Who could be the Attacker?

Identifying the Terrorist Organization responsible for a Terrorist Attack
by using Machine Learning Technology

Content

About Me

LI ZHEMING

Job Title
Data Analyst Trainee
@NTUC LearningHub

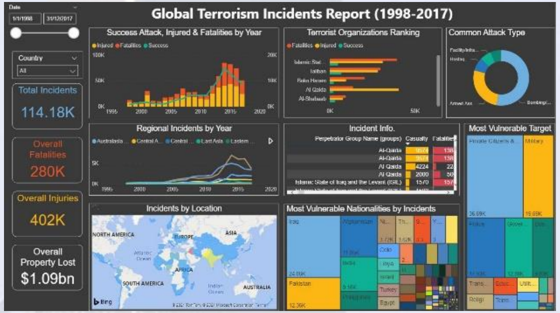
Education
Bachelor of Engineering

Skills

scikit learn, python, Power BI, SQL, Excel

Portfolio Projects

1. **Global Terrorism Incidents Report (1998-2017)**
   Data Analysis and Visualization by Power BI
2. **Power Infrastructure Market Analysis in Asia**
   Interactive Dashboard (Excel & Power Query)
3. **FIFA World Cup Analysis 1930-2014**
   Create Relational Database & Analyze Data with SQL (SQL & Excel)



Representing Organization

Consulting firms

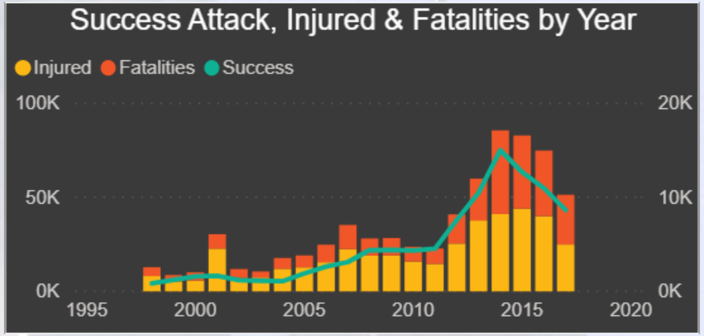
Target Audience

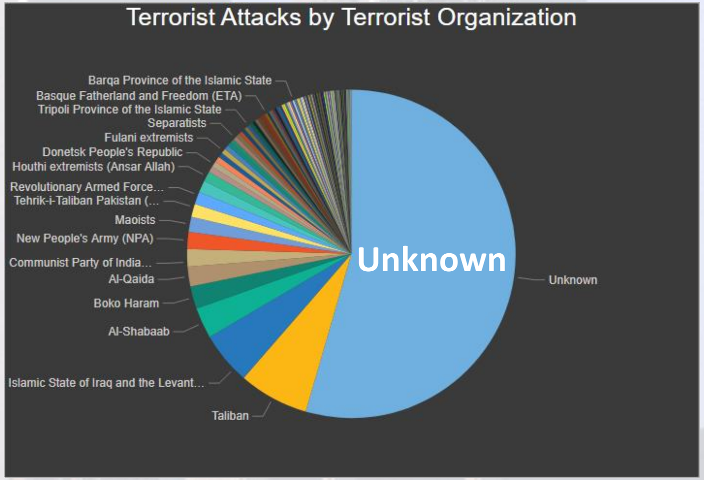
Domestic Security Department

Problem Statement

- In the recent 20 years, Terrorist Attacks show an actively up-trending growth. It's getting threatening.
- Less than 50% of the Attacks are Identifiable for the attackers. We are short of knowledge of these attackers.

- If we could identify the Attacker from the features of an **on-going** Terrorist Attack, we are able to retrieve the Attacker's behavior pattern from historical records and react to it accordingly.
- Furthermore, if we could identify the Attacker of **historical** unknown attacks, it could facilitate the decision making in preventing future Terrorist Attacks.
- **Goal**=> Build a Machine Learning Model to identify the Attacker of a Terrorist Attack as accurate as possible. => Classification Question





Data Collection

Sources:

- *Global Terrorism Database* : https://start.umd.edu/gtd/

Data Collection

Sources:

GTD - *Global Terrorism Database* : https://start.umd.edu/gtd/

1. **Organization**

   National Consortium for the Study of Terrorism and Responses to Terrorism (START) at the University of Maryland

2. **Time period**

   1970-2017, except 1993.

   High Information Integrity from 1998 to 2017.

3. **Contents**

   – 200,000+ **records**

   – 135 **features**

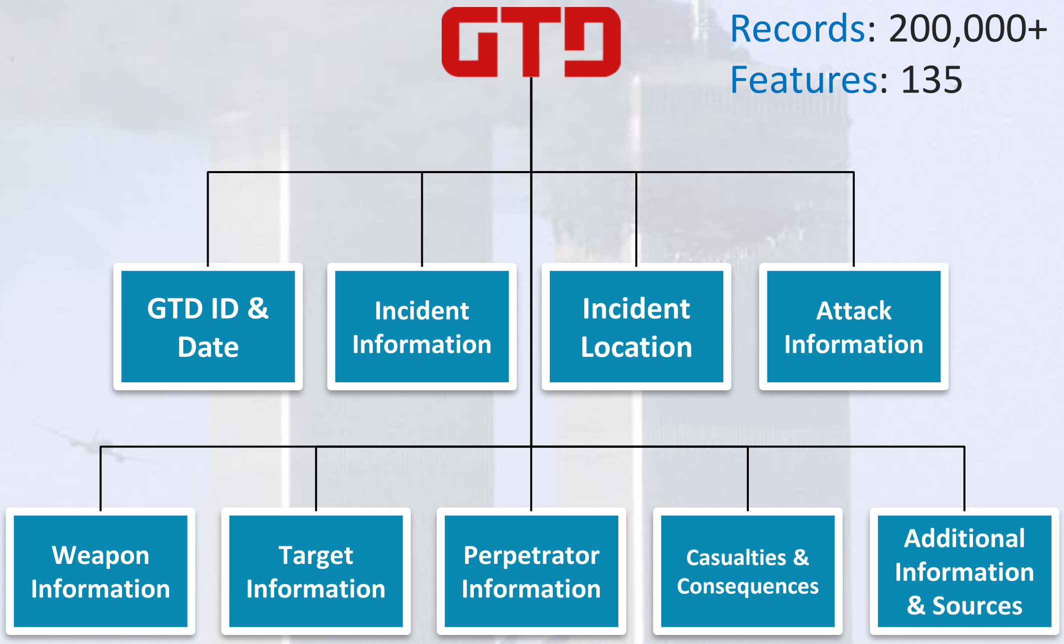
4. **Sources**

   Unclassified media articles

5. **Codebook**

   65 pages

   Reflecting the systematic collection and coding rules for the Global Terrorism Database



1. MODEL

Baseline Model: Logistic Regression

Alternative Model: Support Vector Machine(SVM), Random Forest, Naive Bayes

2. METRICS

Classification Model

⇒ Classification Report(precision,recall,f1-score)

⇒ Confusion Matrix

⇒ ROC Curve

3. TOOLS

scikit learn

pandas

NumPy

jupyter

matplotlib

seaborn

Process Workflow

1. EDA & Data Preparation
2. Data Analysis & Feature Engineering
3. Data Transformation
4. ML Model Training & Evaluation

1. EDA & Data Preparation

Initial Dataset: Global Terrorism Dataset.csv

1. **Choose High Information Integrity Records**

   High Information Integrity from **1998-2017**.

2. **Pick Top 10 Terrorist Organization**

   Except for Unknown Attackers, The top 10 Terrorist organizations have committed over 50% of attacks globally.

