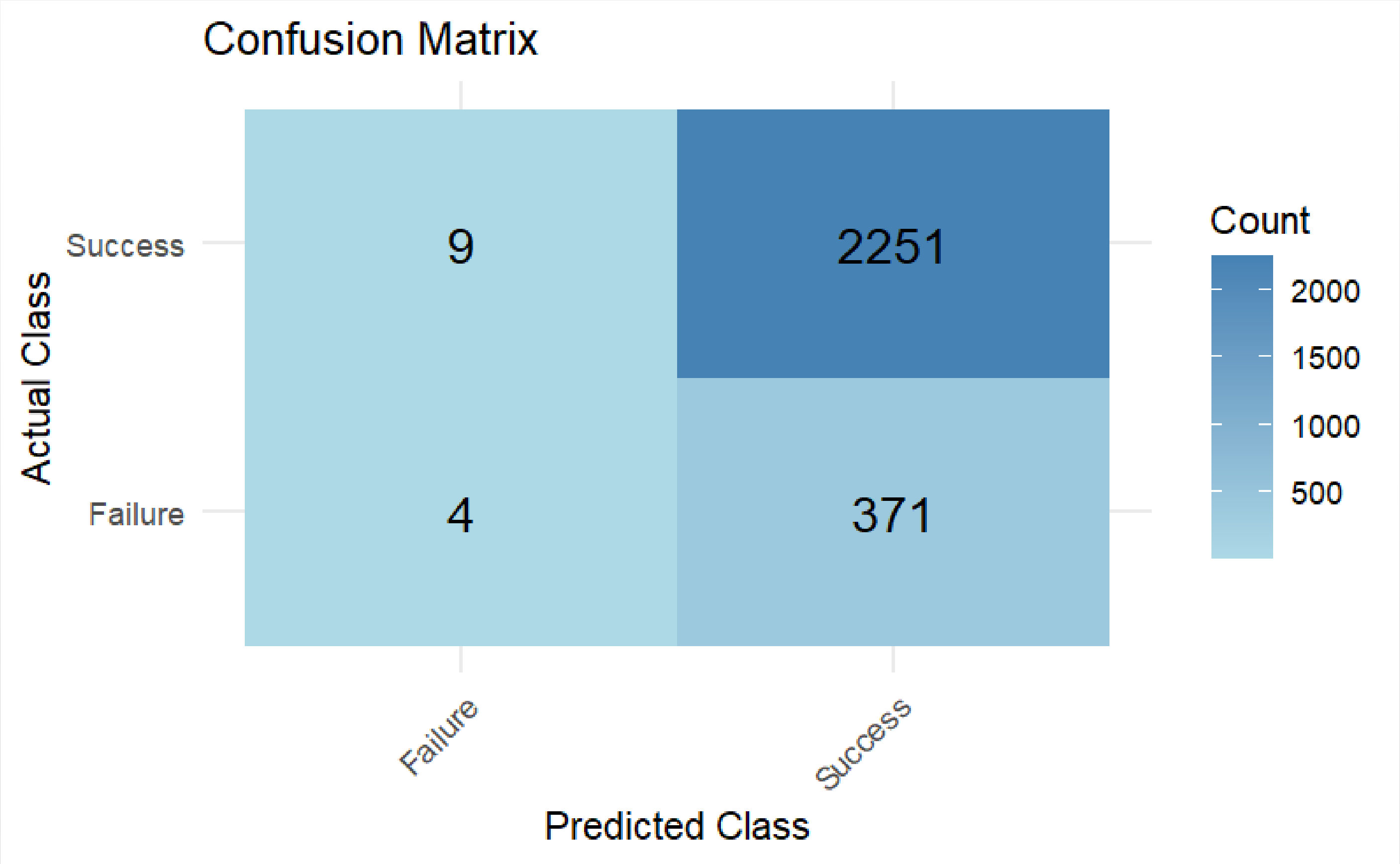
LOGISTIC REGRESSION

- Final model with four predictors: attacktype1, region, suicide, and targtype1

	Dependent variable:
	success
attacktype1Armed Assault	1.776***
	(0.082)
attacktype1Bombing	0.939***
	(0.079)
attacktype1Hijacking	1.730***
	(0.379)
attacktype1Hostage Barricade	2.977***
	(0.595)
attacktype1Hostage Kidnapping	1.684***
	(0.165)
attacktype1Infrastructure Attack	2.089***
	(0.156)
attacktype1Unarmed Assault	1.452***
	(0.199)
regionCent.Amer/Carib	2.119**
	(1.066)
regionS.America	0.957***
	(0.216)
regionE.Asia	-0.284
	(0.271)
regionSE. Asia	1.271***
	(0.158)
regionS.Asia	0.758***
	(0.147)
regionCent.Asia	0.026
	(0.367)
regionWest.Europe	0.157
	(0.177)
regionEast.Europe	0.631***
	(0.199)
regionMidEast N.Afr	0.451***
	(0.151)
regionSub-Saharan Afr	1.134***
	(0.165)
regionAustr/Oceania	0.282
	(0.689)
suicide1	0.236***
	(0.077)
targtype1Government	-0.491***
	(0.119)
targtype1Police	-0.273**
	(0.118)
targtype1Military	0.062
	(0.158)
targtype1Abortion Related	-1.499**
	(0.599)
targtype1Airports	-1.890***
	(0.326)
targtype1Diplomatic	-0.850***
	(0.196)
targtype1Educational	-0.152
	(0.202)
targtype1Food/Water Supply	10.107
	(128.842)
targtype1Journalists	-0.051
	(0.188)
targtype1Maritime	-1.453***
	(0.496)
targtype1NGO	0.227
	(0.393)
targtype1Private Citizens	-0.142
	(0.117)
targtype1Religious	-0.184
	(0.151)
targtype1Telecommunication	1.749*
	(1.016)
targtype1Terrorists	0.681**
	(0.266)
targtype1Tourists	1.115
	(1.028)
targtype1Transportation	-1.127***
	(0.157)
targtype1Utilities	-0.603*
	(0.310)
targtype1Political	-0.076
	(0.220)
Constant	0.171
	(0.184)
Observations	13,174
Log Likelihood	-4,973.741
Akaike Inf. Crit.	10,025.480
Note:	*p<0.1; **p<0.05; ***p<0.01

