

AFRICAN ARMED CONFLICT PREDICTION

A MACHINE LEARNING PERSPECTIVE

A Capstone Project

Presented by: Alexandra Andrade

8 June 2018

CONTENT

- ▶ The Problem
- ▶ Methodology
- ▶ Dataset
- ▶ Modeling stages
- ▶ Results and Analysis
- ▶ Conclusions

THE PROBLEM

In 1970, Edwin Star, anti-war activist and singer famously asked the question (in song:)

“WAR, what is it good for?”

“Absolutely nothing...”

**What about about
completely awesome, utterly
brilliant ML projects????**



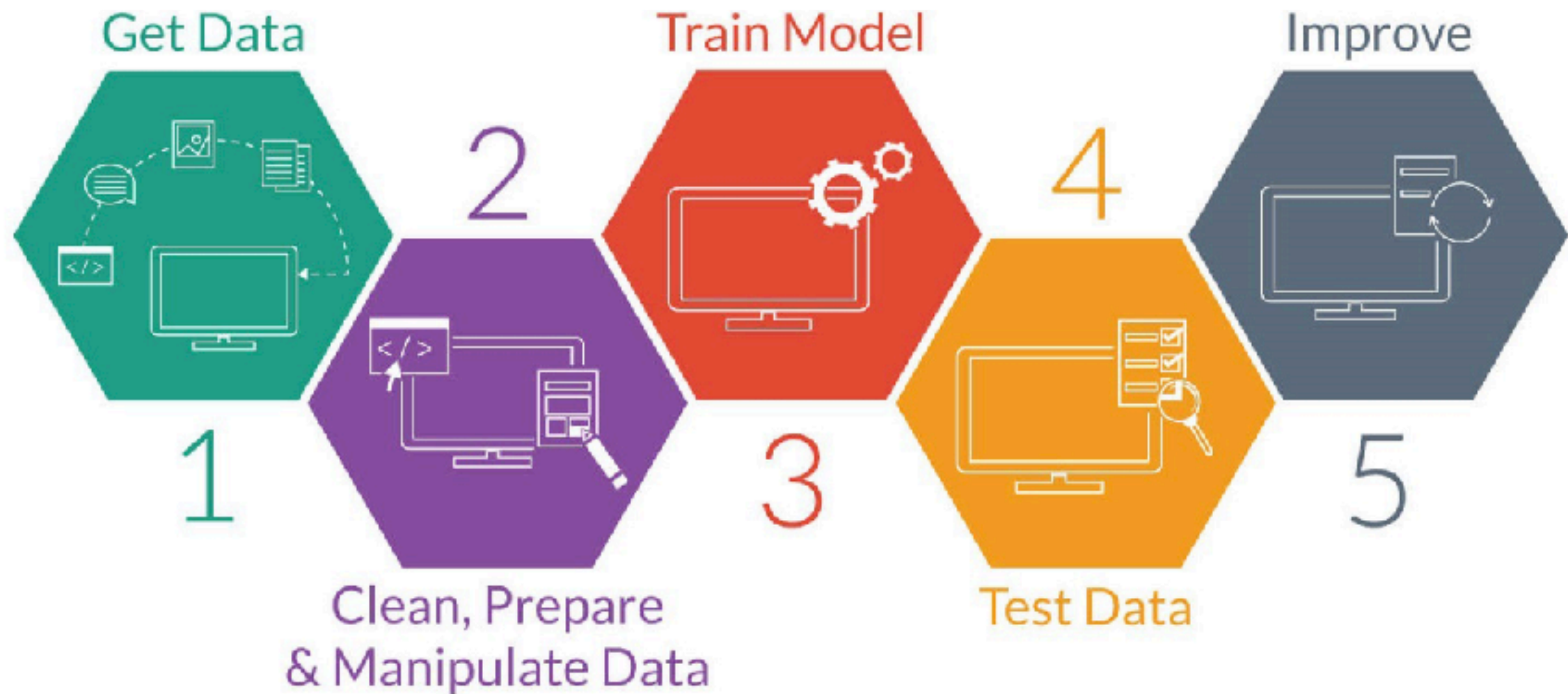
THE BIG QUESTION:

- ▶ Can I predict if a reported event has lead to fatalities?
- ▶ Target: Is the event deadly?
- ▶ Binary Classification Problem

The smaller question:

- ▶ What are the factors that differentiate deadly events from non-deadly ones?