   Subset : **Remove** records with **Unknow** Attackers, **filter out** the records with **top 10** attackers

3. **Remove Irrelevant Features**

   Descriptive text features

4. **Remove Low Information Integrity Features**

Features 135

Records 200,000+

Records 114,000+

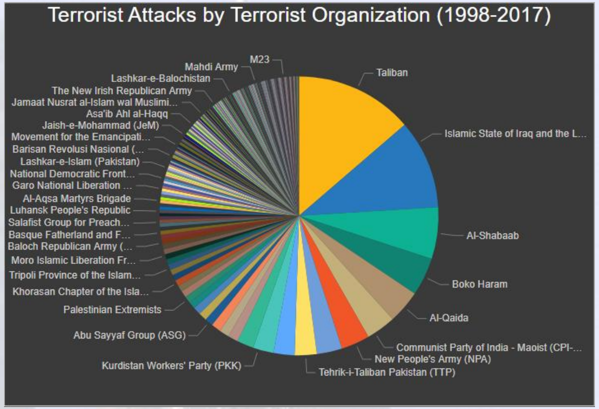
Records 14,000+

Features 62

Features 27

Records 12,000+

ML Dataset



summary
01/01/1998: Hutu Rebels attacked a Burundi military target at Bujumb...
01/01/1998: The breakaway Loyalist Volunteer Force (LVF) claimed res...
01/04/1998: A bomb attack occurred on a police station in Kumanovo,...
01/04/1998: A bomb attack occurred on a police station in Prilep, Mac...
01/05/1998: In one of two related incidents, Hutu Rebels attacked the...
01/05/1998: In one of two related incidents, Hutu Rebels attacked the...
01/06/1998: Hutu Rebels attacked a military post in Maramvya, a villa...

51	ransom	1120 non-null	float64
52	ransomamt	53 non-null	float64
53	ransomamtus	46 non-null	float64
54	ransompaid	47 non-null	float64
55	ransompaidus	45 non-null	float64
56	hostkidoutcome	1077 non-null	float64
57	nreleased	1063 non-null	float64

1. EDA & Data Preparation

Initial Dataset: Global Terrorism Dataset.csv

1. **Choose High Information Integrity Records**

   High Information Integrity from **1998-2017**.

2. **Pick Top 10 Terrorist Organization**

   Except for Unknown Attackers, The top 10 Terrorist organizations have committed over 50% of attacks globally.

   Subset : **Remove** records with **Unknow** Attackers, **filter out** the records with **top 10** attackers

3. **Remove Irrelevant Features**

   Descriptive text features

4. **Remove Low Information Integrity Features**

Features 135

Records 200,000+

Records 114,000+

Records 14,000+

Features 62

Features 27

Records 12,000+

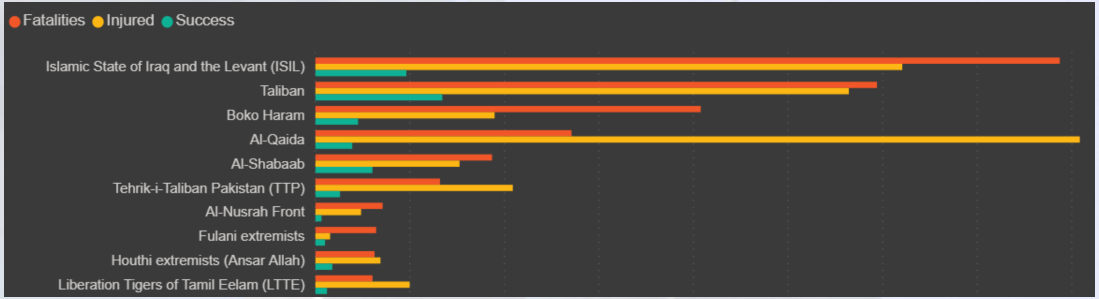
ML Dataset

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12295 entries, 0 to 12294
Data columns (total 27 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   country       12295 non-null  int64
 1   region        12295 non-null  int64
 2   latitude      12295 non-null  float64
 3   longitude     12295 non-null  float64
 4   success       12295 non-null  int64
 5   suicide       12295 non-null  int64
 6   attacktype1   12295 non-null  int64
 7   targtype1     12295 non-null  int64
 8   targsubtype1  12295 non-null  int64
 9   natlty1       12290 non-null  float64
 10  individual    12295 non-null  int64
 11  nperps        11986 non-null  float64
 12  claimed       12295 non-null  int64
 13  weaptype1     12295 non-null  int64
 14  weapsubtype1  10765 non-null  float64
 15  nkill         12295 non-null  int64
 16  nkillus       12295 non-null  int64
 17  nkillter      12295 non-null  int64
 18  nwound        12295 non-null  int64
 19  nwoundus      12295 non-null  int64
 20  nwoundte      12295 non-null  int64
 21  property      12295 non-null  int64
 22  ishostkid     12295 non-null  int64
 23  INT_LOG       12295 non-null  int64
 24  INT_IDEO      12295 non-null  int64
 25  INT_MISC      12295 non-null  int64
 26  INT_ANY       12295 non-null  int64
dtypes: float64(5), int64(22)
memory usage: 2.5 MB
```

1. EDA & Data Preparation

5. Map Text Labels into Numerical Numbers

gname: Name of Terrorist Organization



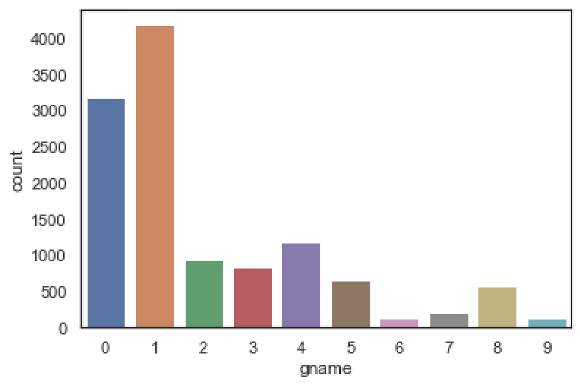

```
gname     10
dtype:  int64
```

```
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
```

6. Explore Label Distribution

This is an **imbalanced** dataset

```
1     4192
0     3181
4     1180
2      939
3      829
5      646
8      566
7      206
6      122
9      120
Name:  gname,  dtype:  int64
```



1. EDA & Data Preparation

7. Drop/Fill up N.A.

```
country           0
region            0
latitude          0
longitude         0
success           0
suicide           0
attacktype1       0
targtype1         0
targsubtype1      0
natlty1           5
individual        0
nperps          309
claimed           0
weaptype1         0
weapsubtype1   1530
nkill             0
nkillus           0
nkillter          0
nwound            0
nwoundus          0
nwoundte          0
property          0
ishostkid         0
INT_LOG           0
INT_IDEO          0
INT_MISC          0
INT_ANY           0
dtype: int64
```