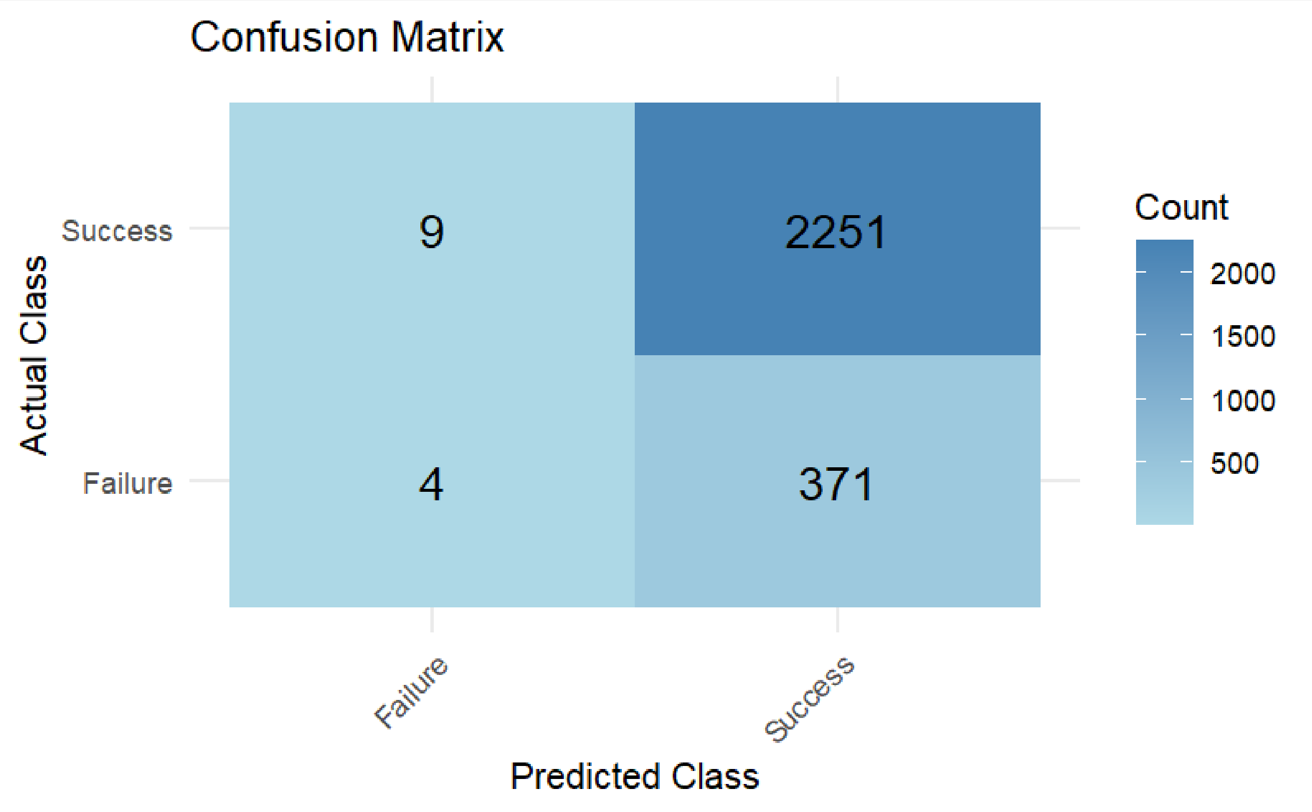
TEST DATA CONFUSION MATRIX

- 80-20 training and test datasets
- Threshold: 0.5
- Accuracy 0.8558
- 95% CI: [0.8418,0.869]



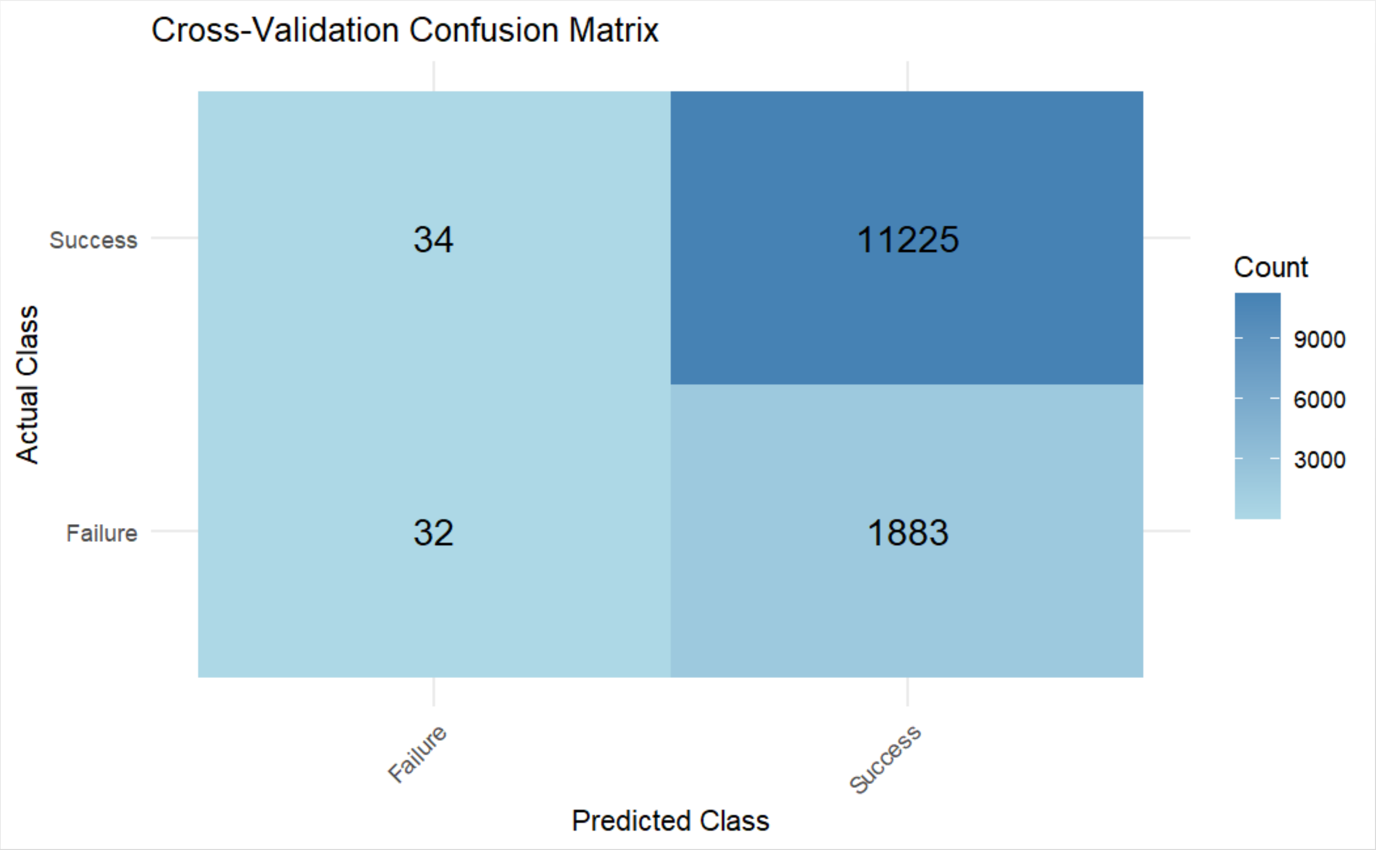
TEST DATA CONFUSION MATRIX

Performance by Class	
Metric	Results
No Information Rate	0.858
P-Value [Acc>NIR]	0.6227
Sensitivity	0.996
Specificity	0.011
PPV	0.859
NPV	0.308
F1	0.922
Prevalence	0.858
Detection Rate	0.854
Detection Prevalence	0.995
Balanced Accuracy	0.503



