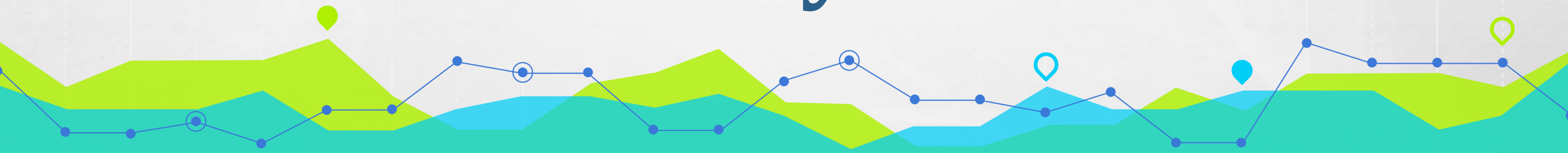
Global Terrorism Analysis



The Confounders

- Harini S.
-Nikita Bhandare
-Nikita Pathak
- Thudcharin W. (Tony)

AGENDA

- ✓ PROJECT GOALS
- ✓ MAIN FINDINGS
- ✓ APPROACH
- ✓ MODEL DESCRIPTION
- ✓ MODEL DETAILS
- ✓ KEY POINTS
- ✓ RECOMMENDATIONS

Situation

- ✓ Government officials focus on capturing the terrorist group and ignore to perceive the long term effect.
- ✓ Terrorist group may falsely or may never claim the responsibility of the attack.

Project Goals

- ✓ Develop a predictive model to determine if the attack is successful.
- ✓ Develop a model to predict the terrorist group based on attack characteristics with an accuracy more than 85%.

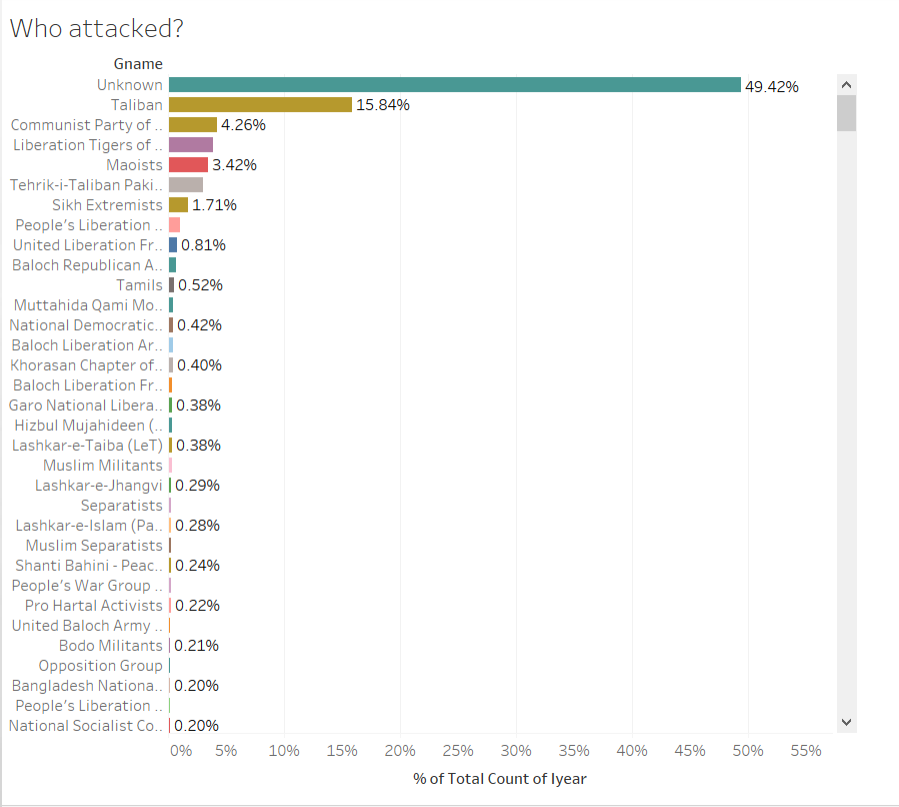
Main Findings

Executing our models can help the police to track the active cells of terrorist groups faster, catch potential perpetrators, help to find real perpetrators and determine the success of the attack.

In the event of an attack, the government can take calculated measures against the attackers based on whether the attack is successful or not.

In South Asia region, 49% of the terrorist attacks are unclaimed.

- Implementing our model reduces the rate of unclaimed attacks by 6.89%.
- Provides an insight to law and order as to who can be behind the attack.



Approach

Dataset Discovery

✓ Global Terrorism Database is an open source database.

-> Maintained by National Consortium for the Study of Terrorism and Response to Terrorism (START).
-> Includes systematic data on domestic as well as international terrorist attacks around the world from 1970 through 2016.

✓ Data is updated annually.

✓ Reviewed the literature provided by START (Study of Terrorism and Response to Terrorism).

