

SAE 5.03

Issa, Iliane, Jules



Sommaire

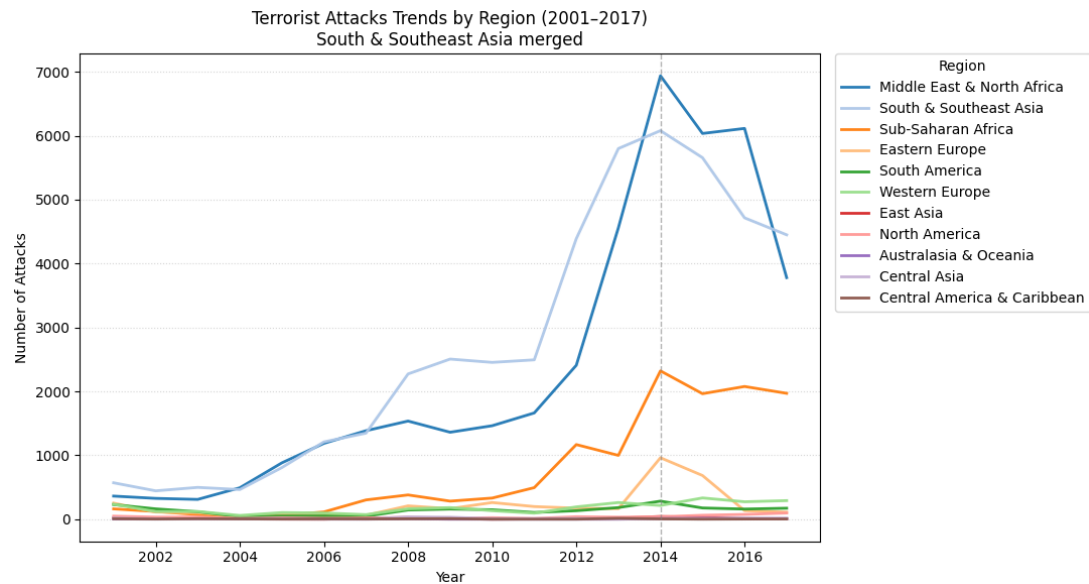
- Choix de la donnée
- EDA
- Modèles
- Conclusion