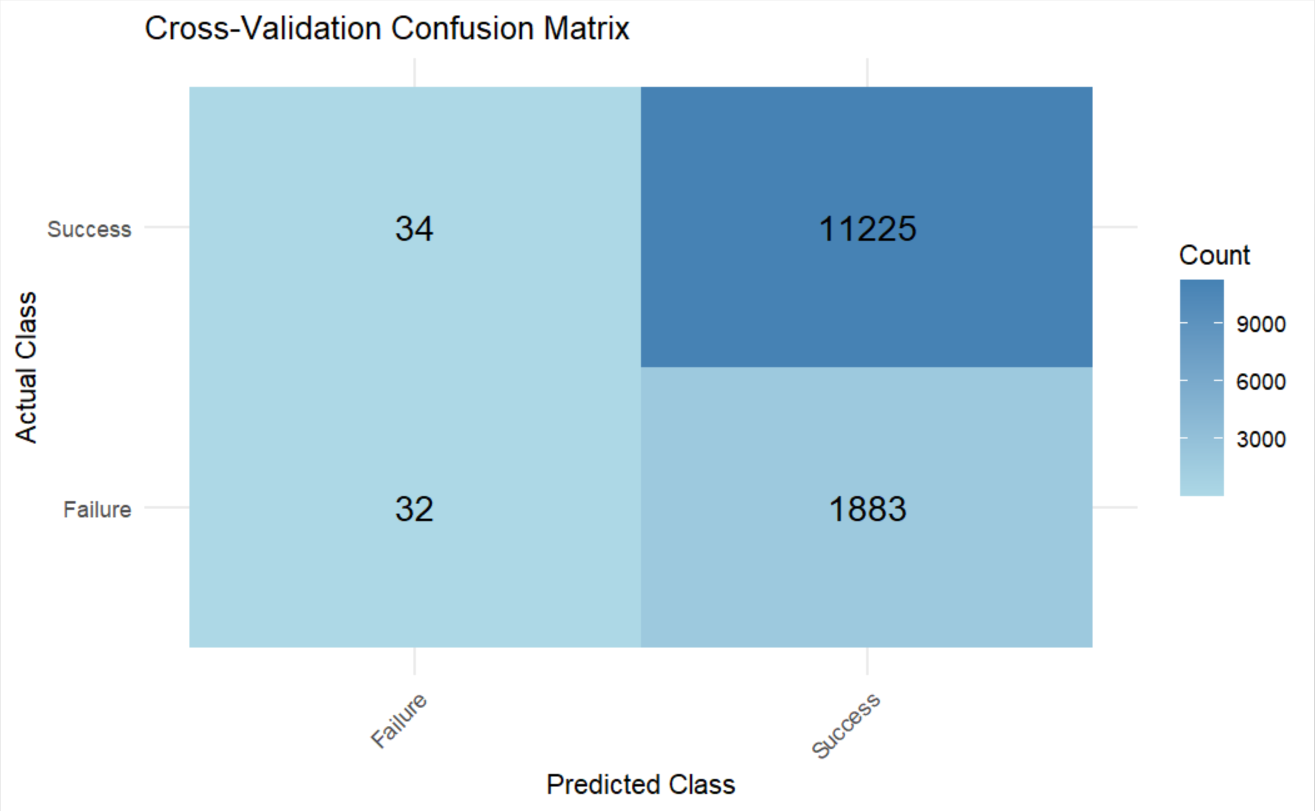
CROSS-VALIDATED DATA

- k=10 folds
- Threshold: 0.5
- Accuracy: 0.854
- 95% CI: [0.8483,0.0.8605]