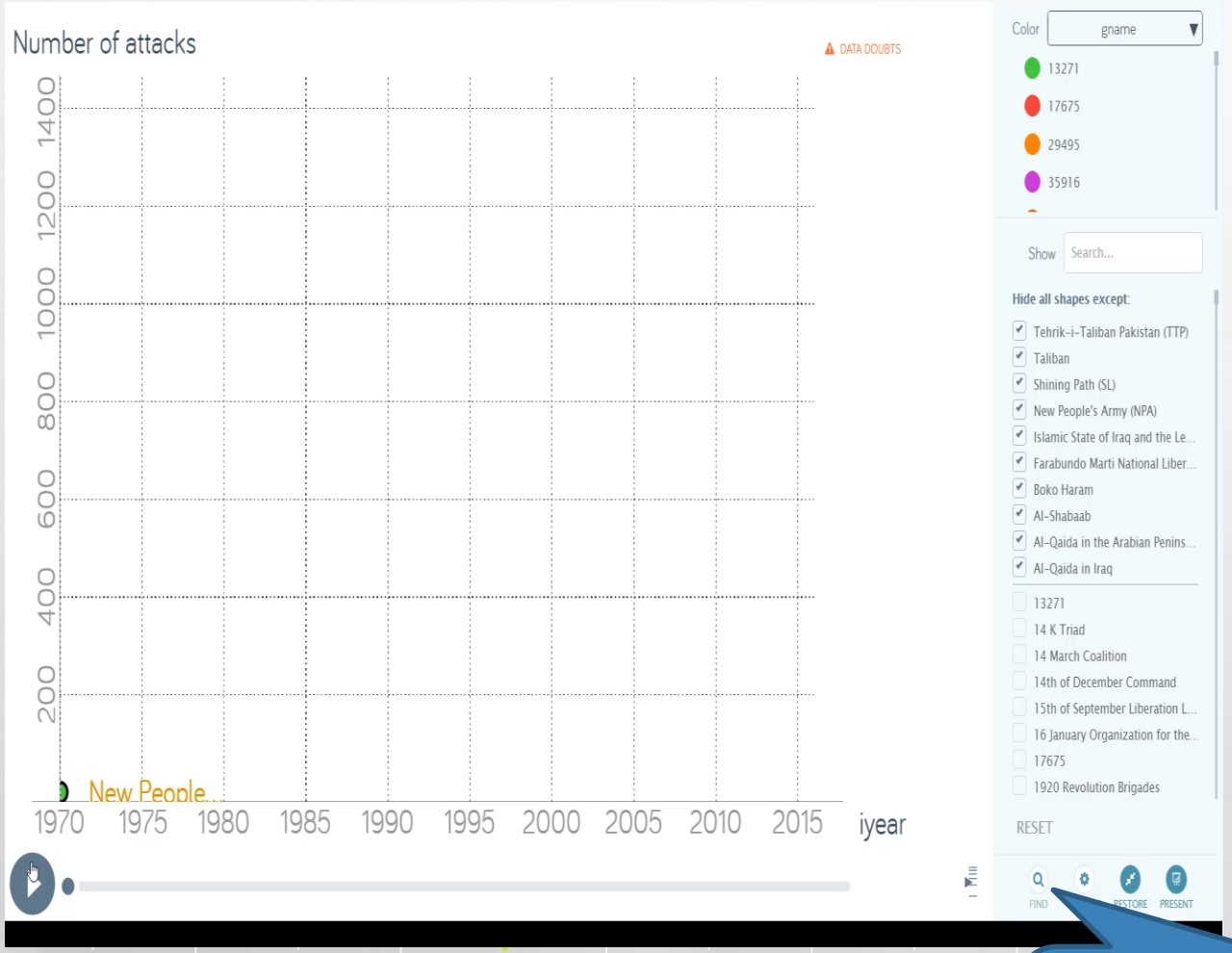
✓ Number of Rows – 170,351 records

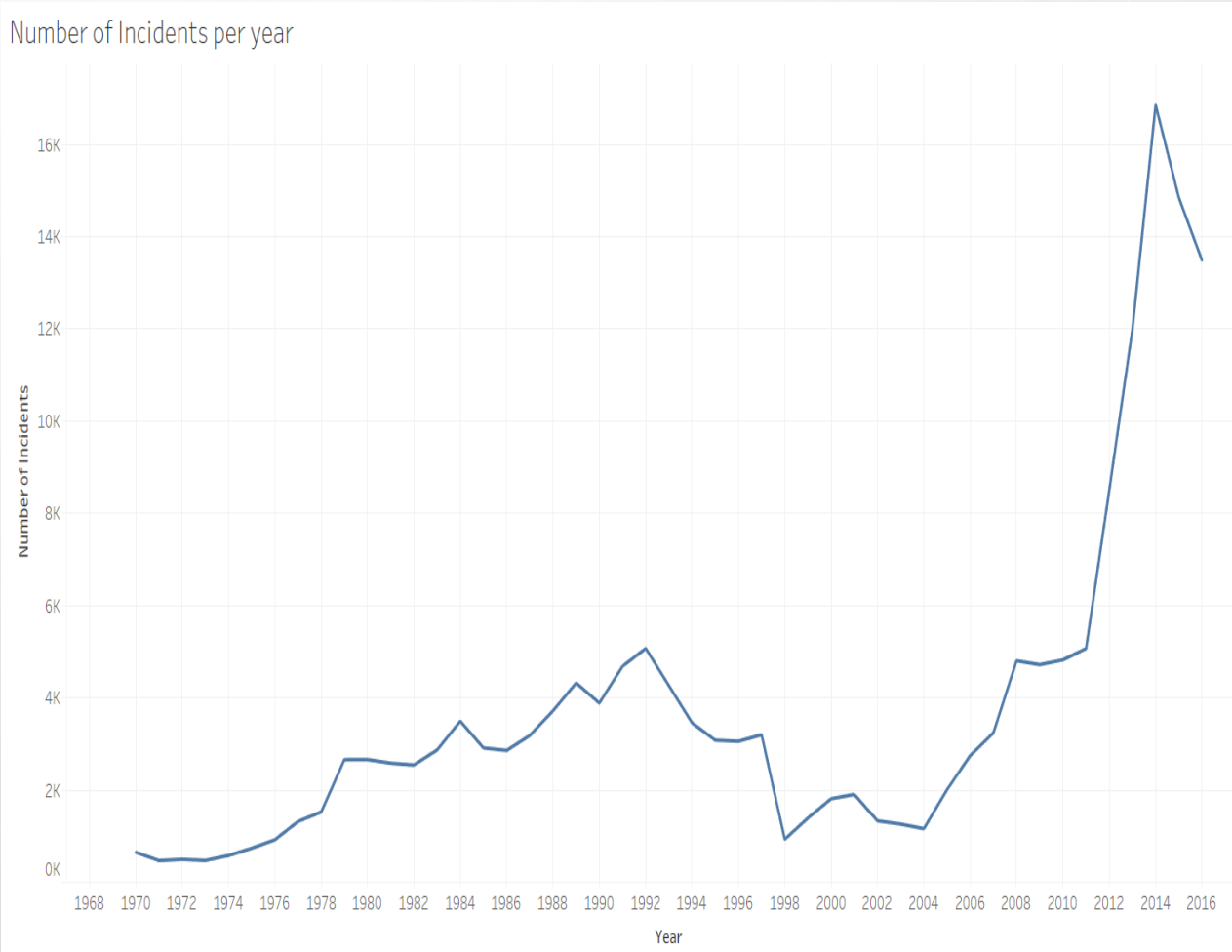
✓ Number of attributes – 135

✓ 1993 data not recorded

Exploratory Analysis

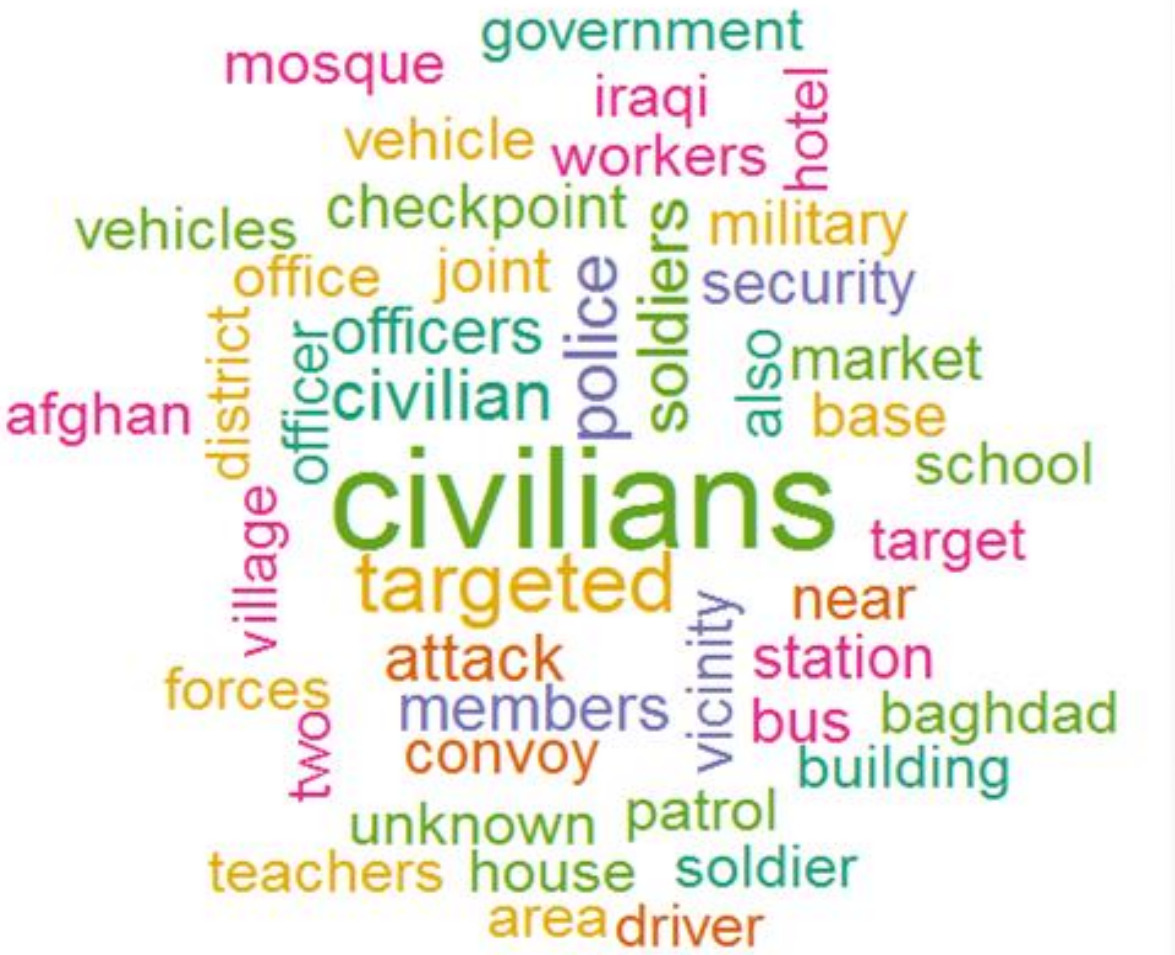
Terrorist Attacks evolving over years



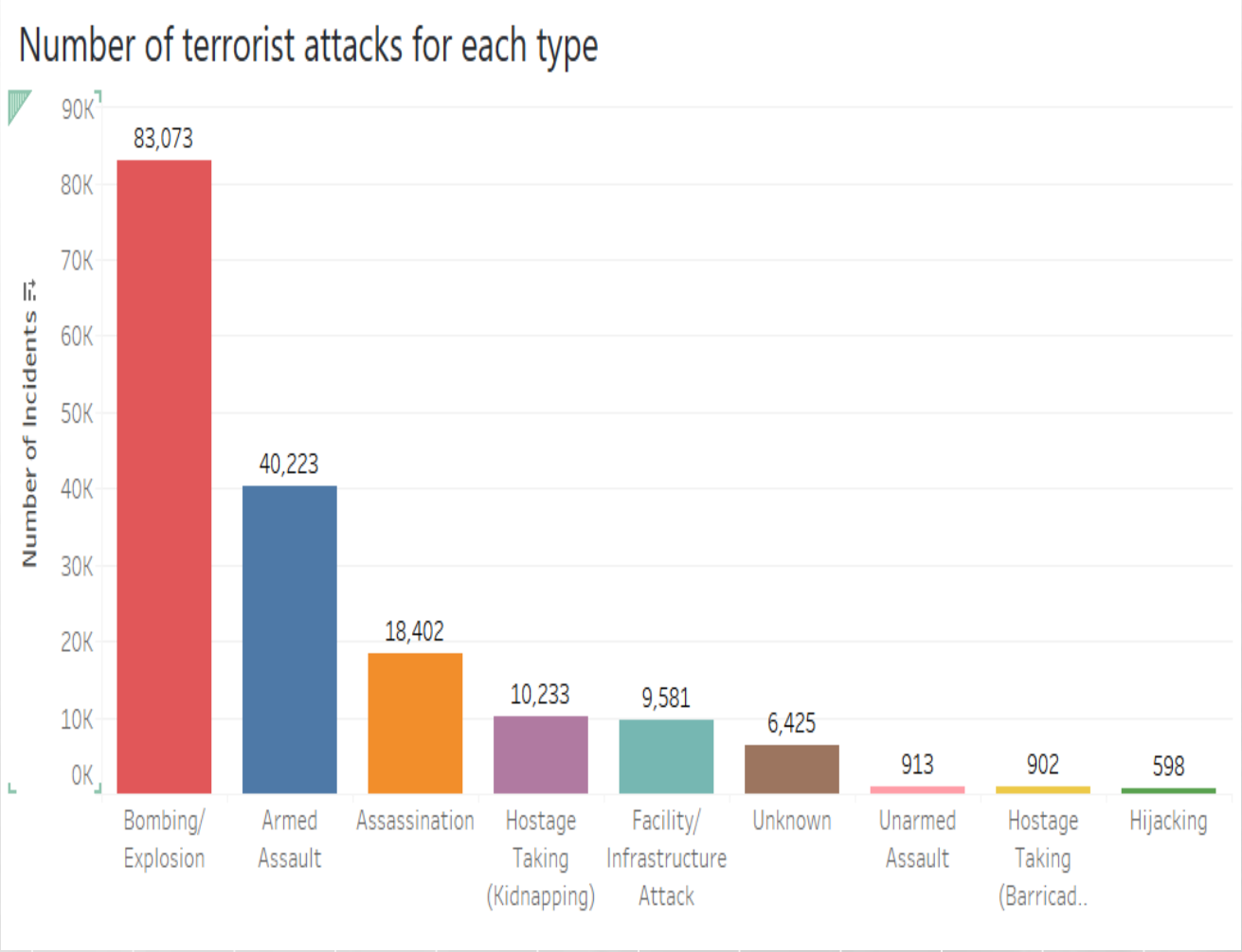


Only a subset of terrorist groups are responsible for majority of attacks

Target of the attack

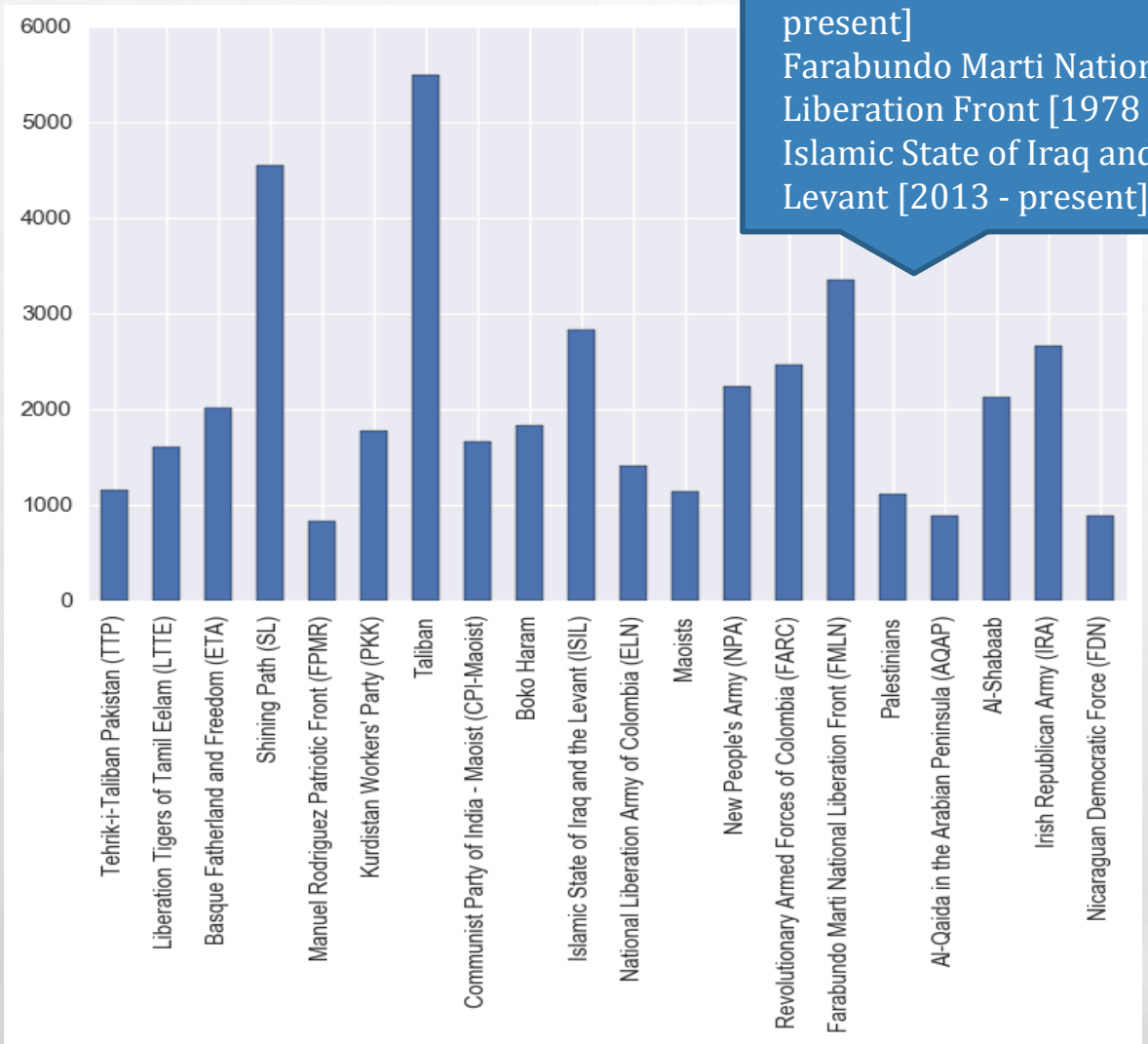


Attack Type

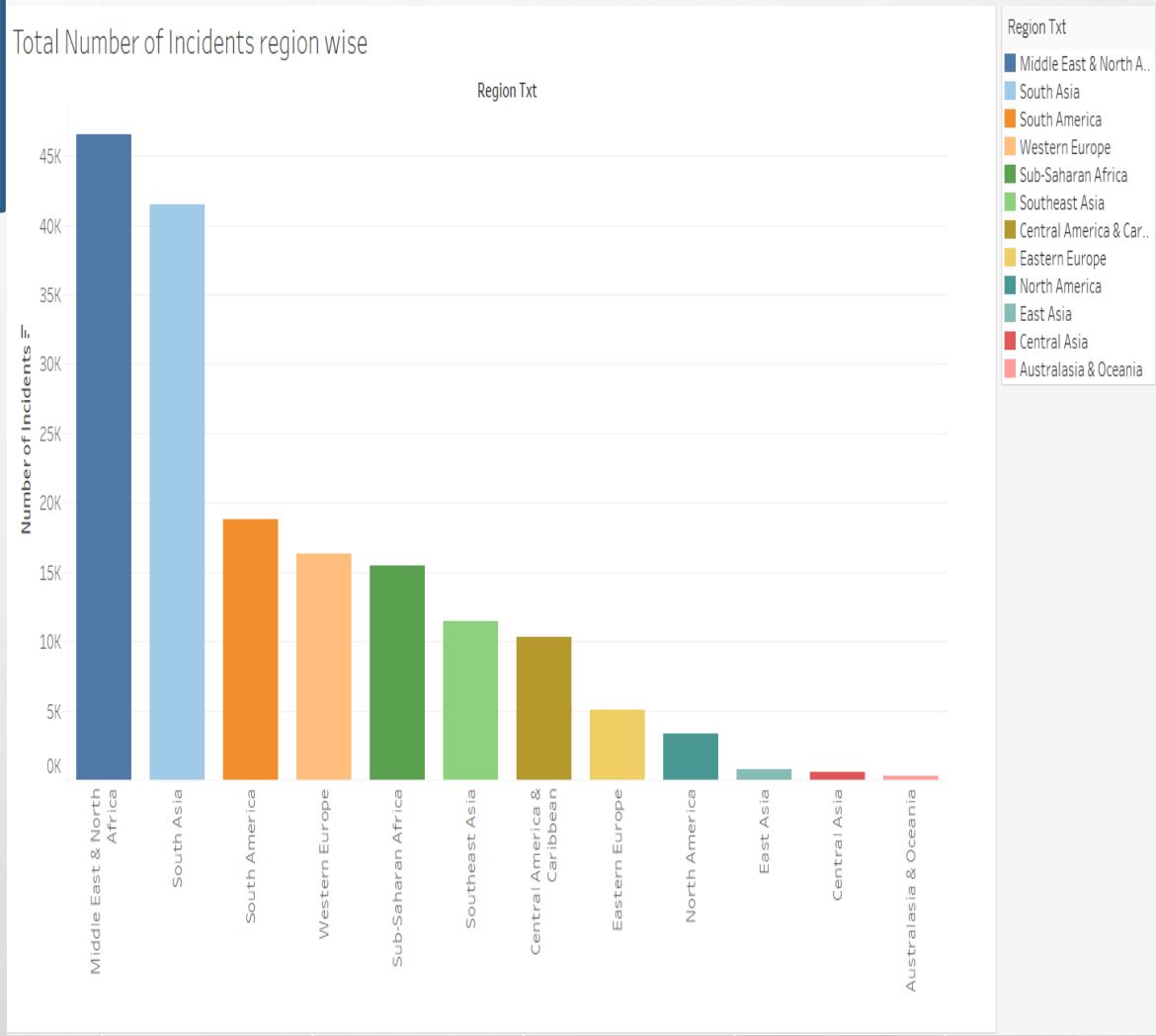


Groups claiming the attacks

Taliban [1995 - present]
Shining Path (SL) [1978 - present]
Farabundo Marti National Liberation Front [1978 - 1994]
Islamic State of Iraq and the Levant [2013 - present]



Region of attack



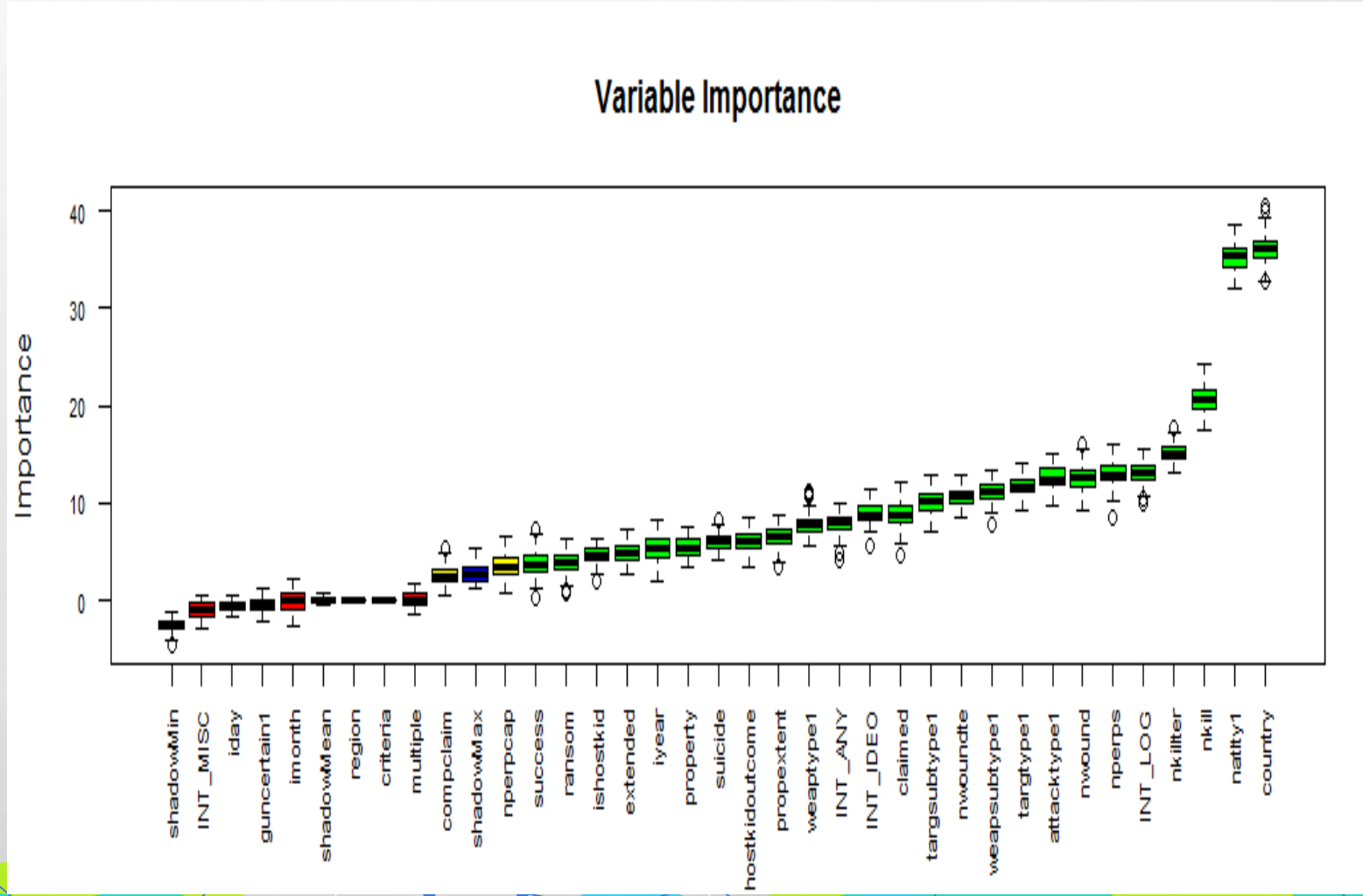
Data Pre-processing