```
country         0
region          0
latitude        0
longitude       0
success         0
suicide         0
attacktype1     0
targtype1       0
targsubtype1    0
natlty1         0
individual      0
nperps          0
claimed         0
weaptype1       0
weapsubtype1    0
nkill           0
nkillus         0
nkillter        0
nwound          0
nwoundus        0
nwoundte        0
property        0
ishostkid       0
INT_LOG         0
INT_IDEO        0
INT_MISC        0
INT_ANY         0
dtype: int64
```

Number of N.A. in each features

2. Data Analysis & Feature Engineering

Four Questions towards Features

1. **Does this feature represent the characteristics of the Terrorist Organizations?**

   Good features will show **different behavior pattern** among Terrorist Organizations

2. **Is this feature measurable in the early stage?**

   If it is measurable, it could help us to **identify** the **attacker** of an **on-going attack** in the early stage.

3. **Is this feature highly correlated with other features? (multicollinearity)**

   Only one feature from the highly correlated features group will be **selected** to **avoid bias.**

4. **Is this a nominal categorical feature?**

   One-Hot Encoding on a nominal feature with huge values will create tremendous new features and cost extra-ordinary processing resources

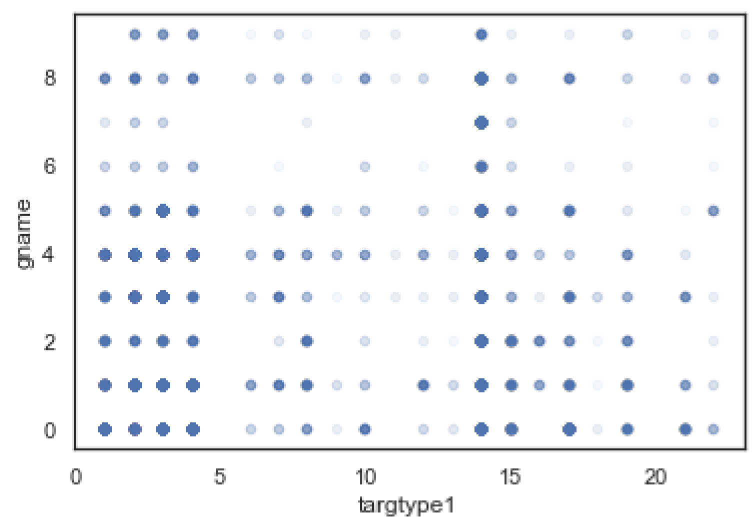
Features
country
region
latitude
longitude
success
suicide
attacktype1
targtype1
targsubtype1
natlty1
individual
nperps
claimed
weaptype1
weapsubtype1
nkill
nkillus
nkillter
nwound
nwoundus
nwoundte
property
ishostkid
INT_LOG
INT_IDEO
INT_MISC
INT_ANY

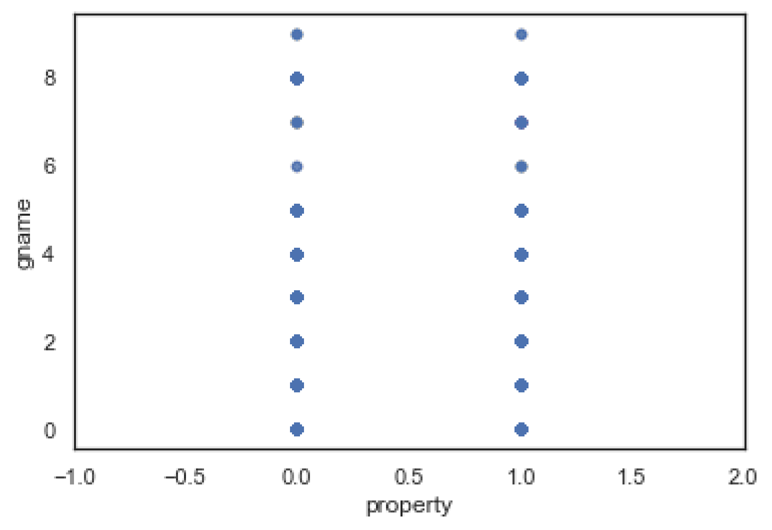
2. Data Analysis & Feature Engineering

1. Does this feature represent the characteristics of the Terrorist Organizations?

Good features will show **different behavior pattern** among Terrorist Organizations

Example:





Good Feature: Target/Victim (nominal)

Bad Feature: Is Property Damage? (binary)

2. Data Analysis & Feature Engineering

2. Is this feature measurable in the early stage?

If it is measurable, it could help us to **identify** the **attacker** of an **on-going attack** in the early stage.

Relies on **Domain Knowledge.**

Example:

1. **nperps (Number of Attackers)**

   We might not be able to know the number of attackers of an on-going attack.

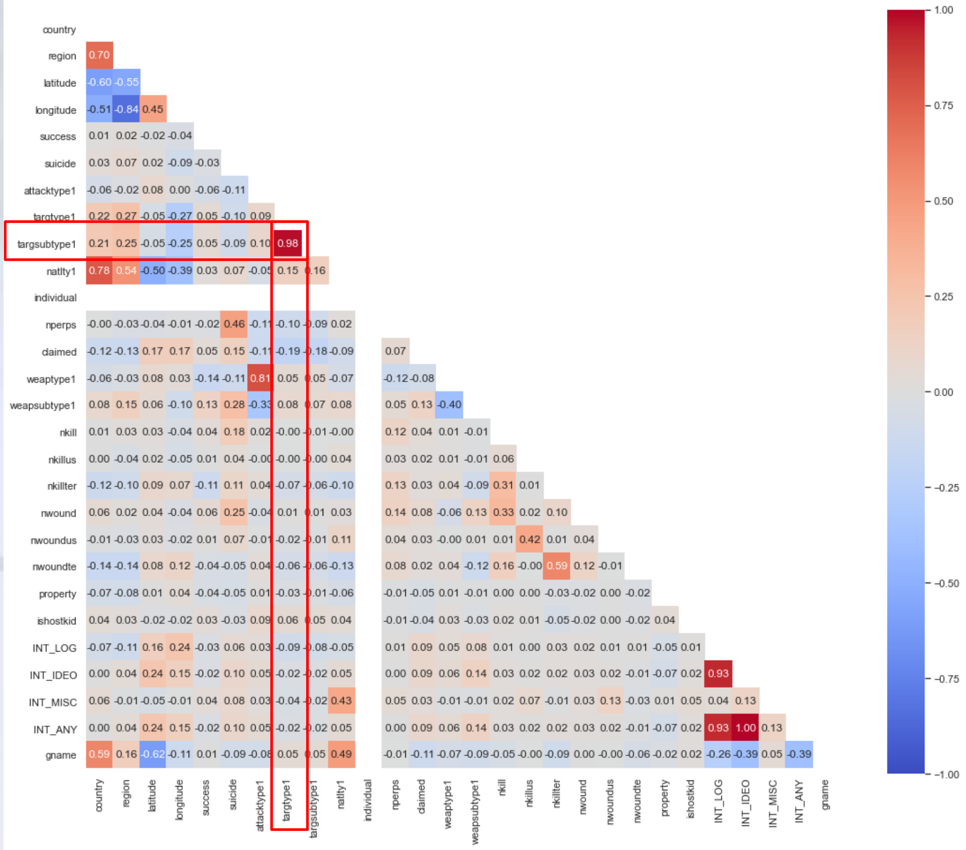
2. **INT_IDEO (International- Ideological)**

   We are not able to know whether the nationalities between attackers and targets are different or not before we identify the attacker

2. Data Analysis & Feature Engineering

3. Is this feature highly correlated with other features? (multicollinearity)

Only one feature from the highly correlated features group will be selected to avoid bias.



2. Data Analysis & Feature Engineering

4. Is this a nominal categorical feature?

One-Hot Encoding on a nominal feature with huge values will create tremendous new features and cost extra-ordinary processing resources

Example:

Attack Type Hierarchy:
- Assassination
- Hijacking
- Kidnapping
- Barricade Incident
- Bombing/Explosion
- Armed Assault
- Unarmed Assault
- Facility/Infrastructure Attack
- Unknown

Good Feature: Attack Type (nominal)

173 = Saudi Arabia
174 = Senegal
175 = Serbia-Montenegro
176 = Seychelles
177 = Sierra Leone
178 = Singapore
179 = Slovak Republic
180 = Slovenia
181 = Solomon Islands
182 = Somalia
183 = South Africa
184 = South Korea
185 = Spain
186 = Sri Lanka
189 = St. Kitts and Nevis
190 = St. Lucia
192 = St. Martin*
195 = Sudan
196 = Suriname
197 = Swaziland
198 = Sweden
199 = Switzerland
200 = Syria
201 = Taiwan
202 = Tajikistan
203 = Tanzania
204 = Togo
205 = Thailand
206 = Tonga*
207 = Trinidad and Tobago
208 = Tunisia
209 = Turkey
210 = Turkmenistan
212 = Tuvalu*
213 = Uganda
214 = Ukraine
215 = United Arab Emirates
216 = Great Britain*
217 = United States
218 = Uruguay
219 = Uzbekistan
220 = Vanuatu
221 = Vatican City
222 = Venezuela
223 = Vietnam
225 = Virgin Islands (U.S.)*
226 = Wallis and Futuna
228 = Yemen
229 = Democratic Republic of the Congo
230 = Zambia
231 = Zimbabwe
233 = Northern Ireland*
235 = Yugoslavia
236 = Czechoslovakia
238 = Corsica*
334 = Asian*
347 = East Timor
349 = Western Sahara
351 = Commonwealth of Independent States*
359 = Soviet Union
362 = West Germany (FRG)
377 = North Yemen
403 = Rhodesia
406 = South Yemen
422 = International
428 = South Vietnam
499 = East Germany (GDR)
520 = Sinhalese*
532 = New Hebrides
603 = United Kingdom
604 = Zaire
605 = People's Republic of the Congo
999 = Multinational*
1001 = Serbia
1002 = Montenegro
1003 = Kosovo
1004 = South Sudan

Bad Feature: Country (nominal)

2. Data Analysis & Feature Engineering

Features	Selected	Q1	Q2	Q3	Q4
country		1	1	1	1
region		1	1	1	1
latitude		1	1	0	0
longitude		1	1	0	0
success		1	0	0	0
suicide		1	1	0	0
attacktype1		1	1	0	1
targtype1		1	1	0	1
targsubtype1		1	1	1	1
natlty1		1	1	0	1
individual		0	0	0	0
nperps		1	0	0	0
claimed		1	1	0	0
weaptype1		1	1	1	1
weapsubtype1		1	1	1	1
nkill		1	1	0	0
nkillus		1	1	0	0
nkillter		1	1	1	0
nwound		1	1	0	0
nwoundus		1	1	0	0
nwoundte		1	0	1	0
property		0	1	0	0
ishostkid		1	1	0	0
INT_LOG		1	0	1	0
INT_IDEO		1	0	1	0
INT_MISC		1	0	1	0
INT_ANY		1	0	1	0

Green: This feature will be selected;

Yellow: This feature will be on hold;

Red: This feature will be dropped;

Orange: This is a special feature (explain latter)

Features 27 → Features 13 / Records 12,000+ →

Revised ML Dataset

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12295 entries, 0 to 12294
Data columns (total 13 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   latitude     12295 non-null  float64
 1   longitude    12295 non-null  float64
 2   success      12295 non-null  int64
 3   suicide      12295 non-null  int64
 4   attacktype1  12295 non-null  int64
 5   targtype1    12295 non-null  int64
 6   claimed      12295 non-null  int64
 7   nkill        12295 non-null  int64
 8   nkillus      12295 non-null  int64
 9   nkillter     12295 non-null  int64
 10  nwound       12295 non-null  int64
 11  nwoundus     12295 non-null  int64
 12  ishostkid    12295 non-null  int64
dtypes: float64(2), int64(11)
memory usage: 1.2 MB
```