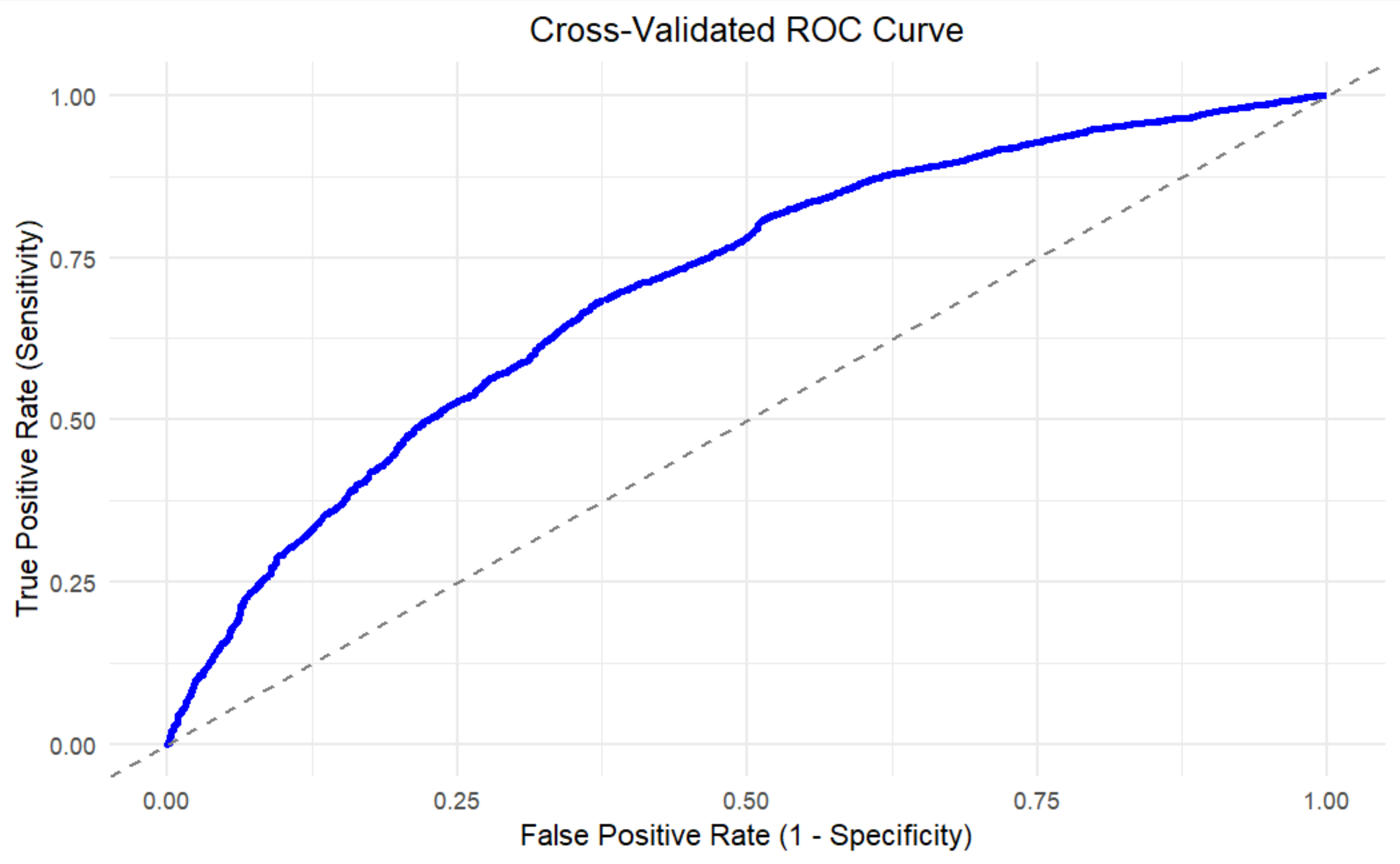


CROSS-VALIDATED DATA

Cross-Validation Performance by Class	
Metric	Result
No Information Rate	0.8545
P-Value [Acc > NIR]	0.5258
Sensitivity	0.997
Specificity	0.017
PPV	0.856
NPV	0.485
F1	0.921
Prevalence	0.855
Detection Rate	0.852
Detection Prevalence	0.995
Balanced Accuracy	0.507