✓ Feature Selection using Boruta

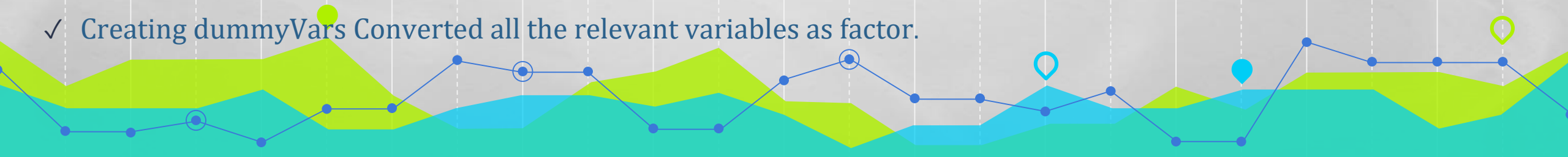
-> Wrapper on Random forest

-> Can be used with any Classification problem

-> Performs top-down search to determine most important variables



- ✓ Removed data before 1998
- ✓ Impute missing values using caret
  - ->Nkill
  - ->Nwound
  - ->Nkillter
  - ->Nwoundte
  - ->Nperps
  - ->Nperpcap
- ✓ Removed unnecessary text fields from data (Motive, Summary, latitude, longitude, provision state, city)
- ✓ Removed columns which ad more than 50% missing values
- ✓ Merged columns crit1,crit2,crit3
- ✓ compclaim, Ransom -> Replaced blank values with -99. Another category was introduced as NA
- ✓ Nalty1, Weaponsubtype1 -> -99 for blank data – unknown
- ✓ Creating dummyVars Converted all the relevant variables as factor.