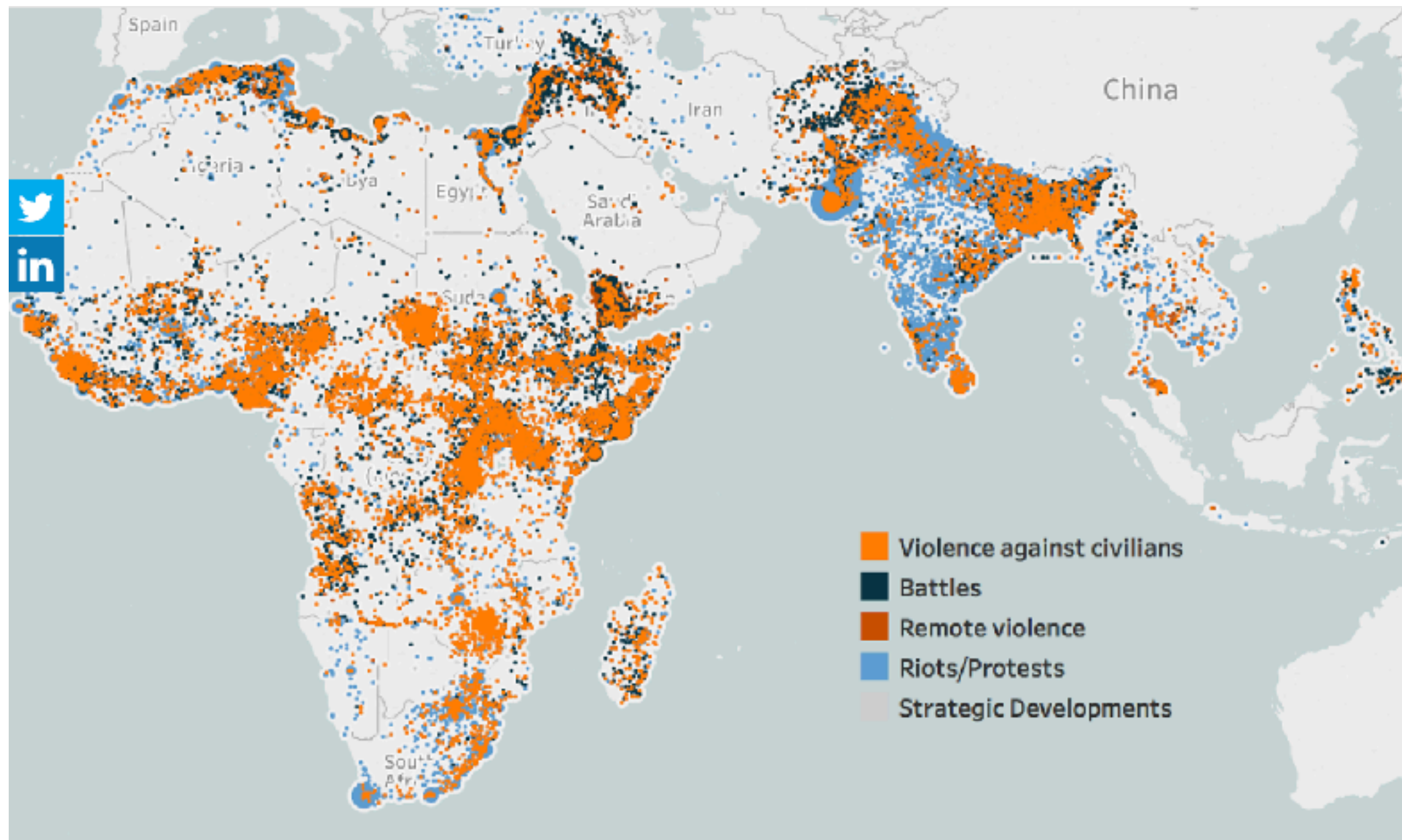
WORKFLOW



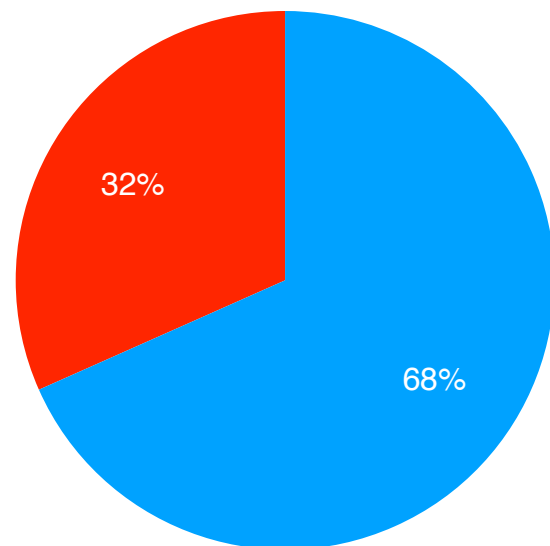
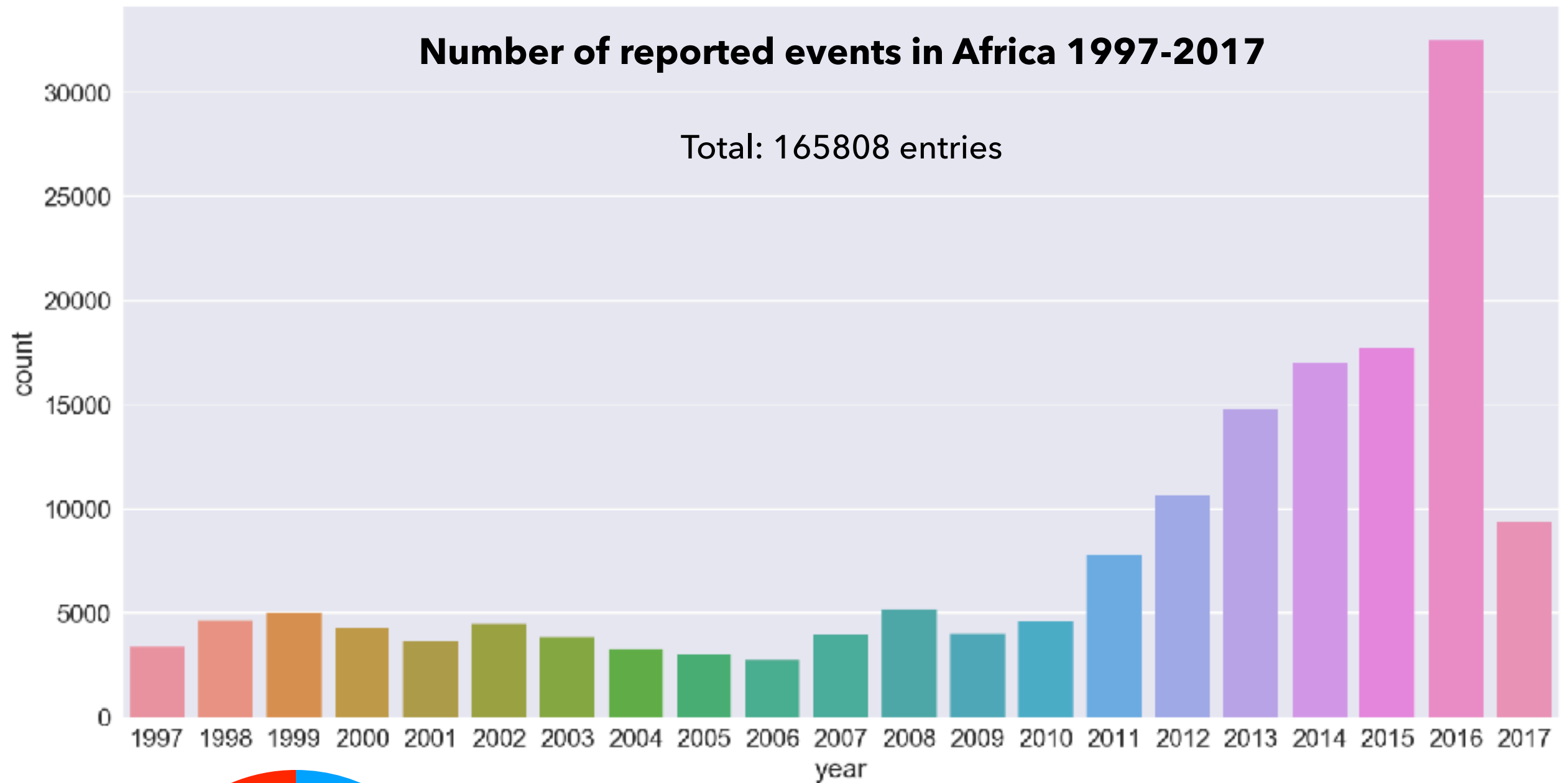
<https://machinelearning-blog.com/2017/11/19/fsgdhfju/>

THE DATASET

- ▶ Taken from Armed Conflict Location & Event Data Project (ACLED).
- ▶ ACLED is a disaggregated conflict collection, analysis and crisis mapping project. ACLED collects the dates, actors, types of violence, locations and fatality count of all reported political violence and protest events across Africa, South East Asia and the Middle East
- ▶ link: <https://www.acleddata.com/>