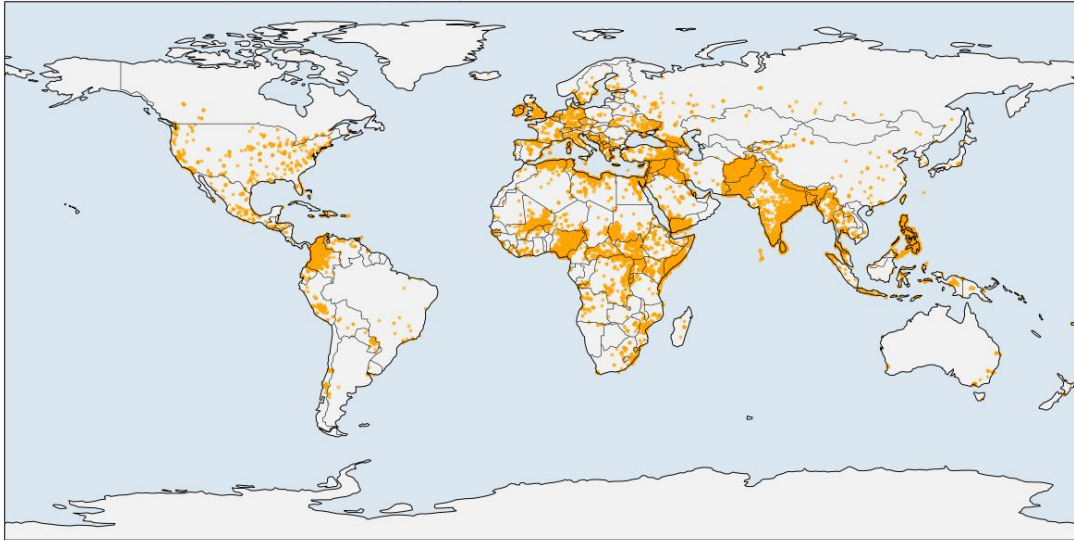
Le Dataset

| | Catégorie | Variables | Type | Description |
|---|--|---|---------------|--|
| 0 | Temps | year, month, day | Quantitatives | Datation précise de l'attaque |
| 1 | Localisation | country, region, latitude, longitude, provstat... | Qualitatives | Localisation géographique complète |
| 2 | Typologie d'attaque | attacktype1, targtype1, weaptype1 | Qualitatives | Type d'attaque, cible, arme utilisée |
| 3 | Victimes (globale / US) | nkill, nwound, nkillus, nwoundus | Quantitatives | Nombre de morts et blessés (total et US) |
| 4 | Acteur terroriste | gname, gsubname, claimmode | Qualitatives | Groupe terroriste et mode de revendication |
| 5 | Motivations / Implication internationale | motive, INT_LOG, INT_IDEO, INT_MISC | Qualitatives | Motivation + implication internationale |
| 6 | Descriptif & sources | summary, dbsource, related | Qualitatives | Description textuelle et sources |

- **Global Terrorism Database**
- **180 000+ attaques**
- **17 variables clés**



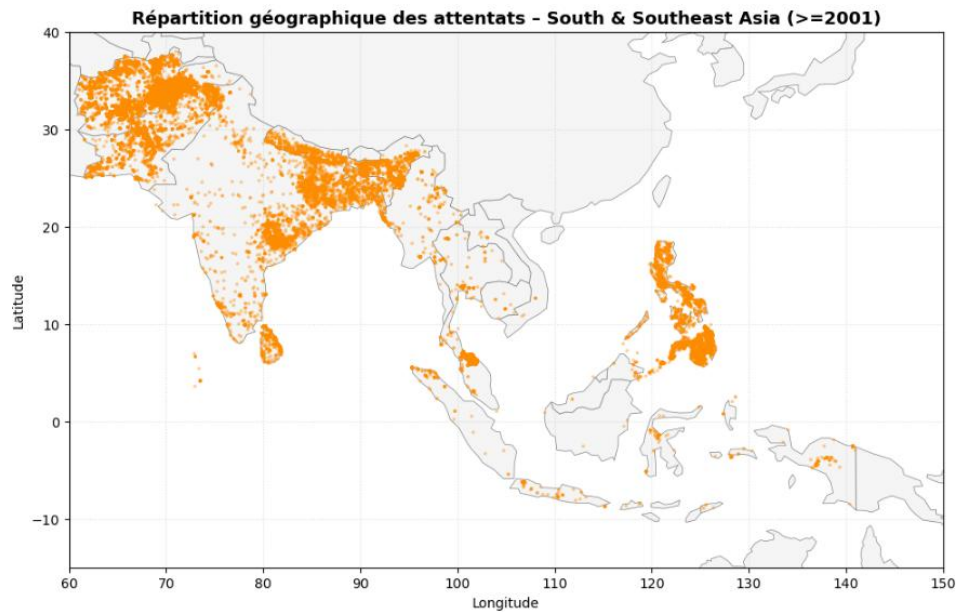
Répartition géographique mondiale des attentats (>=2001)



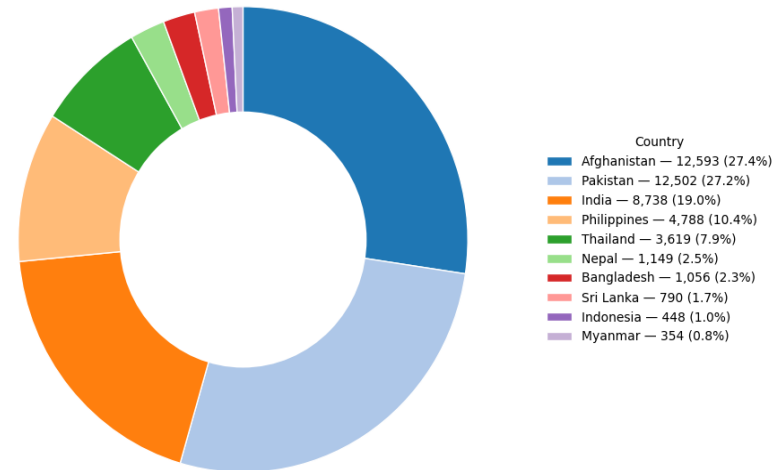
EDA : Vue mondiale

- **Pic 2014-2016**
- **Moyen-Orient / Asie du Sud dominants**
- **Explosifs --> méthode principale**