Model Planning and Building

✓ Two-class Classification

-> Decision Tree

-> Bagging

-> Boosting

-> Random Forest

-> Naïve Bayes

-> Neural Network

Multiclass Classification

-> Decision Tree

-> Bagging

-> Boosting

-> Random Forest

-> Naïve Bayes

-> Neural Network

-> Support Vector Machine

✓ Sampling technique

-> 70% for training data with 10-fold cross validation

-> 30% for testing data

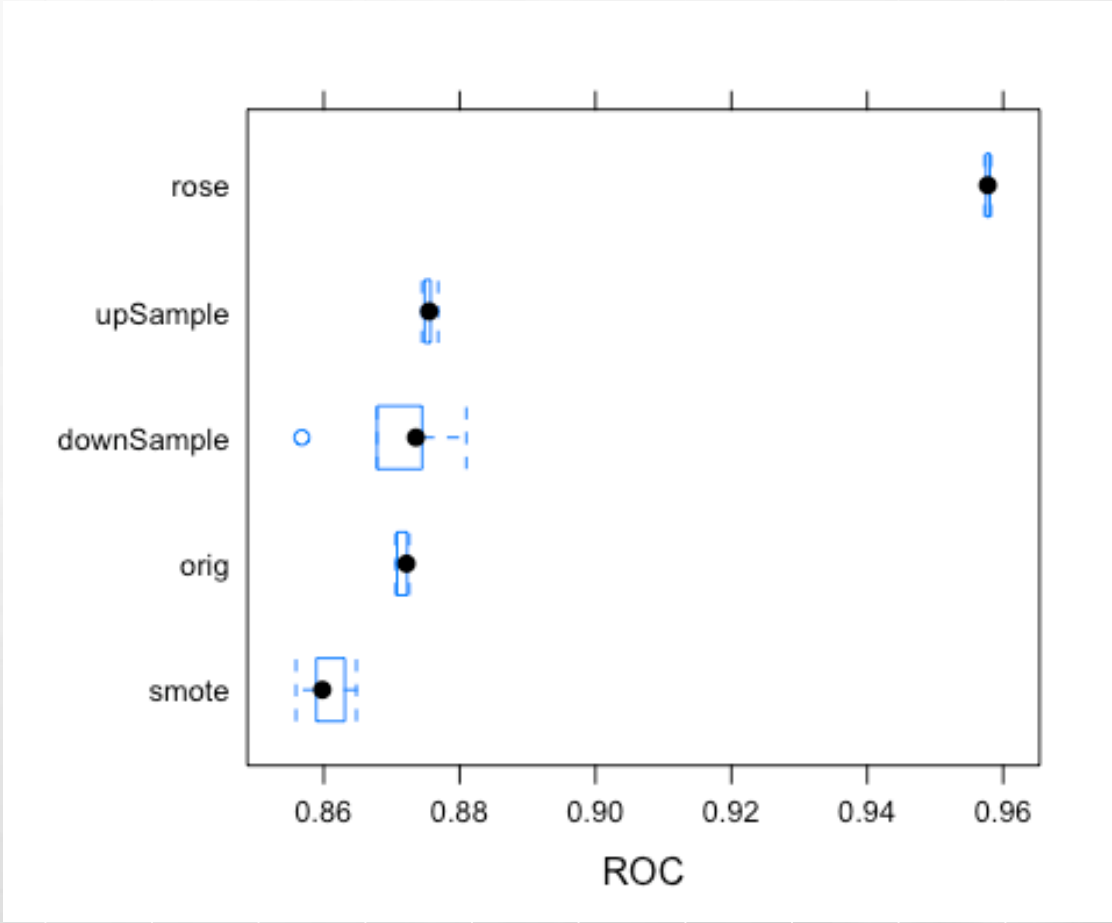
Goal 1

Predict Success of Terror Attack

Imbalance Classification

Fail	Success
12,064	90,850

- ✓ **Method to deal with Imbalanced data sets**
  - -> Down sampling
  - -> Up sampling
  - -> SMOTE
  - -> ROSE
- ✓ **As the result, we decided to use ROSE to handle imbalanced data.**



Model Description

✓ Classification Model

-> Class label: “Success” or “Fail” attack

-> Sample size: 102,914 records

-> Number of predictors: 21 Variables

Model Performance

✓ ROC Plot on training data

-> Decision Tree (rpart)

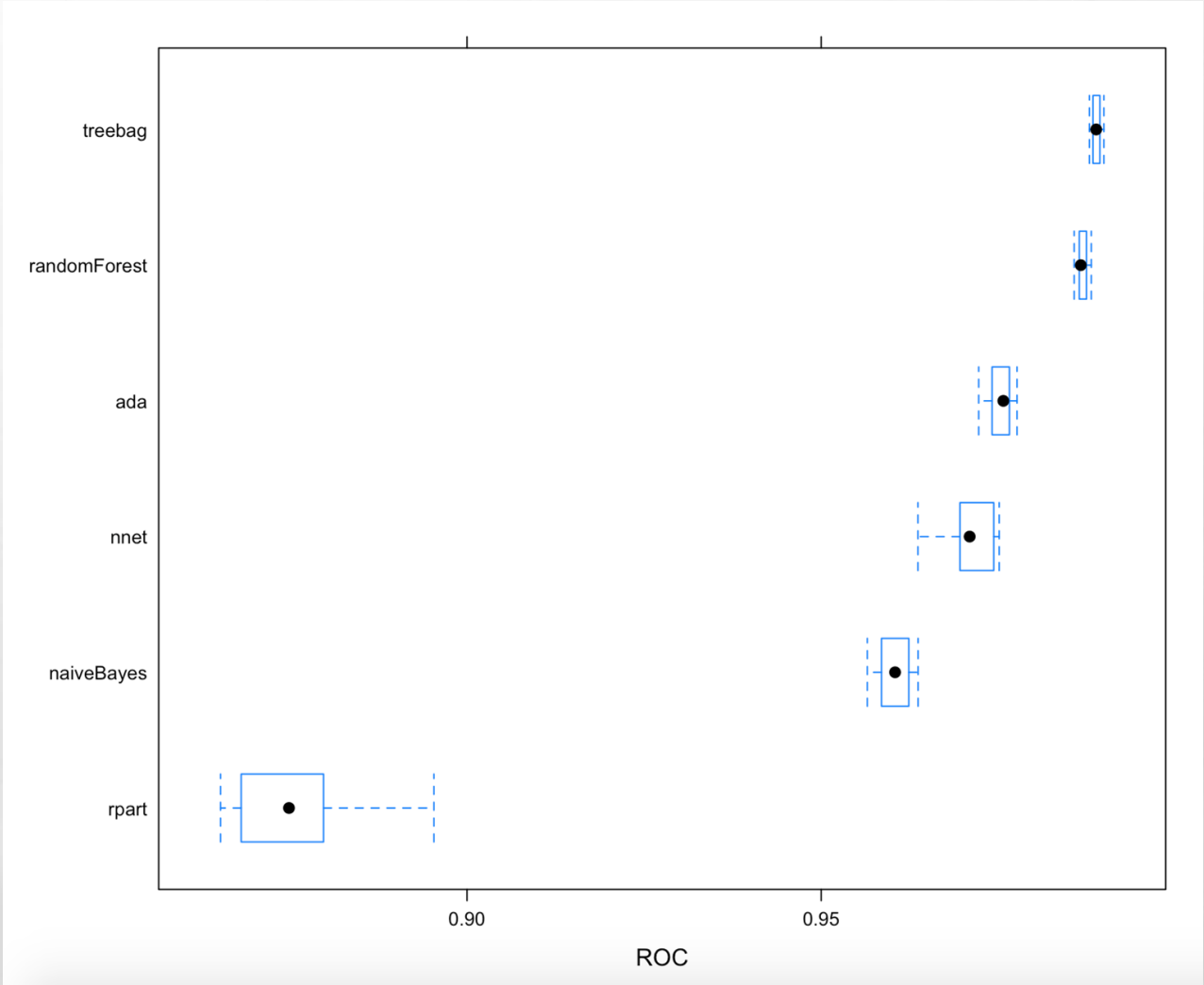
-> Bagging (treebag)

-> Random Forest (rf)

-> Boosting (ada)