THE DATASET



68% of the events do not have associated fatalities

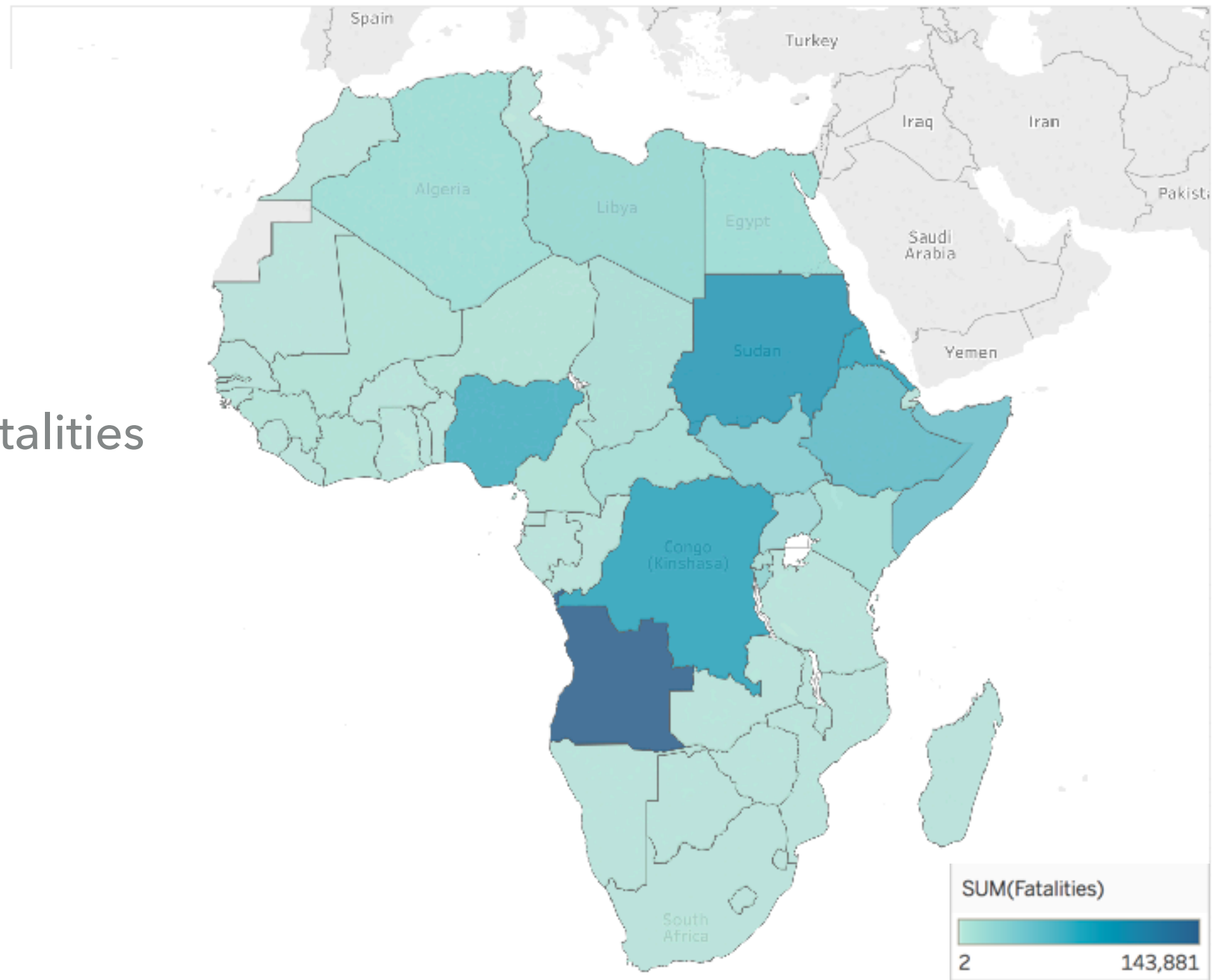
EDA, FE AND FS

► Data includes:

- Country
- Data Source
- Number of Fatalities
- Actors
- Location
- Event Type

► Total:

- 28 cols



Most countries in Africa (50 out of 54) have come entries.