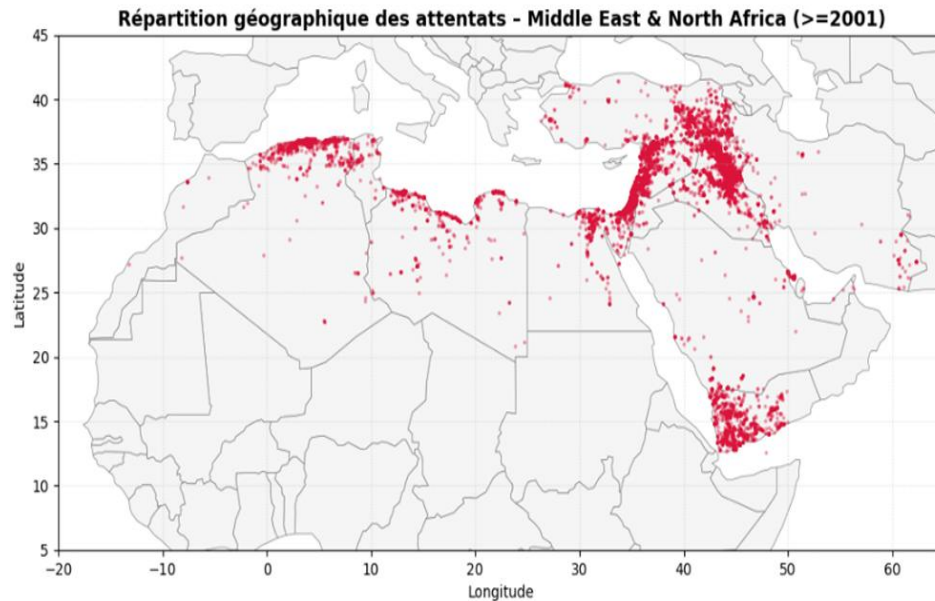
Analyse géographique (Asie)



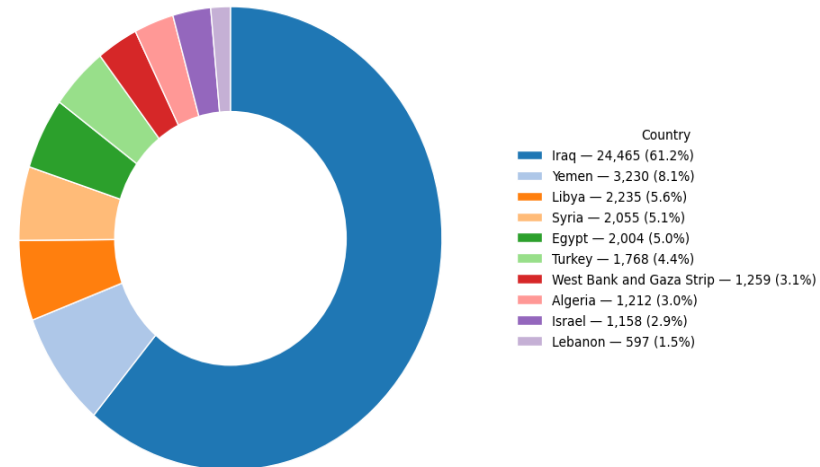
Top 10 Countries in South & Southeast Asia
by Number of Terrorist Attacks (2001-2017)