3. Data Transformation

1. One-Hot Encoding for Nominal Features

Features: attacktype1, targtype1

2. Split the data into training and testing datasets

test_size = 0.2

3. Data Normalization

Standard Scaler :

Logistic Regression, Support Vector Machine(SVM), Decision Tree & Random Forest

Min Max Scaler:

Naive Bayes

Features 13

Features 40

attacktype1_1
attacktype1_2
attacktype1_3
attacktype1_4
attacktype1_5
attacktype1_6
attacktype1_7
attacktype1_8
attacktype1_9

targtype1_1
targtype1_2
targtype1_3
targtype1_4
targtype1_6
targtype1_7
targtype1_8
targtype1_9
targtype1_10
targtype1_11
targtype1_12
targtype1_13
targtype1_14
targtype1_15
targtype1_16
targtype1_17
targtype1_18
targtype1_19
targtype1_21
targtype1_22

4. Machine Learning Model Training & Evaluation

1. MODEL

Baseline Model: Logistic Regression

Alternative Model: Support Vector Machine(SVM), Random Forest, Naive Bayes

4. Machine Learning Model Training & Evaluation

Baseline Model: Logistic Regression

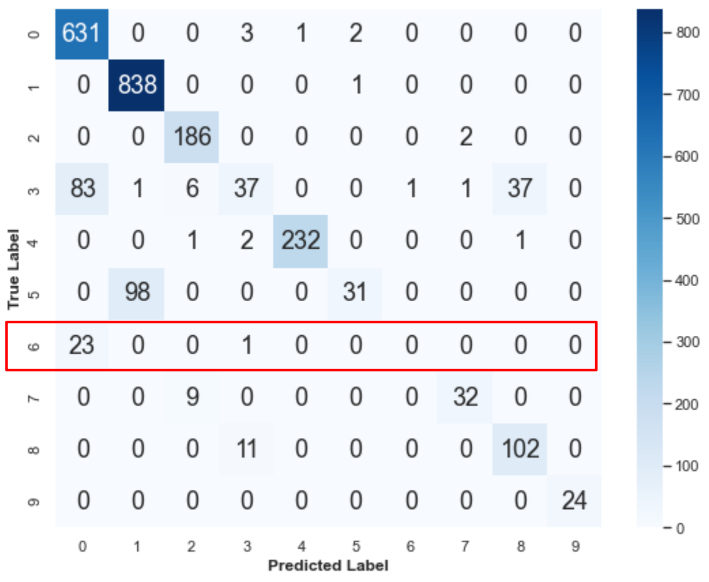
Baseline Model: Logistic Regression

Hyperparameter: (max_iter=5000)

Classification report:

	precision	recall	f1-score	support
0	0.86	0.97	0.92	637
1	0.91	0.99	0.95	880
2	0.90	0.99	0.95	188
3	0.77	0.39	0.52	177
4	1.00	0.98	0.99	236
5	0.80	0.32	0.45	130
6	0.00	0.00	0.00	24
7	0.97	0.78	0.86	41
8	0.80	0.91	0.85	113
9	1.00	0.97	0.98	33
accuracy			0.89	2459
macro avg	0.80	0.73	$\boxed{0.75}$	2459
weighted avg	0.88	0.89	0.87	2459

Classification Report



Confusion Matrix

4. Machine Learning Model Training & Evaluation

Baseline Model: Logistic Regression

Observation

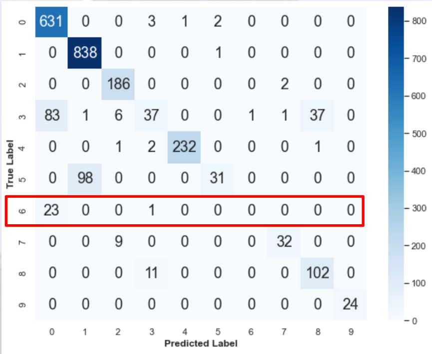
1. Poor performance in classifying Class 3,5
2. Extreme poor performance in classifying Class 6
3. Class 6 tends to be predicted as Class 0

=> Some of the labels with relatively fewer samples are showing poor performance, so **macro f1-score** will be appropriate for this dataset to offset the bias caused by the majority correctly predicted labels.

Classification report:

	precision	recall	f1-score	support
0	0.86	0.97	0.92	637
1	0.91	0.99	0.95	880
2	0.90	0.99	0.95	188
3	0.77	0.39	0.52	177
4	1.00	0.98	0.99	236
5	0.80	0.32	0.45	130
6	0.00	0.00	0.00	24
7	0.97	0.78	0.86	41
8	0.80	0.91	0.85	113
9	1.00	0.97	0.98	33
accuracy			0.89	2459
macro avg	0.80	0.73	0.75	2459
weighted avg	0.88	0.89	0.87	2459

Classification Report



Confusion Matrix

Solution

1. Introduce more features
2. Alternative Model

Introduce more features

Baseline Model: Logistic Regression

Best extra feature: INT_IDEO (the Orange special feature)

Classification report:

	precision	recall	f1-score	support
0	0.98	0.98	0.98	637
1	0.91	0.99	0.95	880
2	0.94	1.00	0.97	188
3	0.86	0.85	0.85	177
4	1.00	0.98	0.99	236
5	0.79	0.32	0.45	130
6	0.70	0.29	0.41	24
7	1.00	0.98	0.99	41
8	0.88	0.92	0.90	113
9	1.00	0.94	0.97	33
accuracy			0.93	2459
macro avg	0.91	0.82	0.85	2459
weighted avg	0.93	0.93	0.92	2459

Classification Report (with INT_IDEO)

Classification report:

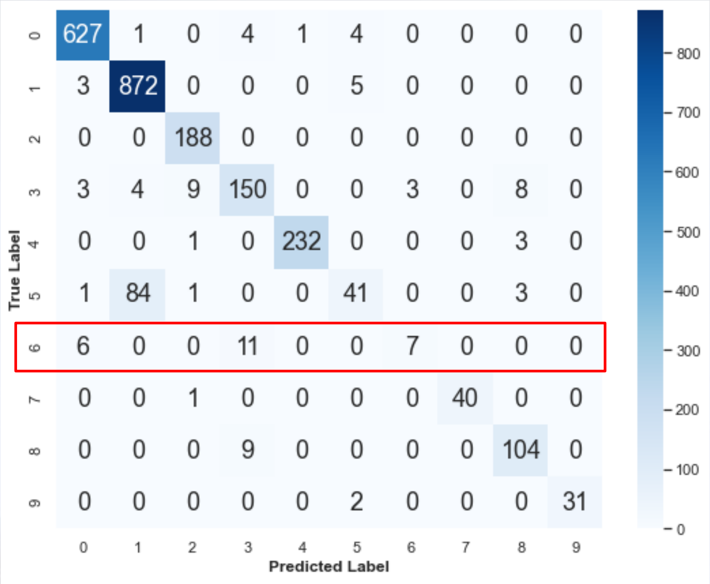
	precision	recall	f1-score	support
0	0.86	0.97	0.92	637
1	0.91	0.99	0.95	880
2	0.90	0.99	0.95	188
3	0.77	0.39	0.52	177
4	1.00	0.98	0.99	236
5	0.80	0.32	0.45	130
6	0.00	0.00	0.00	24
7	0.97	0.78	0.86	41
8	0.80	0.91	0.85	113
9	1.00	0.97	0.98	33
accuracy			0.89	2459
macro avg	0.80	0.73	0.75	2459
weighted avg	0.88	0.89	0.87	2459

Classification Report (Original Model)

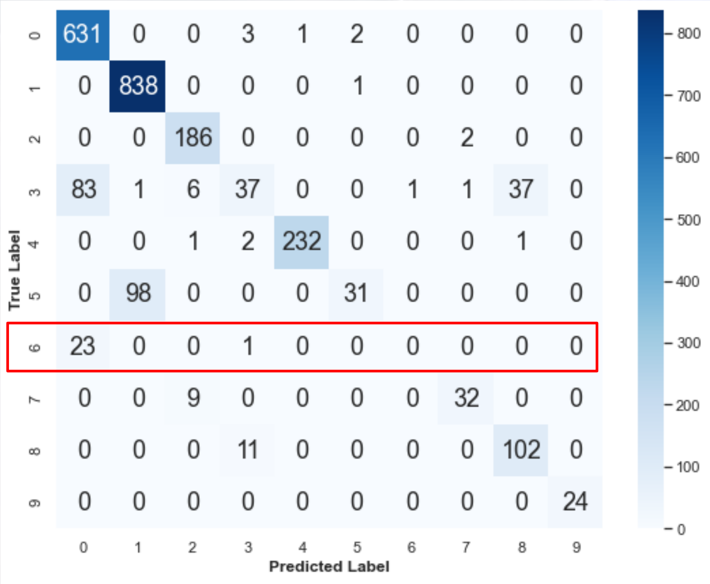
Introduce more features

Baseline Model: Logistic Regression

Best extra feature: INT_IDEO (the Orange special feature)



Confusion Matrix (with INT_IDEO)



Confusion Matrix (Original Model)

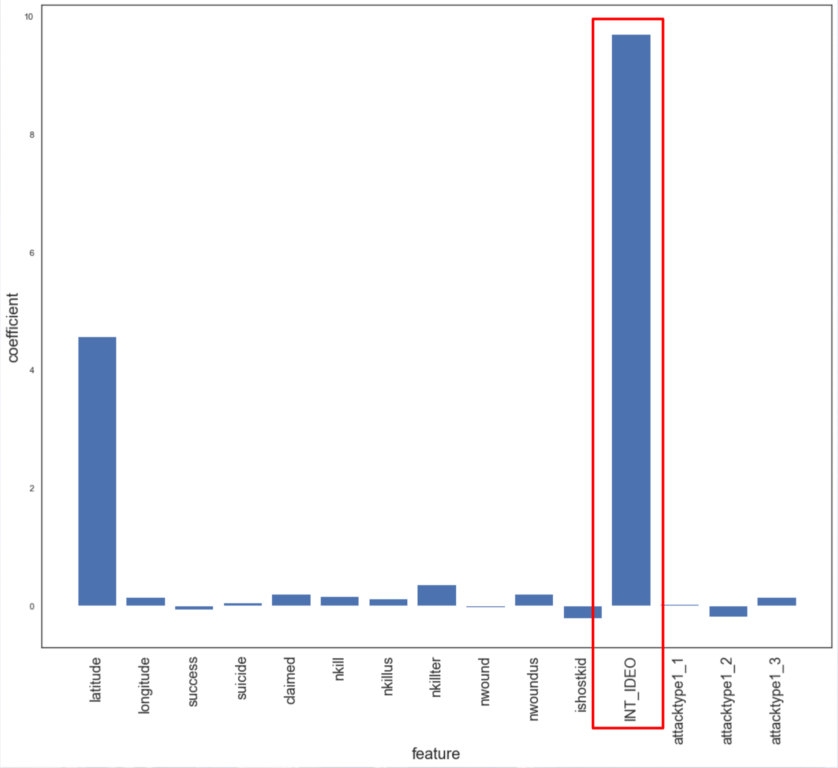
Introduce more features

Baseline Model: Logistic Regression

Best extra feature: INT_IDEO (the Orange special feature)

Observation

1. INT_IDEO plays a strong role in classifying Attackers, especially differentiate Class 6 from Class 0
2. Unfortunately, we couldn't identify this feature's value before we identify the Attackers



Features Coefficient bar chart

Solution

1. ~~Introduce more features~~
2. Alternative Model

4. Machine Learning Model Training & Evaluation

Alternative Model: Support Vector Machine

Alternative Model: Support Vector Machine(SVM)

Best Hyperparameter: (kernel = 'rbf', C = 90, gamma = 0.1668)

Classification report:

	precision	recall	f1-score	support
0	0.87	0.96	0.91	637
1	0.97	0.99	0.98	880
2	0.97	0.98	0.98	188
3	0.80	0.43	0.56	177
4	0.97	0.97	0.97	236
5	0.88	0.78	0.83	130
6	0.46	0.46	0.46	24
7	0.98	0.98	0.98	41
8	0.81	0.89	0.85	113
9	0.94	0.91	0.92	33
accuracy			0.92	2459
macro avg	0.86	0.84	$\boxed{0.84}$	2459
weighted avg	0.91	0.92	0.91	2459

Classification Report

True Label \ Predicted Label	0	1	2	3	4	5	6	7	8	9
0	612	0	0	10	1	2	11	0	0	1
1	3	867	0	0	0	10	0	0	0	0
2	2	0	185	0	0	0	0	1	0	0
3	71	1	3	76	3	1	2	0	20	0
4	1	0	1	0	229	0	0	0	4	1
5	0	26	0	1	1	102	0	0	0	0
6	12	0	0	1	0	0	11	0	0	0
7	0	0	1	0	0	0	0	40	0	0
8	3	0	1	7	1	0	0	0	101	0
9	0	0	0	0	2	1	0	0	0	30

Confusion Matrix

4. Machine Learning Model Training & Evaluation

Alternative Model: Random Forest

Alternative Model: Random Forest

Best Hyperparameter: (criterion = 'gini', max_depth= 27, min_samples_split=5, n_estimators=500)

Classification report:

	precision	recall	f1-score	support
0	0.89	0.97	0.93	637
1	0.98	0.99	0.99	880
2	0.98	1.00	0.99	188
3	0.84	0.56	0.67	177
4	1.00	0.99	0.99	236
5	0.96	0.85	0.90	130
6	0.71	0.42	0.53	24
7	0.97	0.95	0.96	41
8	0.91	0.96	0.93	113
9	0.94	1.00	0.97	33
accuracy			0.94	2459
macro avg	0.92	0.87	0.89	2459
weighted avg	0.94	0.94	0.94	2459

Classification Report

True Label \ Predicted Label	0	1	2	3	4	5	6	7	8	9
0	618	1	0	12	1	0	3	0	0	2
1	0	875	0	0	0	5	0	0	0	0
2	0	0	188	0	0	0	0	0	0	0
3	62	1	2	99	0	0	1	1	11	0
4	0	0	0	2	234	0	0	0	0	0
5	1	19	0	0	0	110	0	0	0	0
6	14	0	0	0	0	0	10	0	0	0
7	0	0	2	0	0	0	0	39	0	0
8	0	0	0	5	0	0	0	0	108	0
9	0	0	0	0	0	0	0	0	0	33