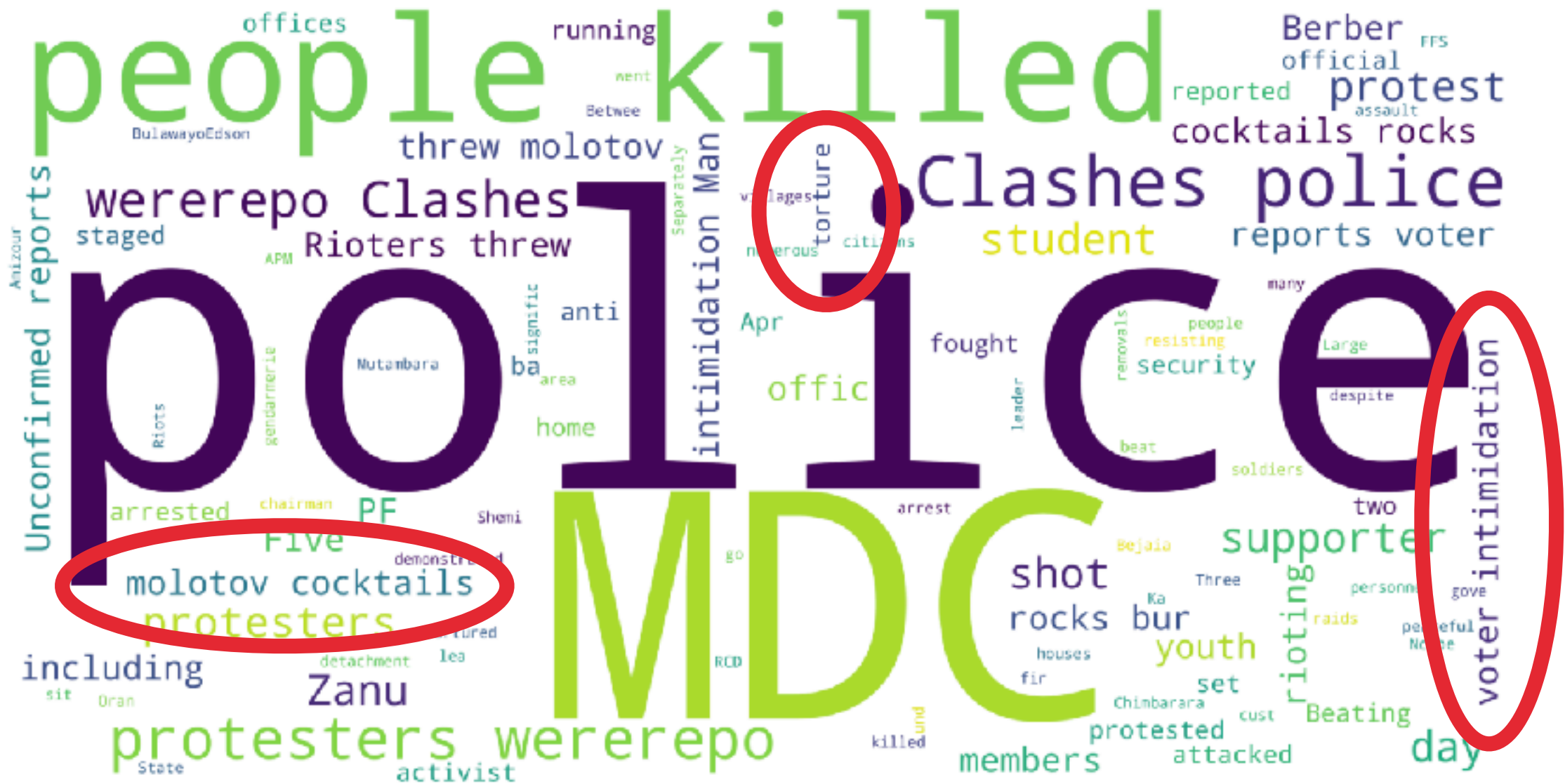
COUNTRIES NOT ON THE LIST

Countries without reported conflicts in the dataset:

- ▶ Cape Verde
- ▶ Comoros
- ▶ Seychelles
- ▶ Sao Tome



WHAT IS IN A WORD?