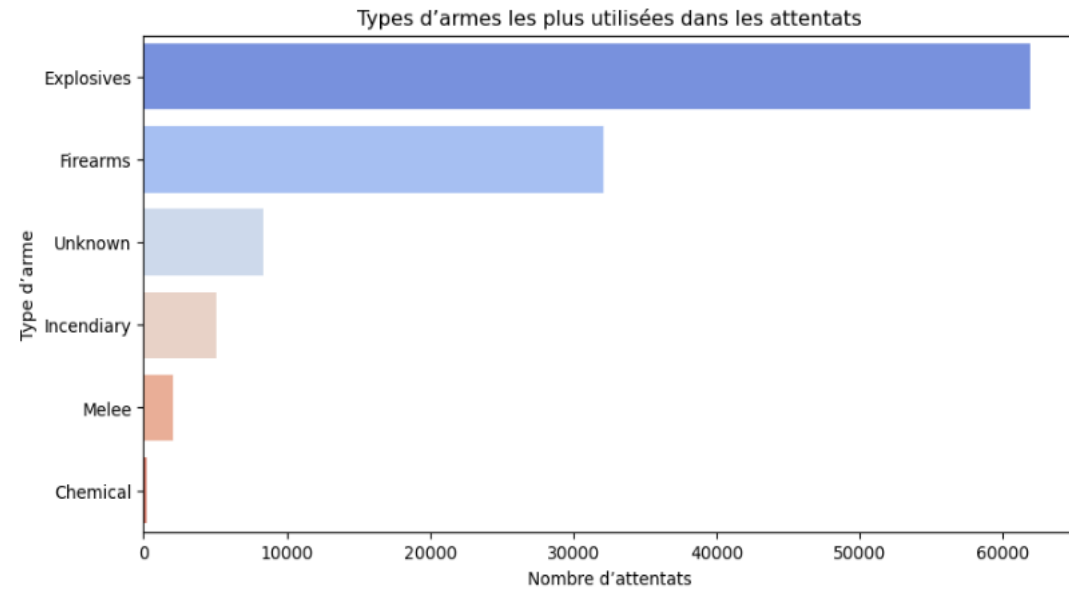
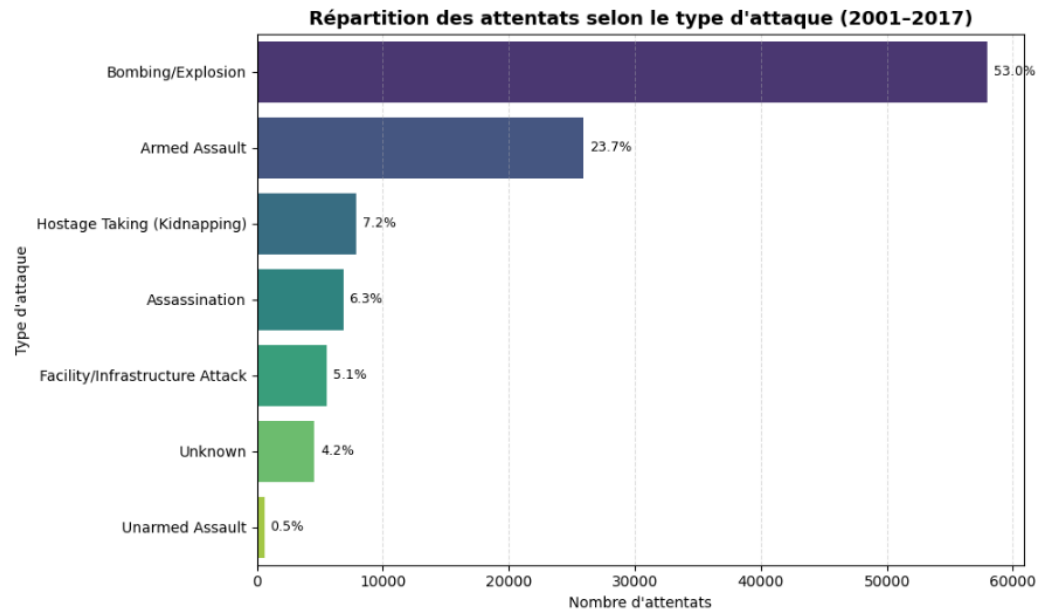
Analyse géographique (MEA&North Africa)



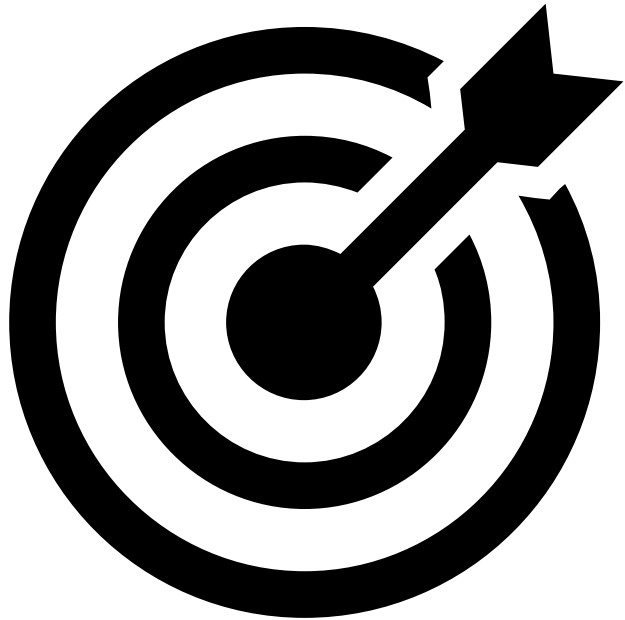
Top 10 Countries in Middle East & North Africa
by Number of Terrorist Attacks (2001-2017)



Typologie des attaques



Modèles réalisés



Problématique posée : Peut-on prédire le type d'attaque ?