Confusion Matrix

4. Machine Learning Model Training & Evaluation

Alternative Model: Naive Bayes

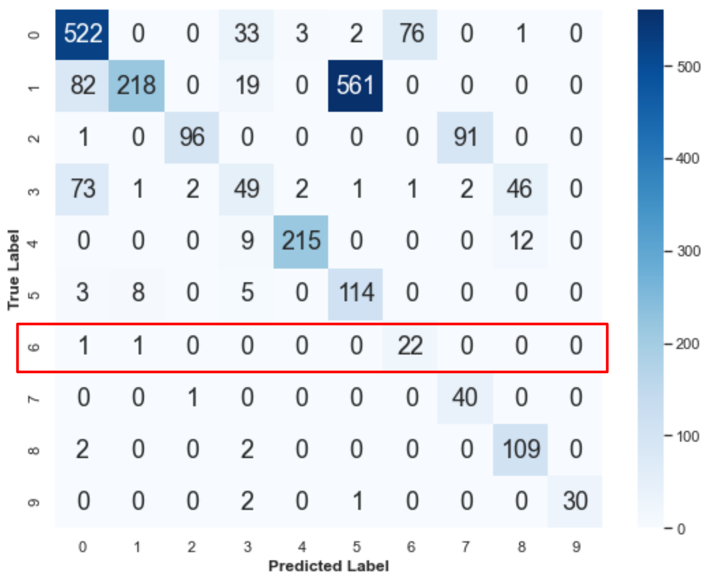
Alternative Model: Naive Bayes

Best Hyperparameter: GaussianNB (var_smoothing = 1.2328467394420658e-05)

Classification report:

	precision	recall	f1-score	support
0	0.76	0.82	0.79	637
1	0.96	0.25	0.39	880
2	0.97	0.51	0.67	188
3	0.41	0.28	0.33	177
4	0.98	0.91	0.94	236
5	0.17	0.88	0.28	130
6	0.22	0.92	0.36	24
7	0.30	0.98	0.46	41
8	0.65	0.96	0.78	113
9	1.00	0.91	0.95	33
accuracy			0.58	2459
macro avg	0.64	0.74	$\boxed{0.60}$	2459
weighted avg	0.80	0.58	0.59	2459