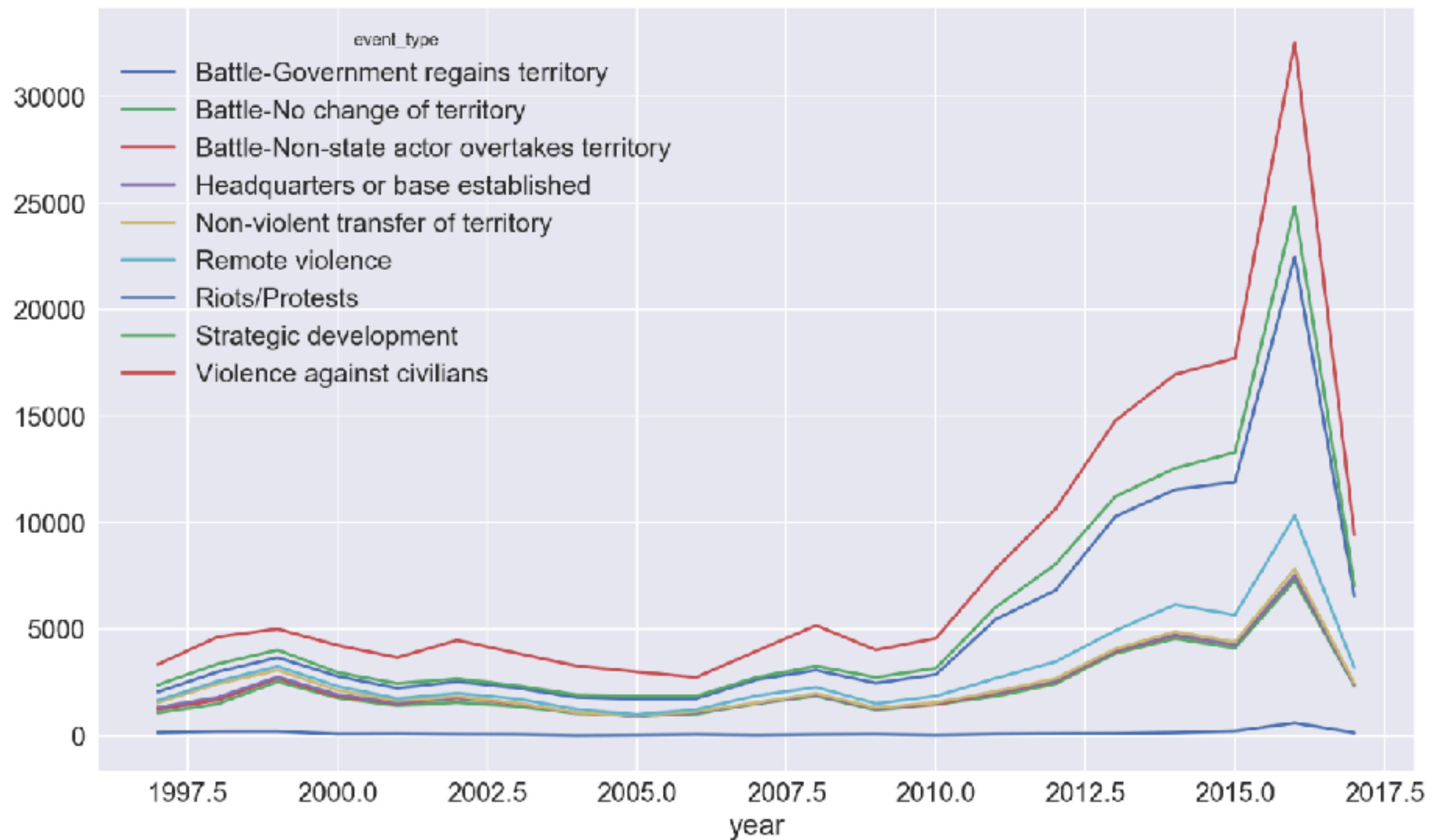


Most used words:

Killed, police, forces, protest, people, attack, government, soldiers, attacked

Sentiment: -0.8

EDA: EVENT TYPE



Are we just reporting things more?

WHO IS ENGAGED IN CONFLICT?