Variable cible : attack_type (9 classes)

Données utilisées :

- caractéristiques temporelles*
- localisation*
- armes & cibles*
- gravité (nkill, nwound)*

Pipeline & Préparation

Transformations appliquées :

- Variables numériques → imputation + standardisation
- Variables catégorielles → encodage + imputation
- Suppression des classes avec < 2 observations
- Train/test 75/25 + stratification



Choix du modèle

Random Forest

Raisons :

- Meilleur modèle sur l'ensemble des tests
- Très bonne généralisation ($CV \approx \text{Test}$)
- Les autres modèles overfitent ou sont instables
- Fiable et robuste pour notre problématique.



Validation croisée 5-fold

Moyennes sur les 5 folds (cross-validation) :

| | |
|--------------------|-----------------|
| accuracy | : 0.819 ± 0.003 |
| f1_weighted | : 0.826 ± 0.003 |
| precision_weighted | : 0.844 ± 0.002 |
| recall_weighted | : 0.819 ± 0.003 |

- *Très bon sur les classes fréquentes*
- *Stabilité du modèle*
- *Influence de la classe dominante*
Bombing/Explosion

Score final sur test set



- *Performant globalement*
- *Classes rares très mal prédites*
- *Déséquilibre majeur du dataset*

| | | | | |
|--------------|-------|-------|-------|-------|
| accuracy | | | 0.815 | 21604 |
| macro avg | 0.506 | 0.620 | 0.517 | 21604 |
| weighted avg | 0.845 | 0.815 | 0.825 | 21604 |

Analyse classe par classe

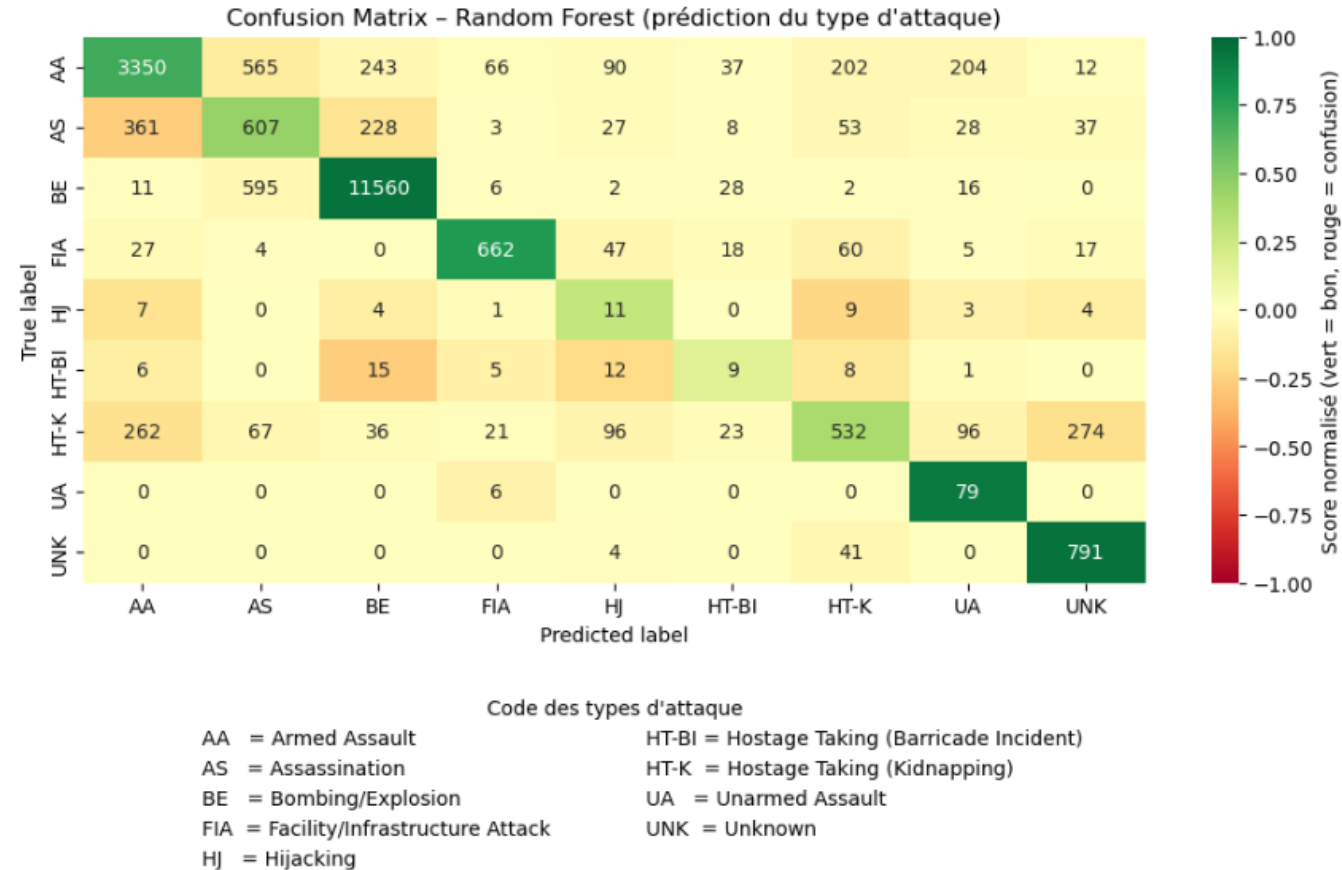
Évaluation finale sur le test set :