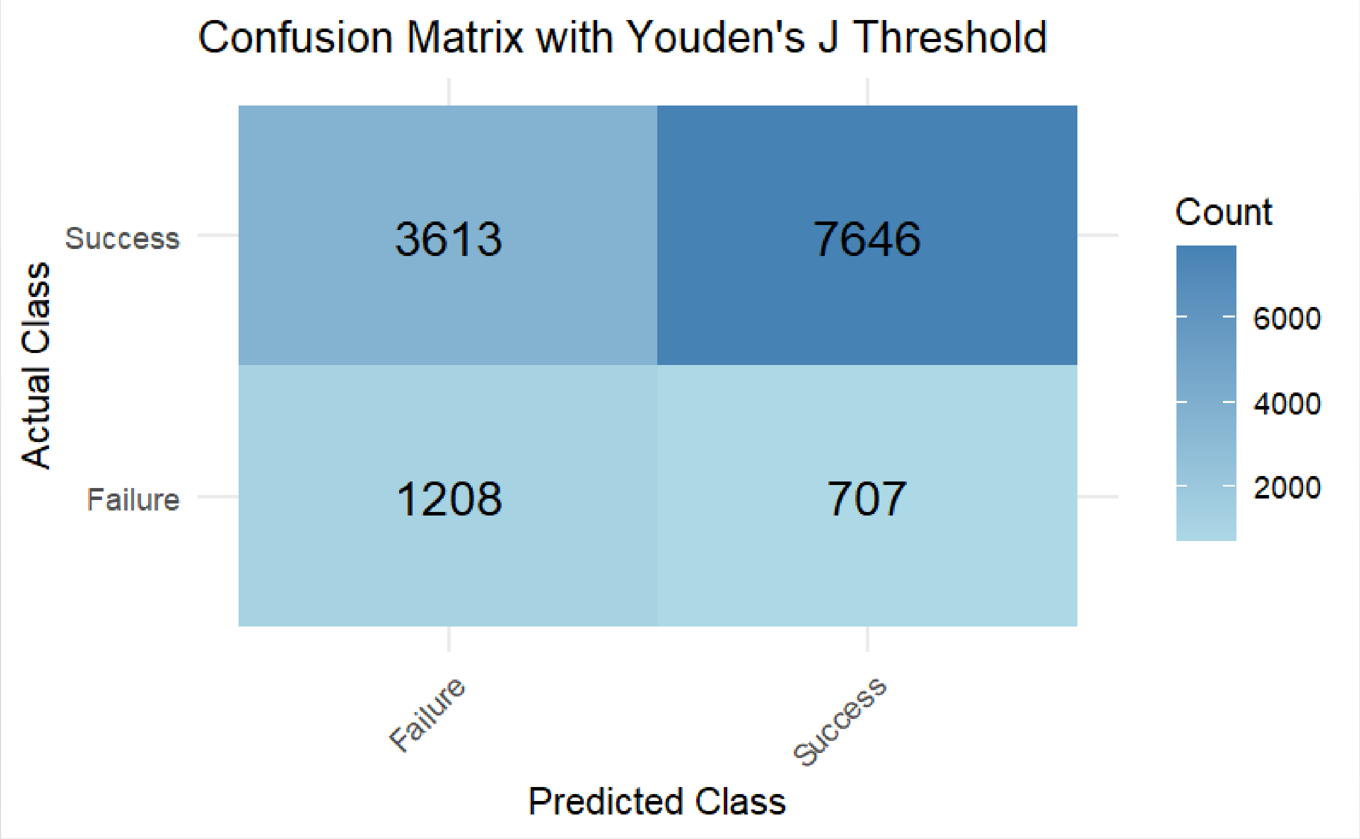
ROC CURVE



AUC=0.70411

OPTIMAL THRESHOLD

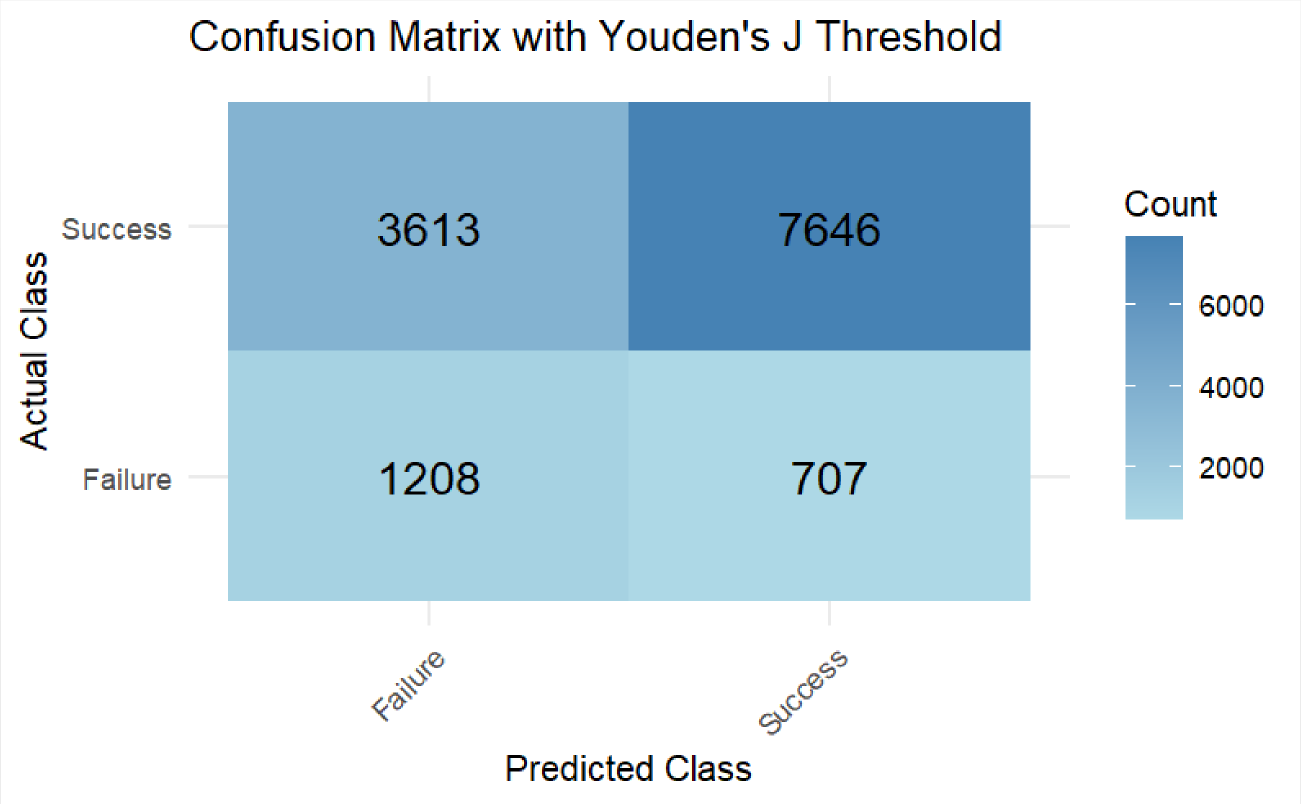
- J statistic threshold: 0.8534
- Accuracy: 0.6721
- 95% CI: [0.664,0.6801]



OPTIMAL THRESHOLD

Performance with Youden's J Threshold (By Class)

Metric	Result
No Information Rate	0.855
P-Value [Acc > NIR]	1
Sensitivity	0.6791
Specificity	0.6308
PPV	0.9154
NPV	0.2506
F1	0.7797
Prevalence	0.8546
Detection Rate	0.5804
Detection Prevalence	0.6341
Balanced Accuracy	0.6550



CONCLUSIONS

CONCLUSIONS - GEOGRAPHY

With N. America as reference level:

- E. Asia, W. Europe, Oceania have zero or negative coeffs.

regionCent.Amer/Carib	2.119**
	(1.066)
regionS.America	0.957***
	(0.216)
regionE.Asia	-0.284
	(0.271)
regionSE. Asia	1.271***
	(0.158)
regionS.Asia	0.758***
	(0.147)
regionCent.Asia	0.026
	(0.367)
regionWest.Europe	0.157
	(0.177)
regionEast.Europe	0.631***
	(0.199)
regionMidEast N.Afr	0.451***
	(0.151)
regionSub-Saharan Afr	1.134***
	(0.165)
regionAustr/Oceania	0.282
	(0.689)

CONCLUSIONS - GEOGRAPHY

With N. America as reference level:

- SE. Asia, Carribean, Sub-S. Africa have substantially positive (>1.5) coeffs.

regionCent.Amer/Carib	2.119**
	(1.066)
regionS.America	0.957***
	(0.216)
regionE.Asia	-0.284
	(0.271)
regionSE. Asia	1.271***
	(0.158)
regionS.Asia	0.758***
	(0.147)
regionCent.Asia	0.026
	(0.367)
regionWest.Europe	0.157
	(0.177)
regionEast.Europe	0.631***
	(0.199)
regionMidEast N.Afr	0.451***
	(0.151)
regionSub-Saharan Afr	1.134***
	(0.165)
regionAustr/Oceania	0.282
	(0.689)

CONCLUSIONS - GEOGRAPHY

Observation:

Regions w. higher economic development scores have decreased log odds of a successful terrorist attack.

Conclusion:

Countries dedicating budget spending to military and police as well as security and surveillance may hinder the efforts of terrorism.

regionCent.Amer/Carib	2.119**
	(1.066)
regionS.America	0.957***
	(0.216)
regionE.Asia	-0.284
	(0.271)
regionSE. Asia	1.271***
	(0.158)
regionS.Asia	0.758***
	(0.147)
regionCent.Asia	0.026
	(0.367)
regionWest.Europe	0.157
	(0.177)
regionEast.Europe	0.631***
	(0.199)
regionMidEast N.Afr	0.451***
	(0.151)
regionSub-Saharan Afr	1.134***
	(0.165)
regionAustr/Oceania	0.282
	(0.689)