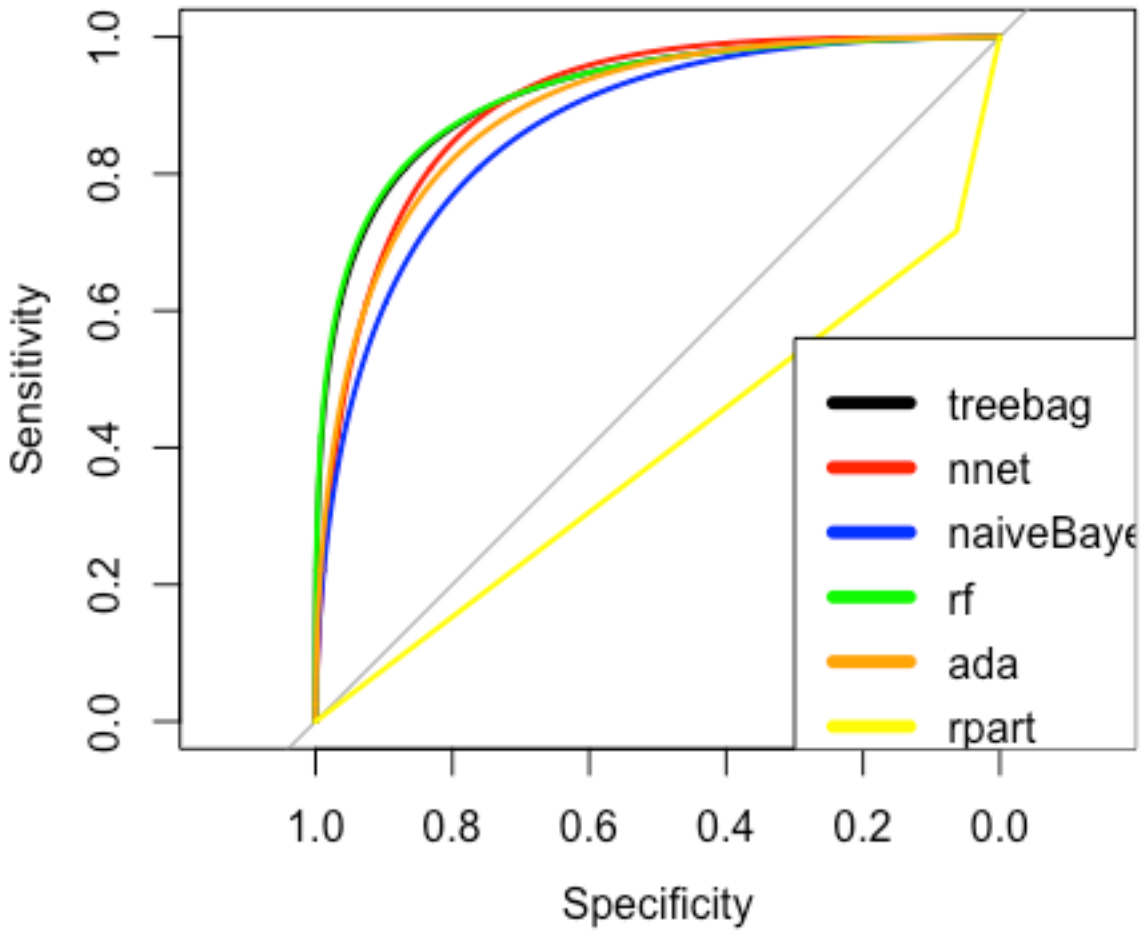
-> Neural Network (nnet)

-> Naive Bayes



Model Performance

ROC Curve on testing data



AUC & Accuracy

	AUC	Accuracy
TreeBag	0.9095	0.7203
nnet	0.9079	0.6782
randomForest	**0.9202**	**0.807**
ada	0.8955	0.5358
rpart	0.3897	0.3603
naiveBayes	0.8678	0.6263

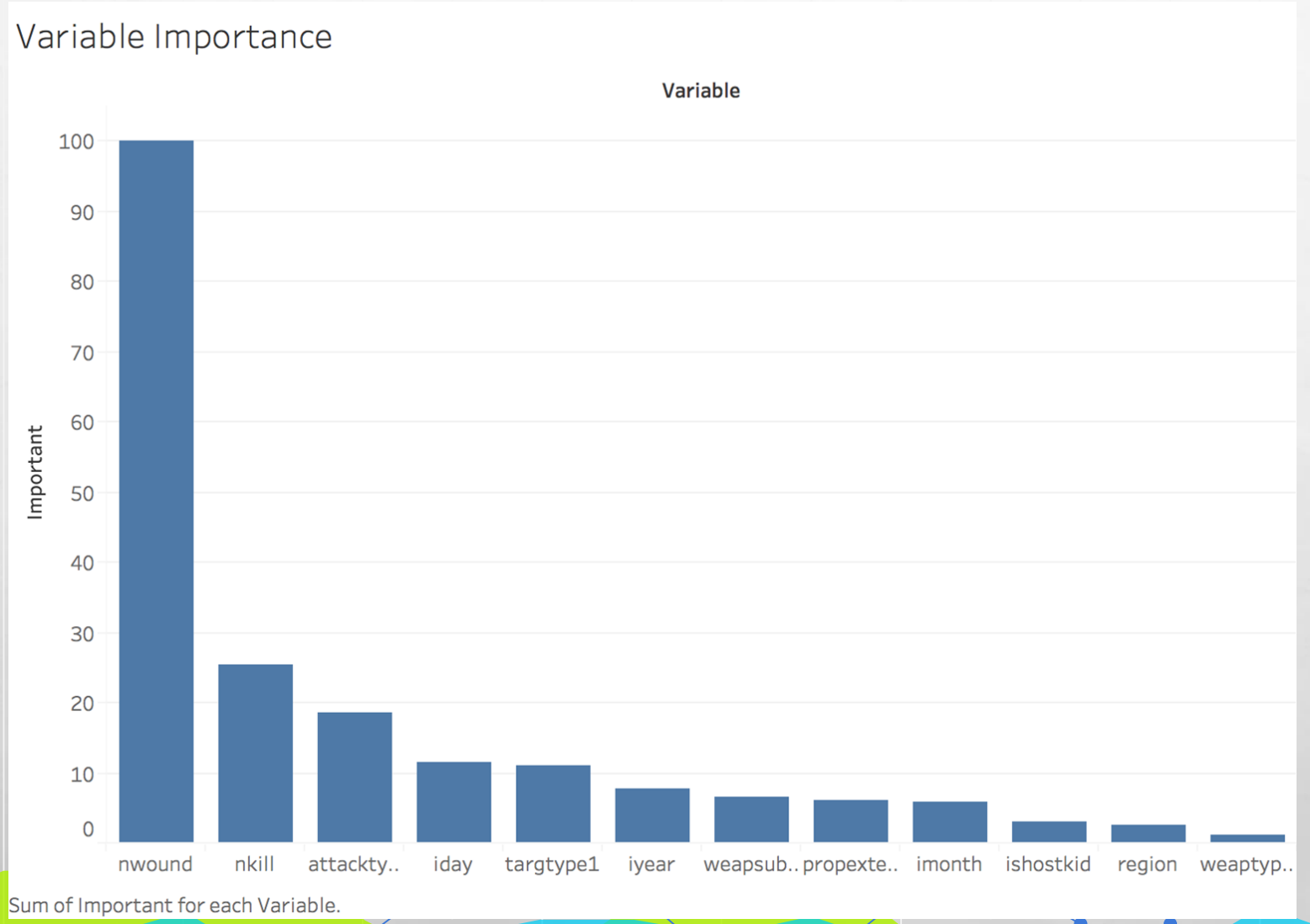
Model Performance

Confusion Matrix (RandomForest)

Metric	Value
Accuracy	0.807
Sensitivity (True Positive Rate)	0.8964
Specificity (True Negative Rate)	0.7952

```
Confusion Matrix and Statistics

          Reference
Prediction  FAIL SUCCESS
   FAIL     3244    5583
   SUCCESS   375   21672

               Accuracy : 0.807
                 95% CI : (0.8026, 0.8114)
    No Information Rate : 0.8828
    P-Value [Acc > NIR] : 1

                  Kappa : 0.4258
 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.8964
            Specificity : 0.7952
         Pos Pred Value : 0.3675
         Neg Pred Value : 0.9830
             Prevalence : 0.1172
         Detection Rate : 0.1051
   Detection Prevalence : 0.2859
      Balanced Accuracy : 0.8458
```

Variable Importance



	Overall
nwound	100.00000
nkill	25.39744
attacktype1	18.69201
iday	11.62108
targtype1	11.12302
iyear	7.69391
weapsubtype1	6.57854
propextent	6.16215
imonth	5.95752
ishostkid	2.97965
region	2.52150
weaptype1	1.06118

Goal 2

Predict Terrorist Group Name

Drilling down to South Asia Region

- ✓ Dividing the dataset according to geographic region of the attack.
- ✓ Manually selecting “real terrorist groups” by using search engines.
- ✓ Created 5 different classes as the top 5 terrorist groups with more number of incidents in South Asia region.
- ✓ Created a separate category “other” which included all the other terrorist groups apart from top 5.
- ✓ Dropped all attacks marked as doubtful.

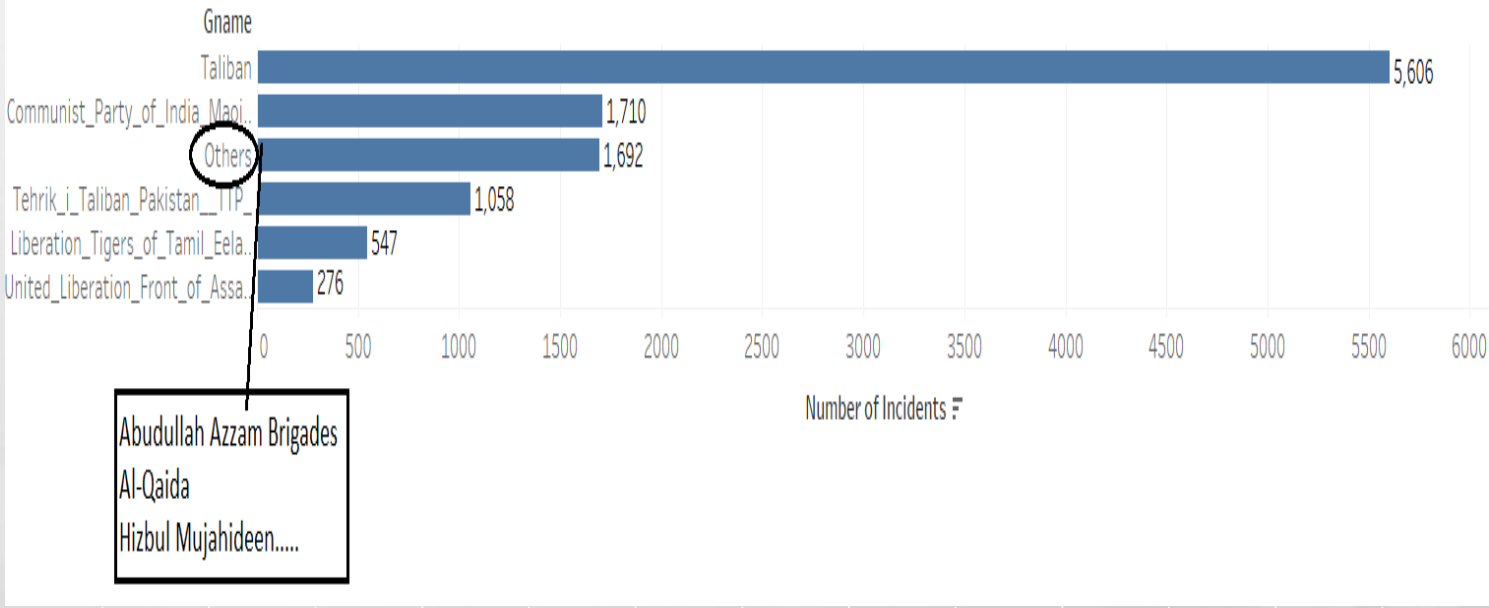
Model Description

✓ Classification Model

-> 6 Class label: “Communist Party of India_Maoist _CPI_Maoist_”, “Taliban”, “United National Liberation Front _UNLF_”, “Tehrik_i_Taliban Pakistan _TTP_”, “Liberation Tigers of Tamil Eelam _LTTE_”, “others”