| | precision | recall | f1-score | support |
|-------------------------------------|-----------|--------|----------|---------|
| Armed Assault | 0.833 | 0.702 | 0.762 | 4769 |
| Assassination | 0.330 | 0.449 | 0.381 | 1352 |
| Bombing/Explosion | 0.956 | 0.946 | 0.951 | 12220 |
| Facility/Infrastructure Attack | 0.860 | 0.788 | 0.822 | 840 |
| Hijacking | 0.038 | 0.282 | 0.067 | 39 |
| Hostage Taking (Barricade Incident) | 0.073 | 0.161 | 0.101 | 56 |
| Hostage Taking (Kidnapping) | 0.587 | 0.378 | 0.460 | 1407 |
| Unarmed Assault | 0.183 | 0.929 | 0.306 | 85 |
| Unknown | 0.697 | 0.946 | 0.803 | 836 |

| | | | |
|--------------|--------------------|---------------|----------------|
| Très bon | Bombing /Explosion | Armed Assault | |
| Très mauvais | Assassination | Hijacking | Hostage Taking |

Matrice de confusion

- *Explosion quasi parfaite*
- *Confusions fréquentes entre Armed Assault / Assassination*
- *Classes rares invisibles (aucun signal)*



Analyse des erreurs du modèle

Confusion entre types proches
Armed Assault / Hostage Taking

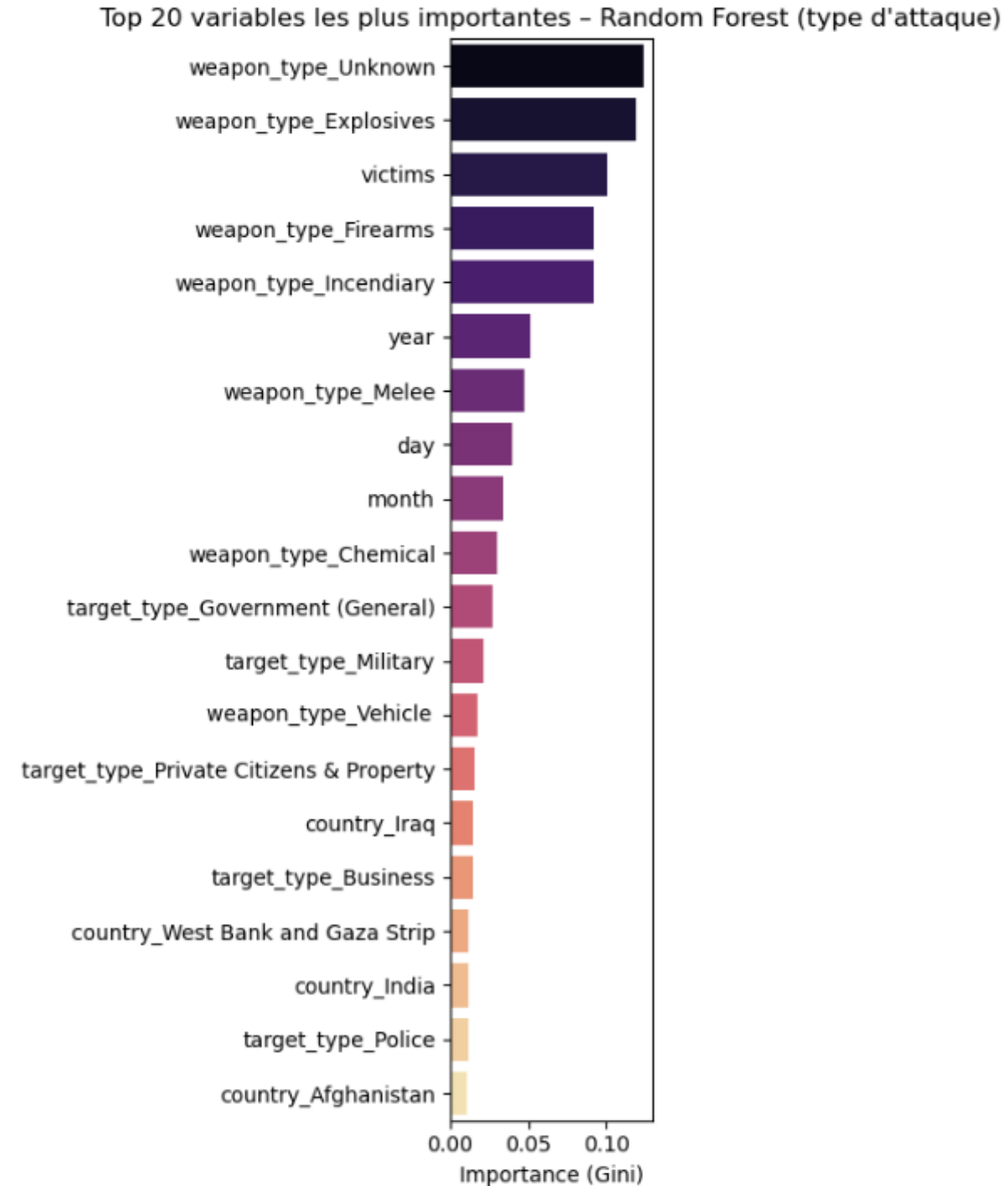
Classes rares mal apprises
trop peu d'exemples

Peu de victimes = signal faible
pas assez discriminant

| | region | country | weapon_type | target_type | year | month | day | victims | true | pred |