CONCLUSIONS - TARGET

With Business as reference level:

- All statistically significant groups are either state-owned targets or militarized in some capacity (excl. Abortion).

targtype1Police	-0.273**
	(0.118)
targtype1Military	0.062
	(0.158)
targtype1Abortion Related	-1.499**
	(0.599)
targtype1Airports	-1.890***
	(0.326)
targtype1Diplomatic	-0.850***
	(0.196)
targtype1Educational	-0.152
	(0.202)
targtype1Food/Water Supply	10.107
	(128.842)
targtype1Journalists	-0.051
	(0.188)
targtype1Maritime	-1.453***
	(0.496)
targtype1NGO	0.227
	(0.393)
targtype1Private Citizens	-0.142
	(0.117)
targtype1Religious	-0.184
	(0.151)
targtype1Telecommunication	1.749*
	(1.016)
targtype1Terrorists	0.681**
	(0.266)
targtype1Tourists	1.115
	(1.028)
targtype1Transportation	-1.127***
	(0.157)
targtype1Government	-0.491***
	(0.119)

CONCLUSIONS - TARGET

With Business as reference level:

- All but 1 of the statistically significant groups have negative coefficients.

targtype1Police	-0.273**
	(0.118)
targtype1Military	0.062
	(0.158)
targtype1Abortion Related	-1.499**
	(0.599)
targtype1Airports	-1.890***
	(0.326)
targtype1Diplomatic	-0.850***
	(0.196)
targtype1Educational	-0.152
	(0.202)
targtype1Food/Water Supply	10.107
	(128.842)
targtype1Journalists	-0.051
	(0.188)
targtype1Maritime	-1.453***
	(0.496)
targtype1NGO	0.227
	(0.393)
targtype1Private Citizens	-0.142
	(0.117)
targtype1Religious	-0.184
	(0.151)
targtype1Telecommunication	1.749*
	(1.016)
targtype1Terrorists	0.681**
	(0.266)
targtype1Tourists	1.115
	(1.028)
targtype1Transportation	-1.127***
	(0.157)
targtype1Government	-0.491***
	(0.119)

CONCLUSIONS - TARGET

Observation:

Government/state targets have significantly decreased log odds of a successful attack.

Conclusion:

Government/state facilities (airports, military bases, ports) have significantly stronger security and faster response times than civilian targets, leading to less dangerous (successful) attacks.

targtype1Police	-0.273**
	(0.118)
targtype1Military	0.062
	(0.158)
targtype1Abortion Related	-1.499**
	(0.599)
targtype1Airports	-1.890***
	(0.326)
targtype1Diplomatic	-0.850***
	(0.196)
targtype1Educational	-0.152
	(0.202)
targtype1Food/Water Supply	10.107
	(128.842)
targtype1Journalists	-0.051
	(0.188)
targtype1Maritime	-1.453***
	(0.496)
targtype1NGO	0.227
	(0.393)
targtype1Private Citizens	-0.142
	(0.117)
targtype1Religious	-0.184
	(0.151)
targtype1Telecommunication	1.749*
	(1.016)
targtype1Terrorists	0.681**
	(0.266)
targtype1Tourists	1.115
	(1.028)
targtype1Transportation	-1.127***
	(0.157)
targtype1Government	-0.491***
	(0.119)

CONCLUSIONS - MODELS

- Our best model's accuracy of 0.65 is worse than the NIR of 0.855.
- In context, the NIR may not be applicable to this problem as a solution.
- This dataset has a very high class imbalance.
  - ~6:1
- Predicting every incident to be a success (high level threat) may strain resources thin.
- Our model may not be perfect, but dismissing it because of NIR ignores the context of the issue at hand.

Performance with Youden's J Threshold (By Class)

Metric	Result
No Information Rate	0.855
P-Value [Acc > NIR]	1
Sensitivity	0.6791
Specificity	0.6308
PPV	0.9154
NPV	0.2506
F1	0.7797
Prevalence	0.8546
Detection Rate	0.5804
Detection Prevalence	0.6341
Balanced Accuracy	0.6550

LIMITATIONS – J STATISTIC

Controversy surrounding models using the J-Statistic:

- Balances the overall predictive accuracy of the model.
- May smooth some metrics at the cost of other performance metrics.
- This balance is especially critical in national security applications, where the cost of false negatives (i.e., failing to predict a successful attack) is significantly greater than the cost of false positives.

CV Log Reg

Sensitivity	0.997
Specificity	0.017
PPV	0.856
NPV	0.485

J-Statistic Model

Sensitivity	0.6791
Specificity	0.6308
PPV	0.9154
NPV	0.2506