Classification Report

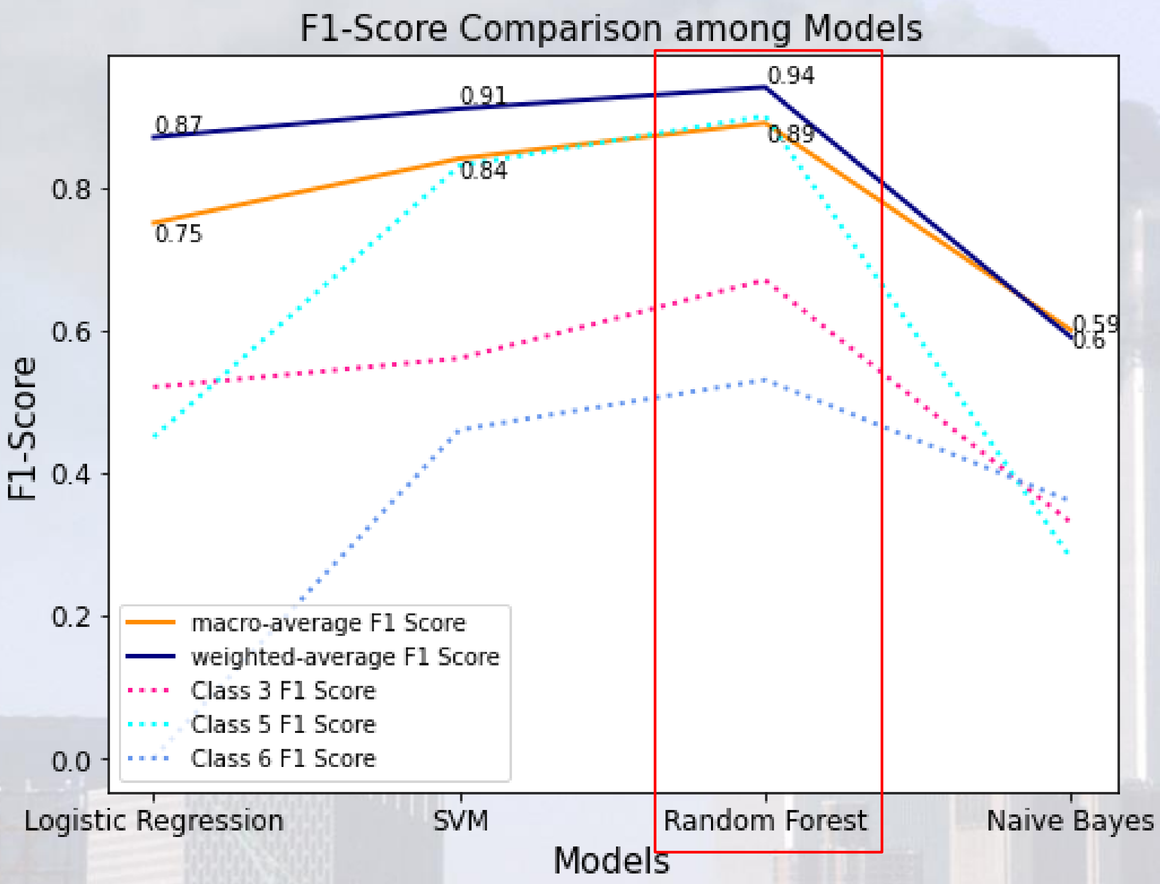


Confusion Matrix

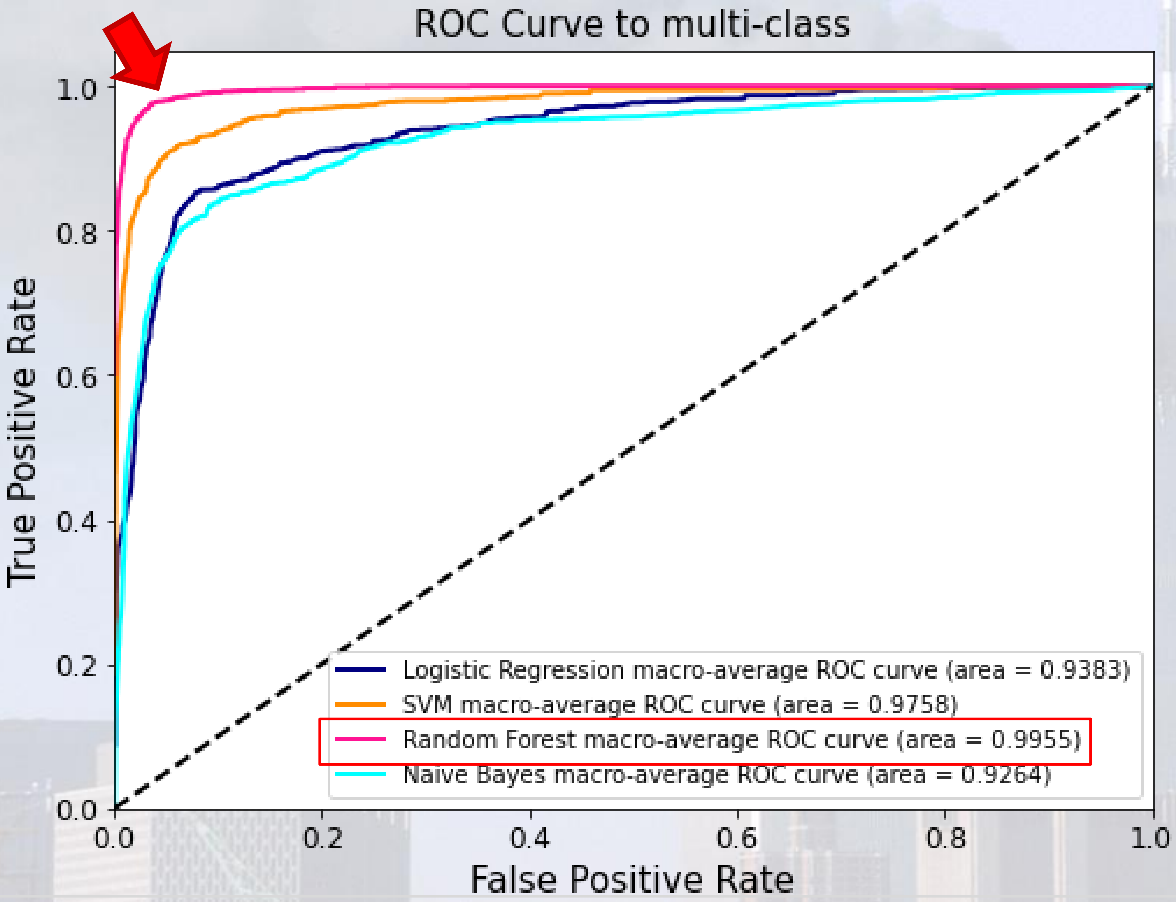
Results

1. F1-Score
2. ROC Curve
3. Class 6

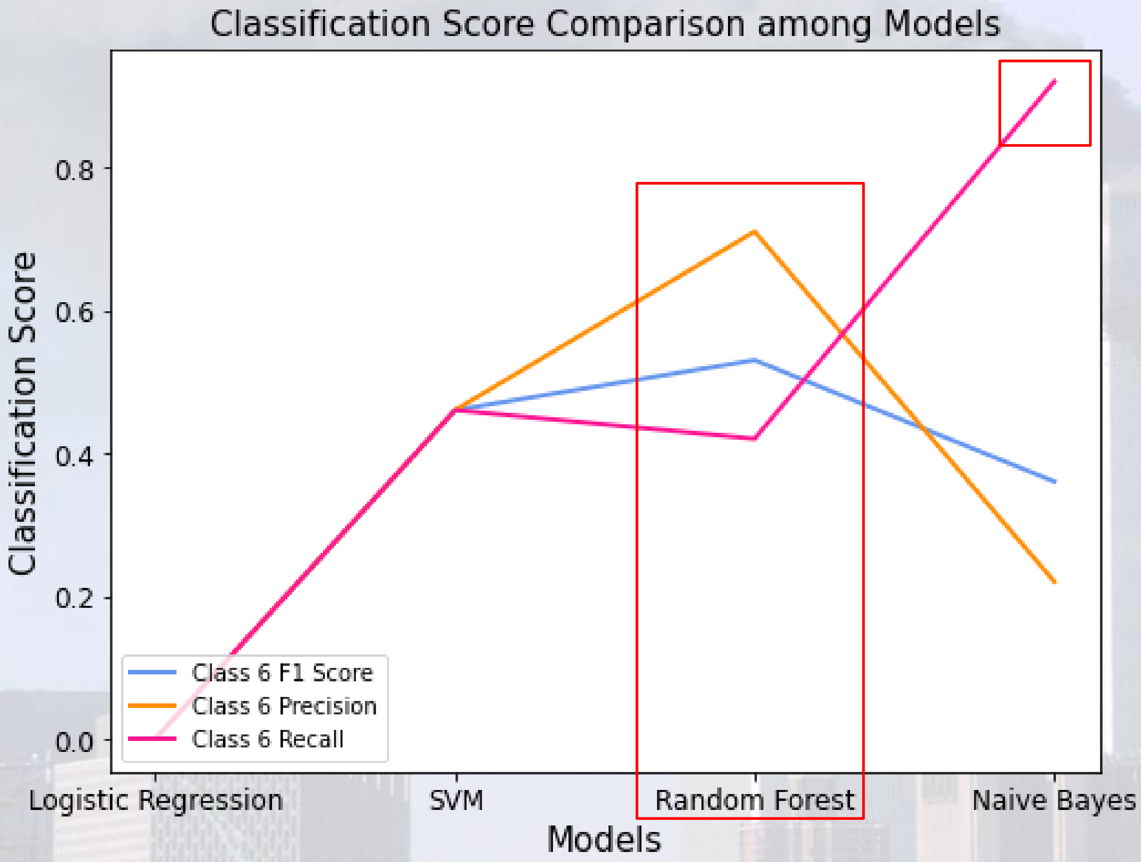
1. F1-Score


F1-Score Comparison among Models
F1-Score
Models
0.0
0.2
0.4
0.6
0.8
Logistic Regression
SVM
Random Forest
Naive Bayes
0.87
0.91
0.94
0.89
0.84
0.75
0.59
0.6
macro-average F1 Score
weighted-average F1 Score
Class 3 F1 Score
Class 5 F1 Score
Class 6 F1 Score

2. ROC Curve


ROC Curve to multi-class
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Logistic Regression macro-average ROC curve (area = 0.9383)
SVM macro-average ROC curve (area = 0.9758)
Random Forest macro-average ROC curve (area = 0.9955)
Naive Bayes macro-average ROC curve (area = 0.9264)

3. Class 6


Classification Score Comparison among Models
Classification Score
0.0
0.2
0.4
0.6
0.8
Class 6 F1 Score
Class 6 Precision
Class 6 Recall
Logistic Regression
SVM
Random Forest
Naive Bayes
Models

Conclusions

Random Forest is the best Model to perform this task:

1. Accuracy : **94%**
2. Macro Average Precision: **92%**
3. Macro Average Recall: **87%**
4. Macro Average F1-Score: **89%**
5. ROC-AUC-Score:**0.9955**

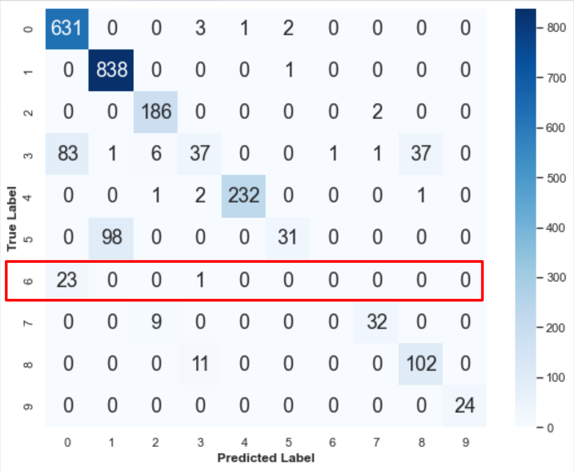
High performance in Identifying Attackers in both **precision** and **recall** rate, which enable **Relevant Department** to take **immediate reaction** against an on-going terrorist attack accordingly.

Classification report:

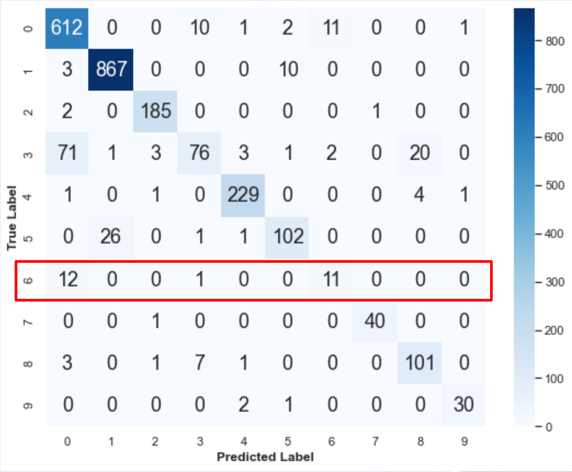
	precision	recall	f1-score	support
0	0.89	0.97	0.93	637
1	0.98	0.99	0.99	880
2	0.98	1.00	0.99	188
3	0.84	0.56	0.67	177
4	1.00	0.99	0.99	236
5	0.96	0.85	0.90	130
6	0.71	0.42	0.53	24
7	0.97	0.95	0.96	41
8	0.91	0.96	0.93	113
9	0.94	1.00	0.97	33
accuracy			0.94	2459
macro avg	0.92	0.87	0.89	2459
weighted avg	0.94	0.94	0.94	2459

Interesting insights

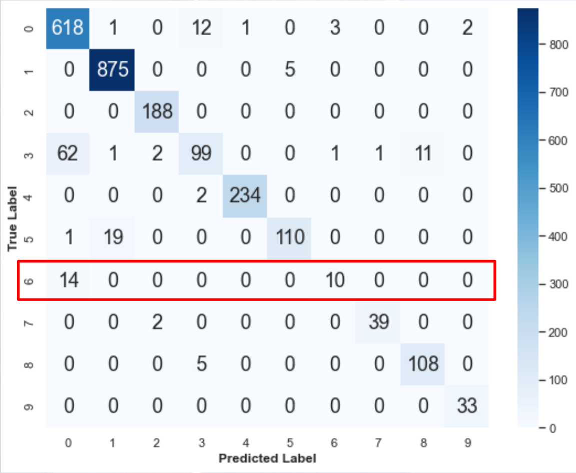
Why Class 6 always trend to be classified as Class 0 in all Models?