-> Sample size: 10,219 records

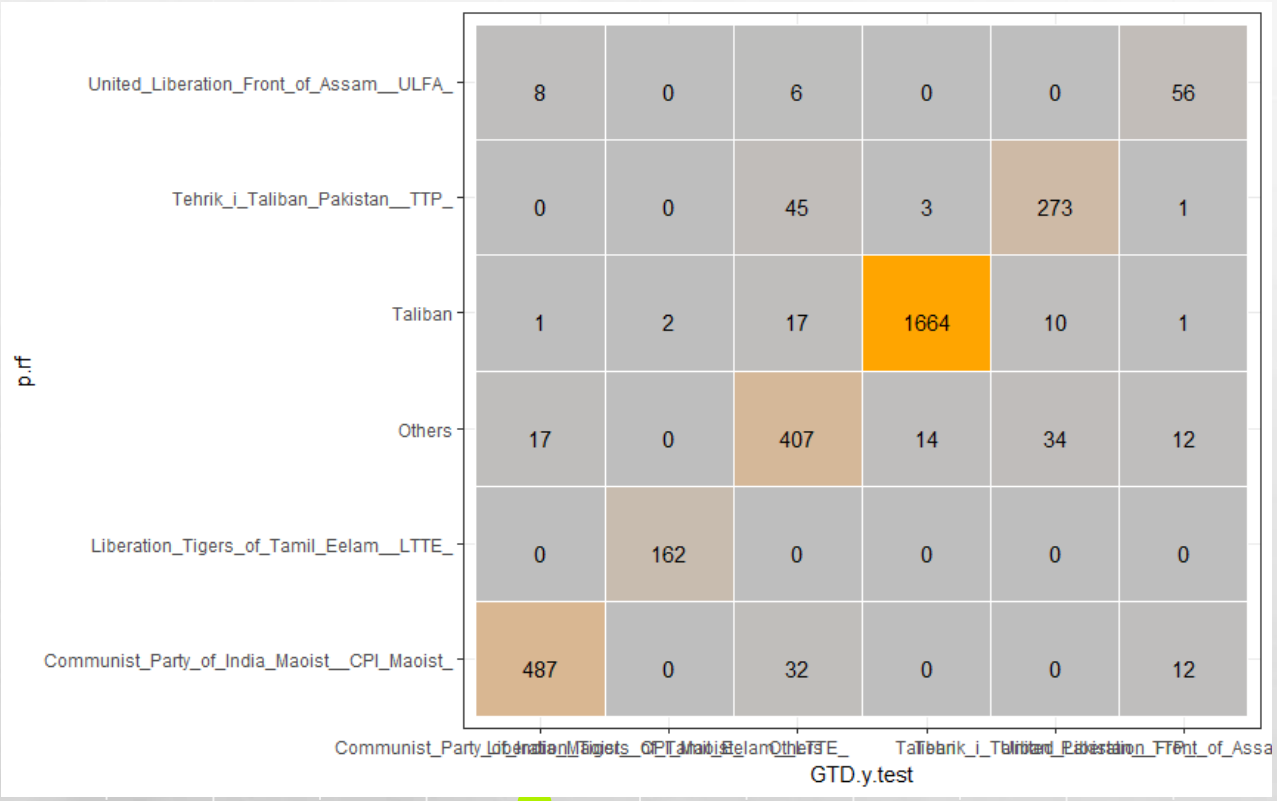
-> Number of predictors: 26 Variables



Model Performance

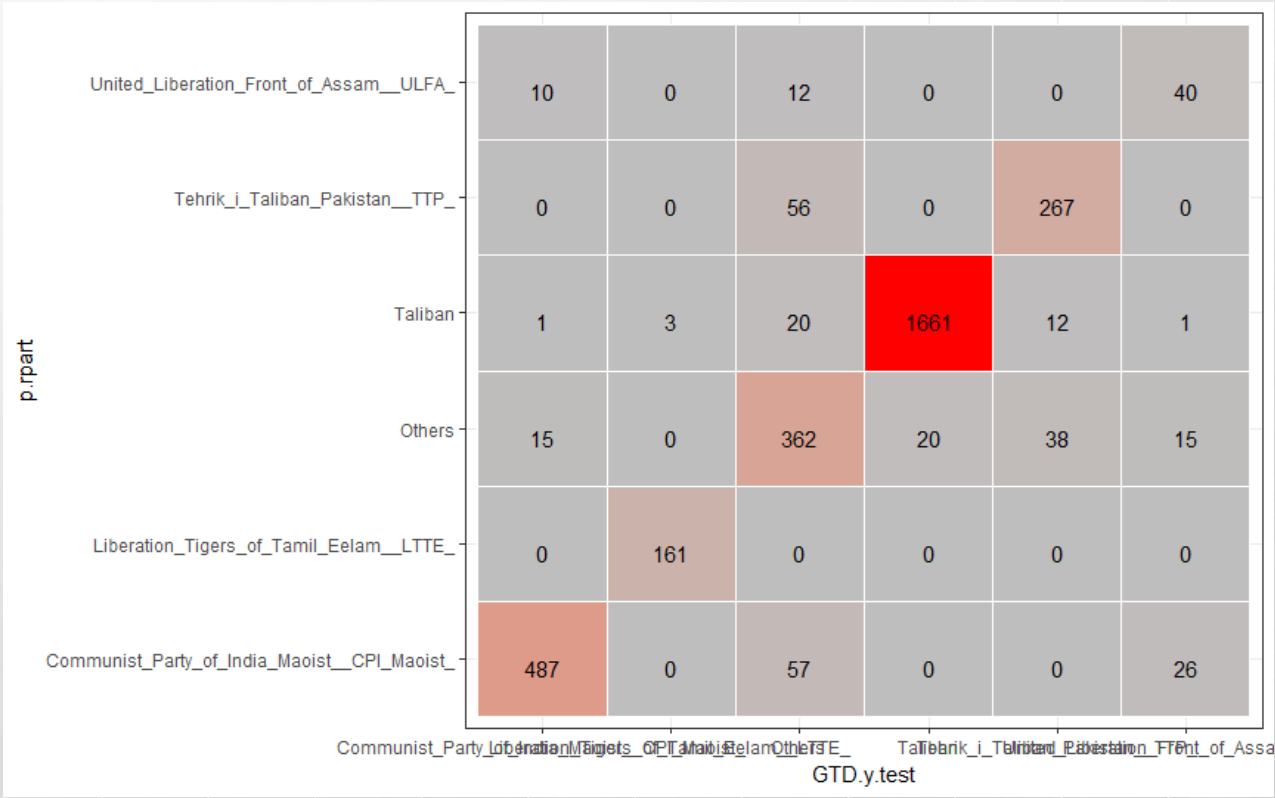
✓ Random Forest **(Best Model)**

-> Accuracy – 93.41%



✓ Decision Tree

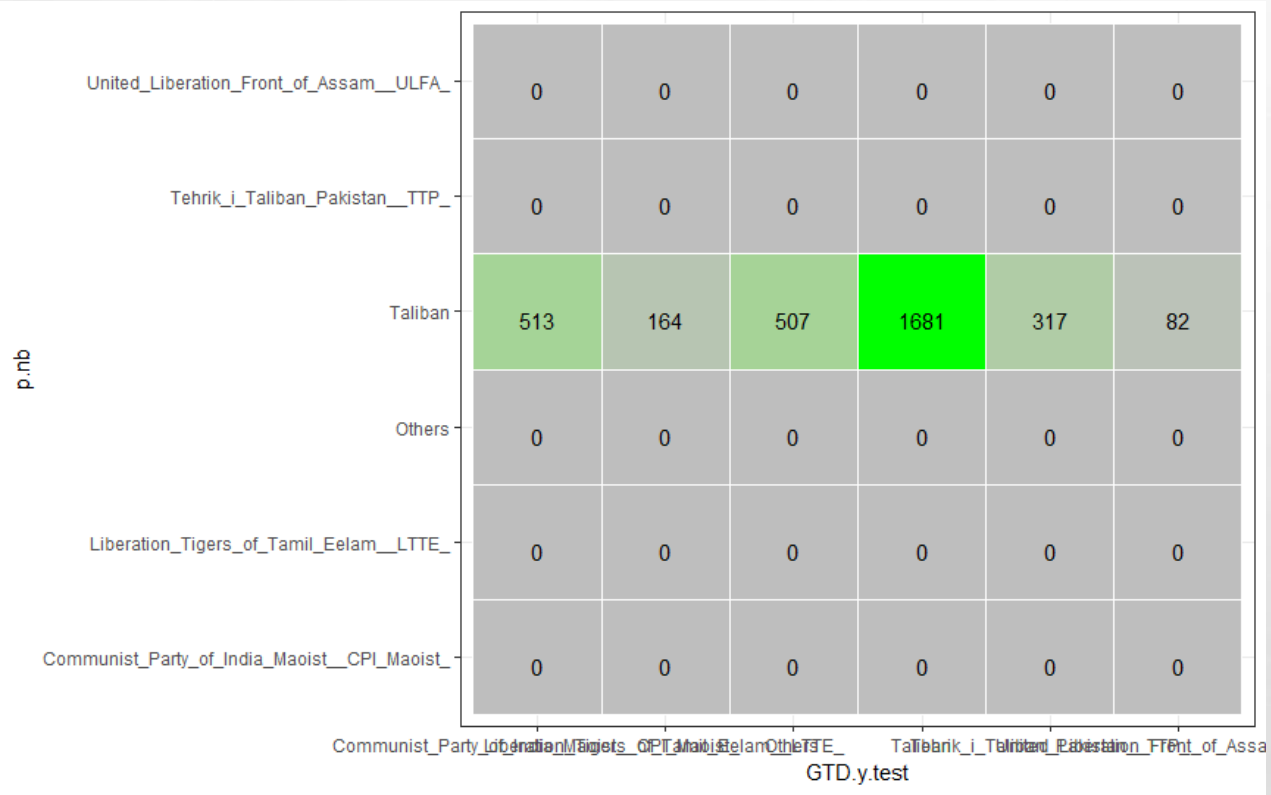
-> Accuracy – 91.24%



Model Performance

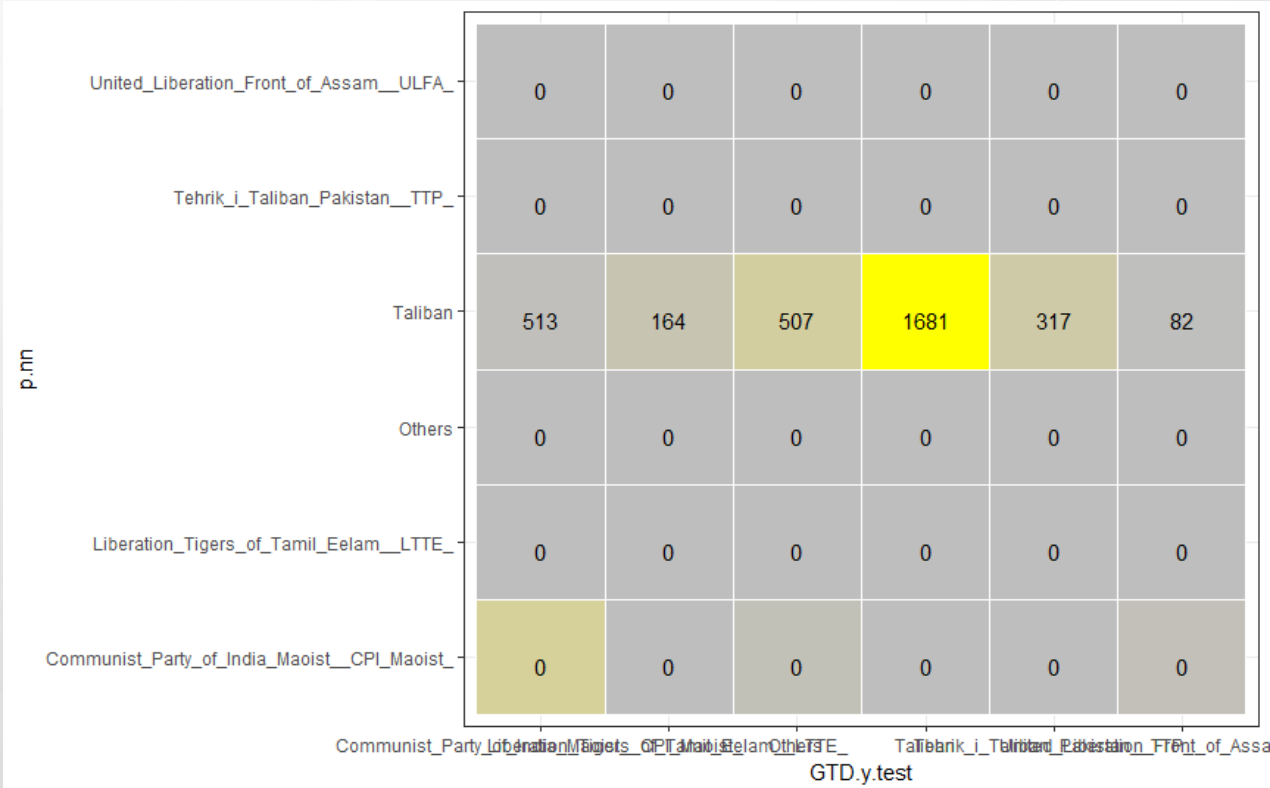
✓ Naïve Bayes

-> Accuracy – 51.5%



✓ Neural Network

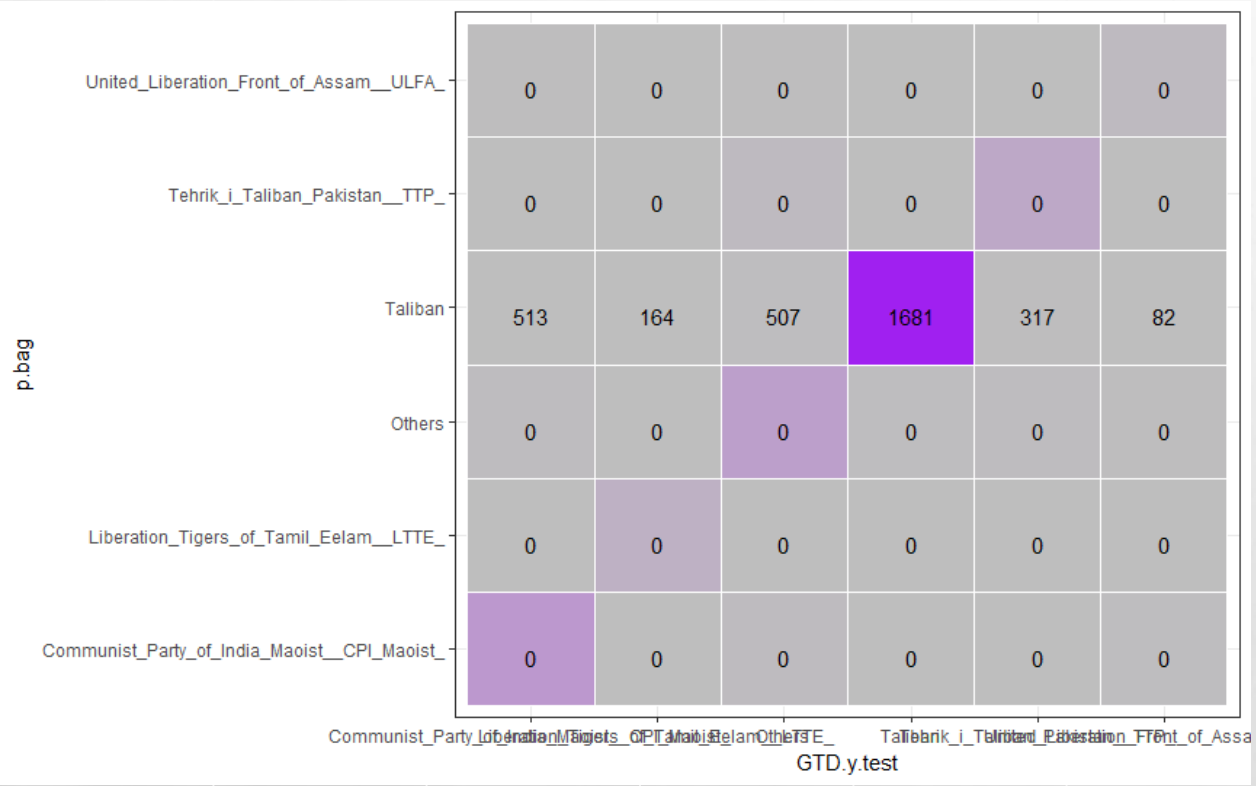
-> Accuracy –65.9%



Model Performance

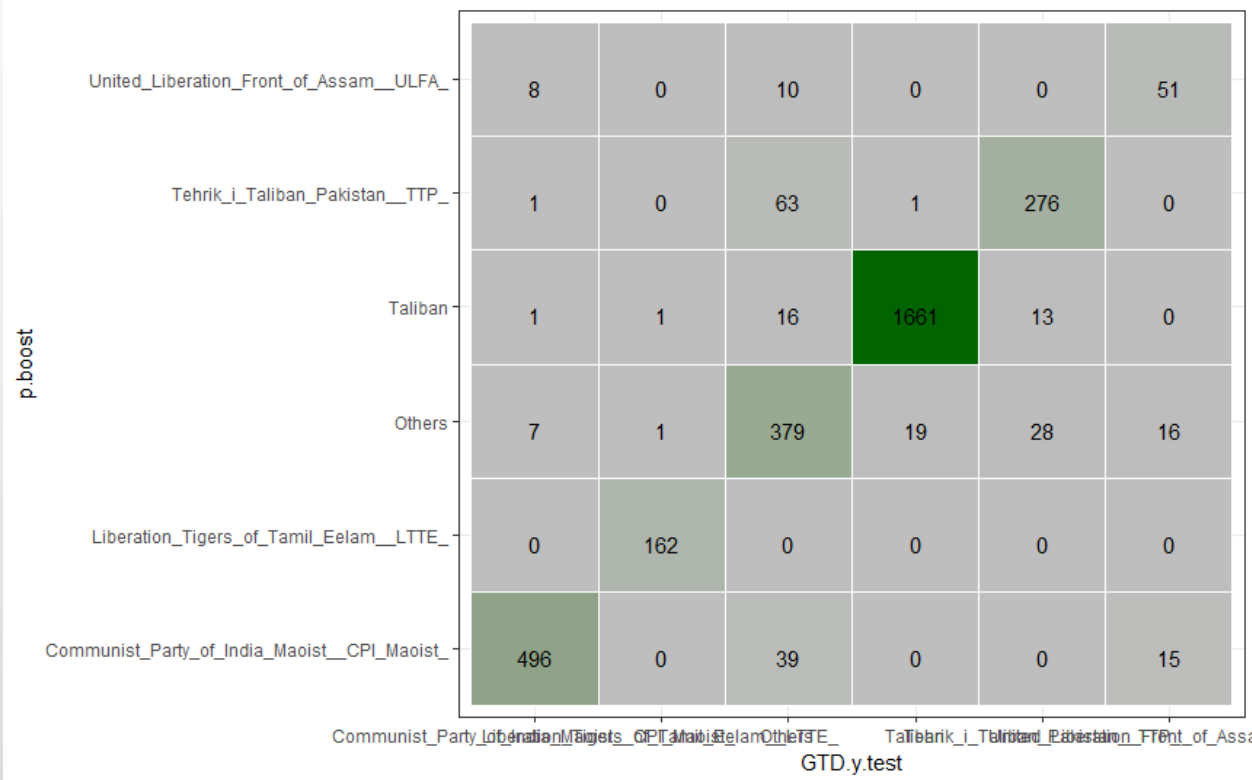