|--------|----------------------------|---------|-------------|----------------------|------|-------|-----|---------|-------------------------------------|-----------------------------|
| 135951 | Middle East & North Africa | Lebanon | Firearms | Military | 2014 | 8 | 2 | 0.0 | Hostage Taking (Kidnapping) | Armed Assault |
| 159035 | Middle East & North Africa | Iraq | Firearms | Journalists & Media | 2016 | 2 | 17 | 0.0 | Assassination | Hostage Taking (Kidnapping) |
| 137947 | Middle East & North Africa | Libya | Firearms | Police | 2014 | 9 | 17 | 1.0 | Assassination | Armed Assault |
| 105650 | Middle East & North Africa | Egypt | Firearms | Police | 2012 | 2 | 9 | 1.0 | Hostage Taking (Barricade Incident) | Armed Assault |
| 141376 | Middle East & North Africa | Yemen | Explosives | Government (General) | 2014 | 12 | 8 | 0.0 | Bombing/Explosion | Assassination |

Importance des variables

- *weapon_type = variable dominante*
- *Puis victims / target_type / region*
- *Le modèle repose sur les choix tactiques + contexte spatio-temporel*



Conclusion

Nos données

Dataset GTD (2001–2017), variables sur lieux, types et victimes.

EDA

Tendances, zones les plus touchées, typologies

Modèles

Random Forest $\approx 80\%$, tests cohérents.

Synthèse

Prédictions fiables sur classes fréquentes, cohérence géographique.



Merci