Logistic Regression



SVM



Random Forest

Interesting insights

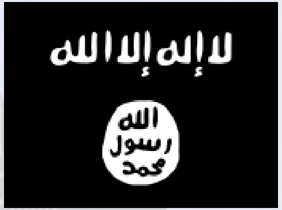
Why Class 6 always trend to be classified as Class 0 in all Models?

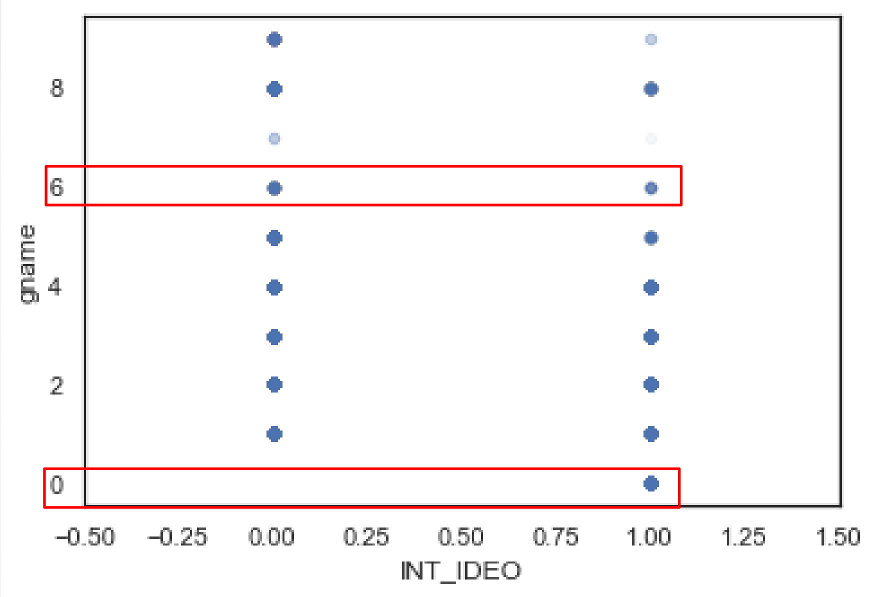
Class 6 : Al-Nusrah Front

Tend to attack both Domestic and International targets.

Class 0: Islamic State of Iraq and the Levant (ISIL)

Tend to focus on Domestic targets





Interesting insights

Why Class 6 always trend to be classified as Class 0 in all Models?

WIKIPEDIA
The Free Encyclopedia

History [edit]

Origin [edit]

Upon the outbreak of the Syrian Civil War in 2011, Islamic State of Iraq's leader Abu Bakr al-Baghdadi and al-Qaeda's central command authorized the Syrian Abu Mohammad al-Golani to set up a Syrian offshoot of al Qaeda in August 2011, to bring down the Assad government and establish an Islamic state there. Golani and six colleagues crossed the border from Iraq into Syria, and reached out to Islamists released from Syria's Sednaya military prison in May–June 2011 who were already active in fighting against Assad's security forces. The six men who founded Nusra alongside Julani were Saleh al-Hamawi (Syrian), Abu Maria Al-Qahtani (Iraqi), Mustafa Abd al-Latif al-Saleh (kunya:Abu Anas al-Sahaba) (Jordanian/Palestinian), Iyad Tubasi (kunya: Abu Julaybib) (Jordanian/Palestinian), Abu Omar al-Filistini (Palestinian) and Anas Hassan Khattab (Syria).[107][46][108]

Future Opportunities

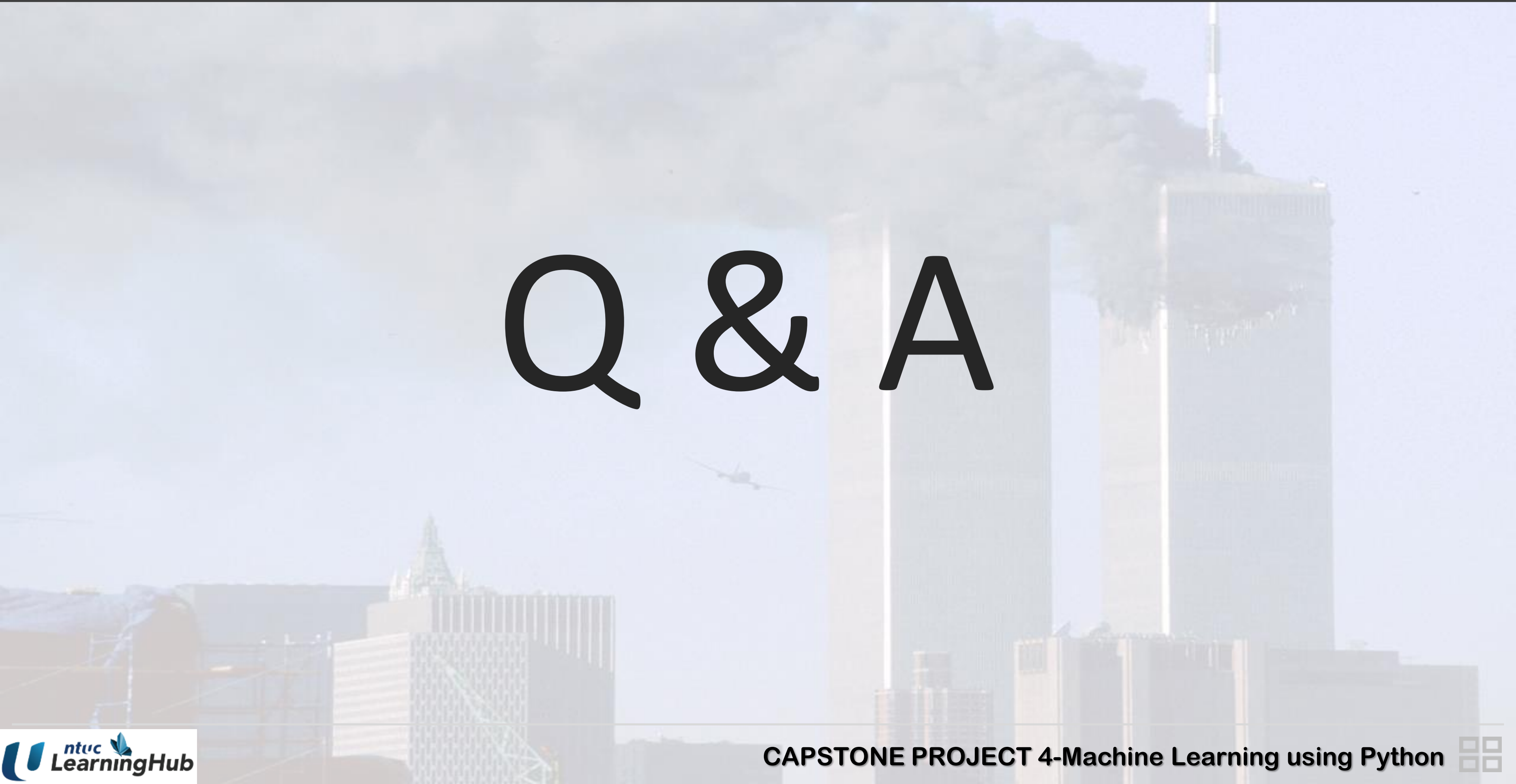
1. **Import more Attacker-known Terrorist Attack records into Machine Learning Dataset**

   To classify more attackers and get closer to real-life situation

2. **Fine Tune Feature Selection**

   Reduce the number of input features with minimum sacrifice of classification performance

3. **Predict potential threatening Attackers**



Q & A

Thank you!

Li Zheming