✓ Bagging

-> Accuracy – 93.01%



✓ Boosting

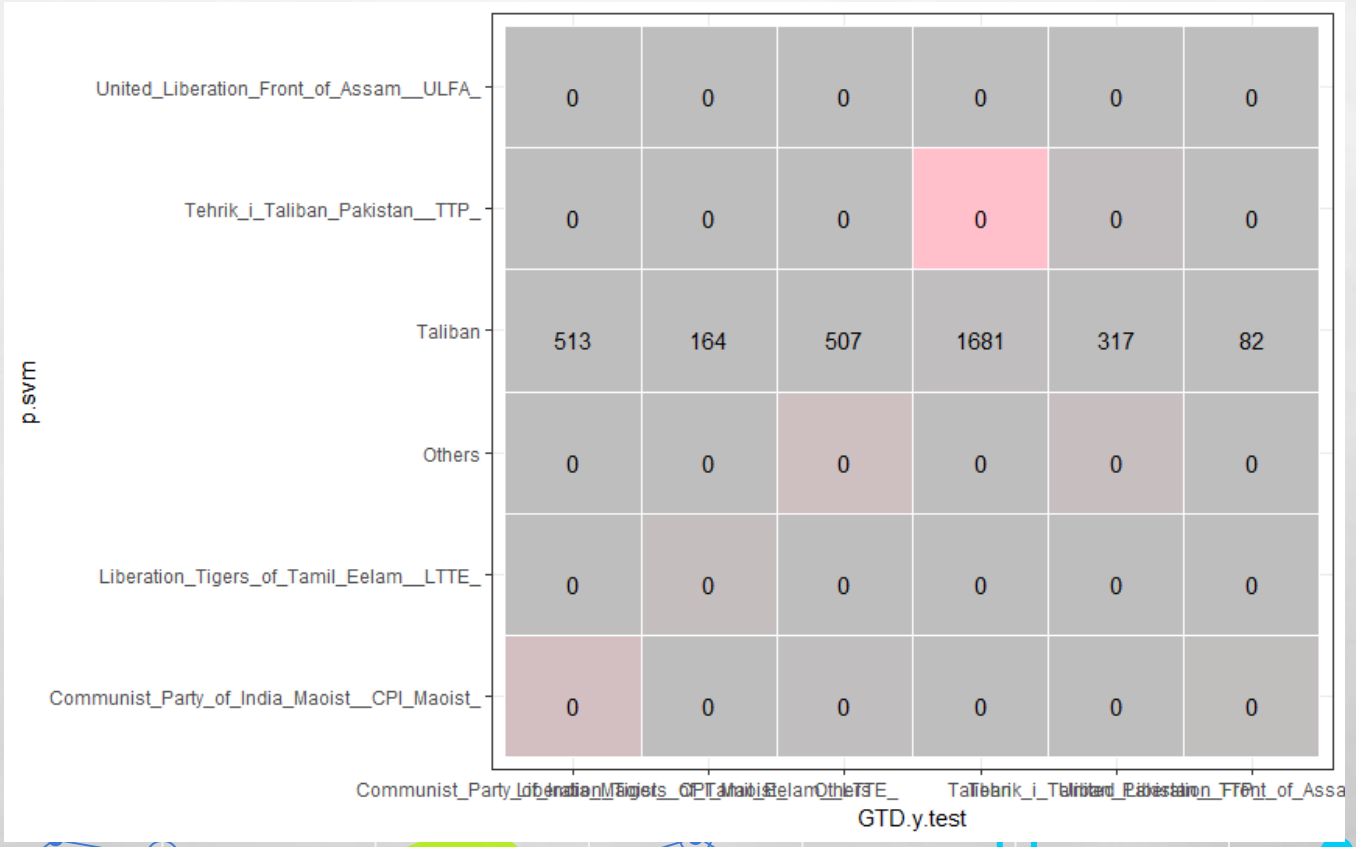
-> Accuracy –92.68%



Model Performance

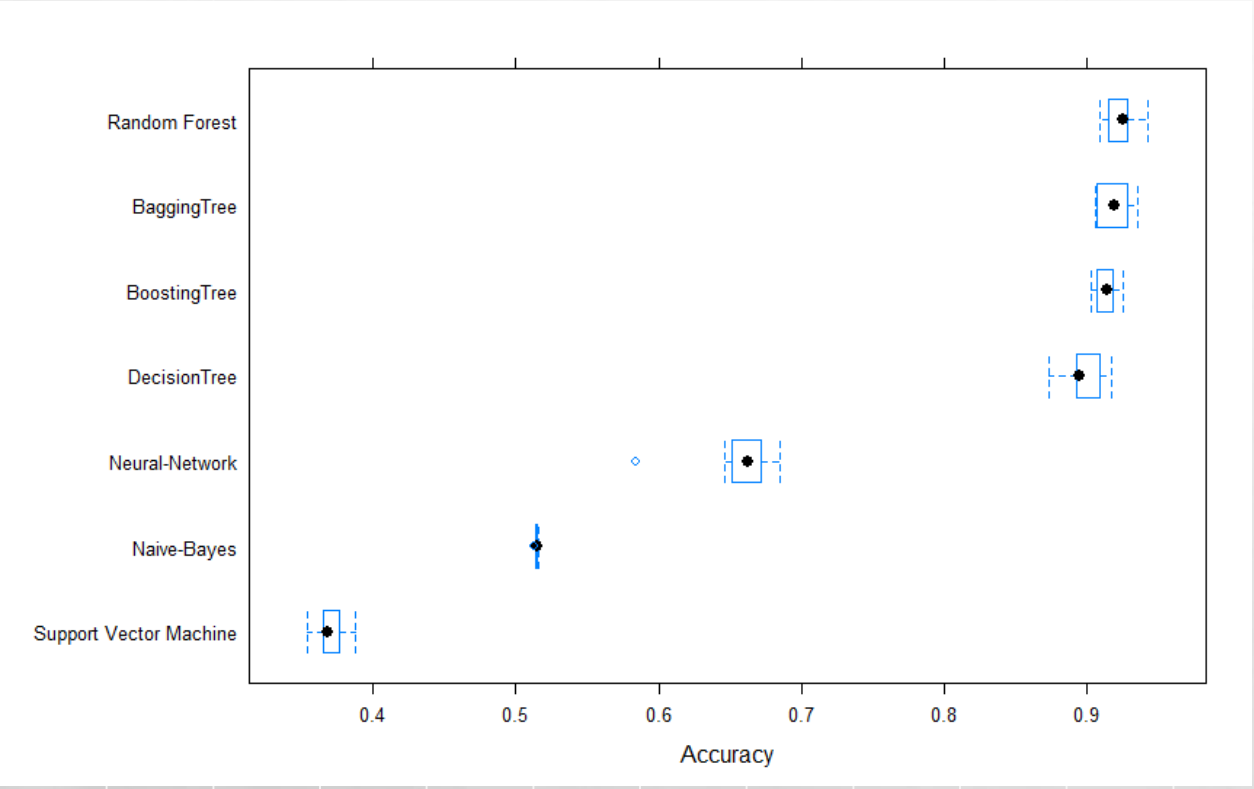
✓ Support Vector Machine

-> Accuracy – 36.4%

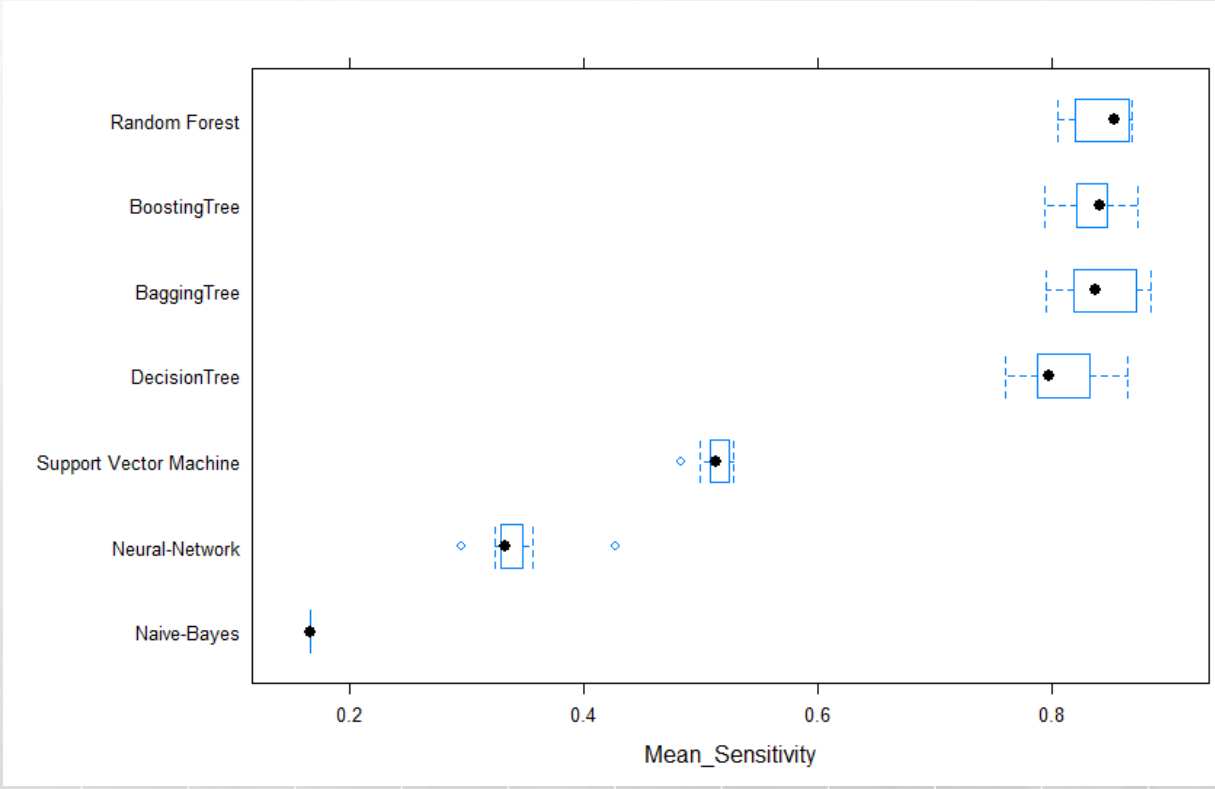


Model Comparison

✓ Box Whisker plot for Accuracy



✓ Box Whisker plot for Sensitivity



Key Points

✓Predictive model for determining the success of the attack:

-> Helps law and order to take calculated steps based on the probability of success rate

✓ Predictive model for determining the terrorist group name:

-> Help law and order predict the Terrorist group name for **unclaimed** attacks.

-> Useful to catch potential perpetrators terrorist groups for **falsely claimed** attacks.

Recommendations

- ✓ Implement the model as a pilot in order to learn on performance and precision of the model
- ✓ Extend the terrorist group predictive model for all the remaining regions of the world.
- ✓ Alter the model in a way that it can predict the probability of success for a particular group.

THANK YOU!!