Government and security forces



Rebel Groups



Rioters



Political and Identity Militias



Protesters and Civilians

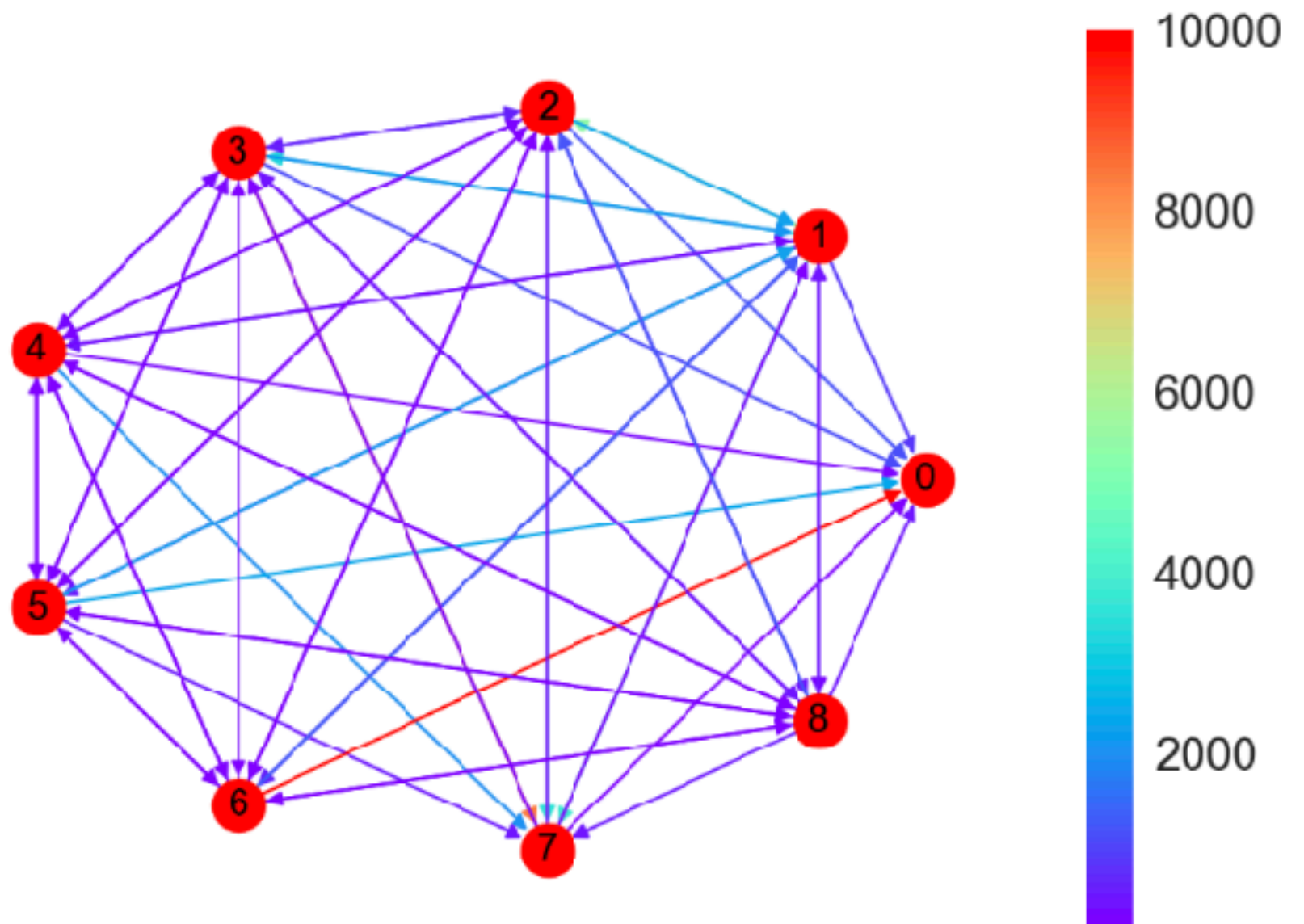


External Forces

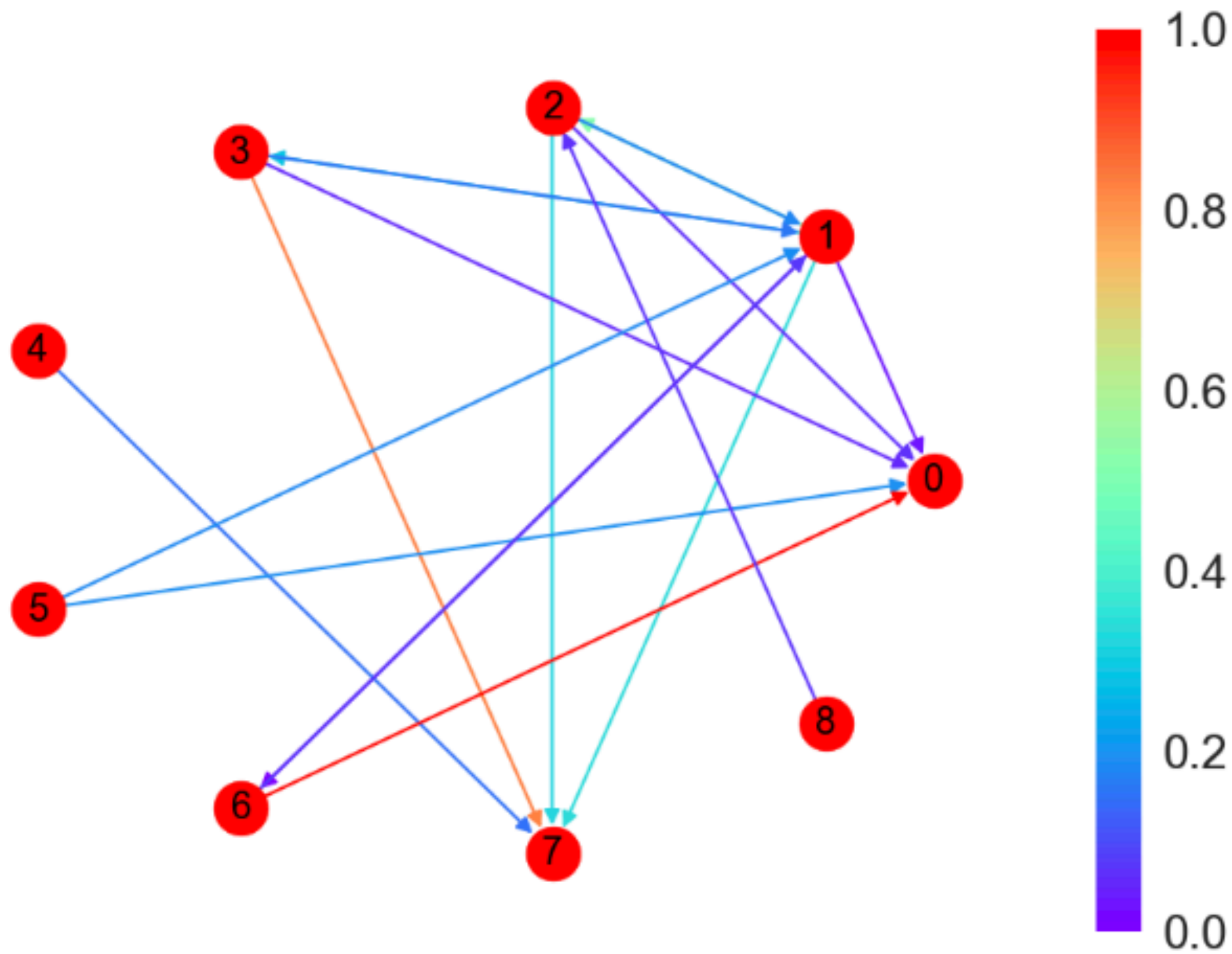
NETWORKX

Explanation:

- 1- Government and State Security
- 2- Rebel Groups
- 3- Political Militias
- 4- Identity Militias
- 5- Rioters
- 6- Protesters
- 7- Civilians
- 8- External/Other Forces

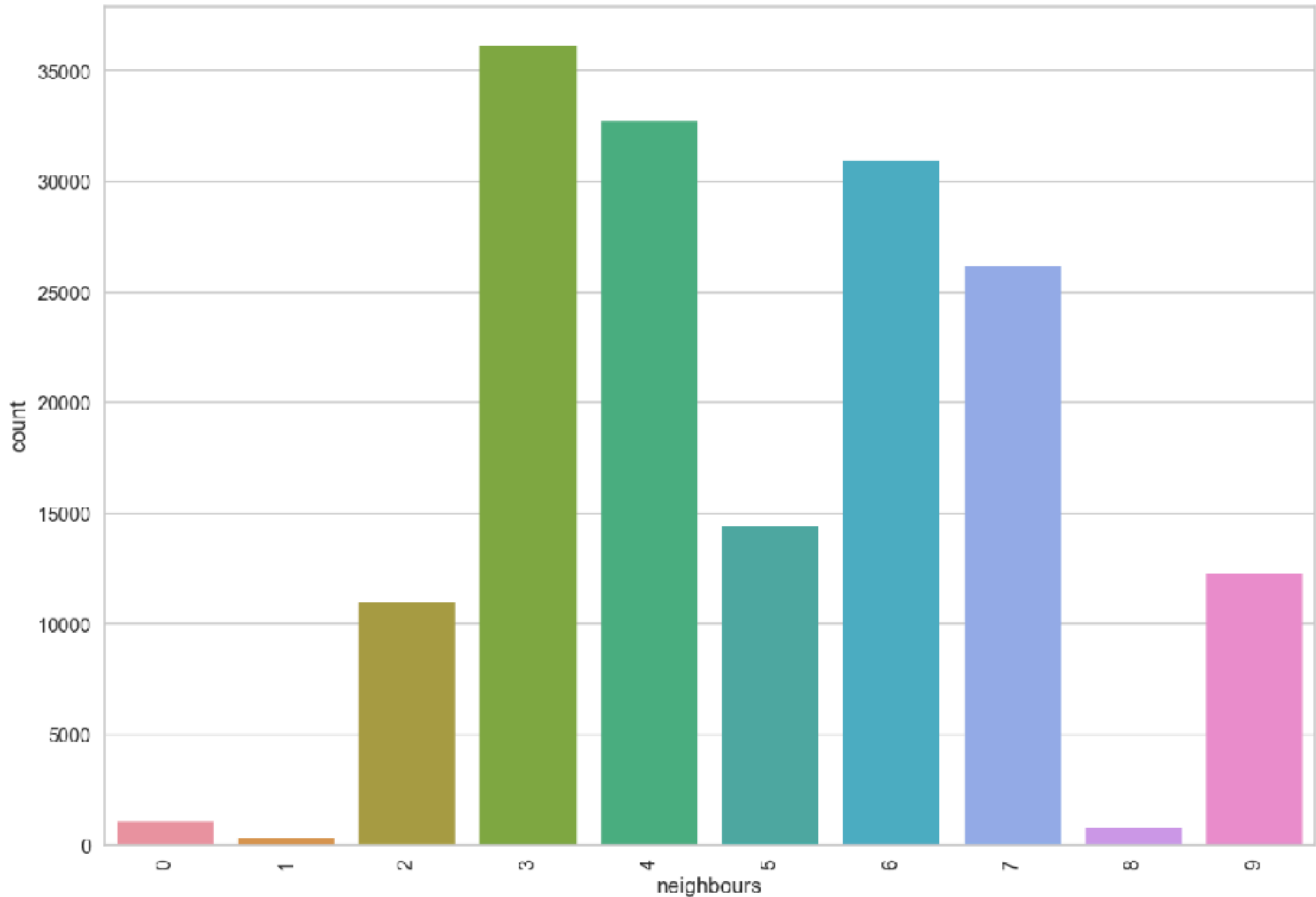


FEATURE ENGINEERING

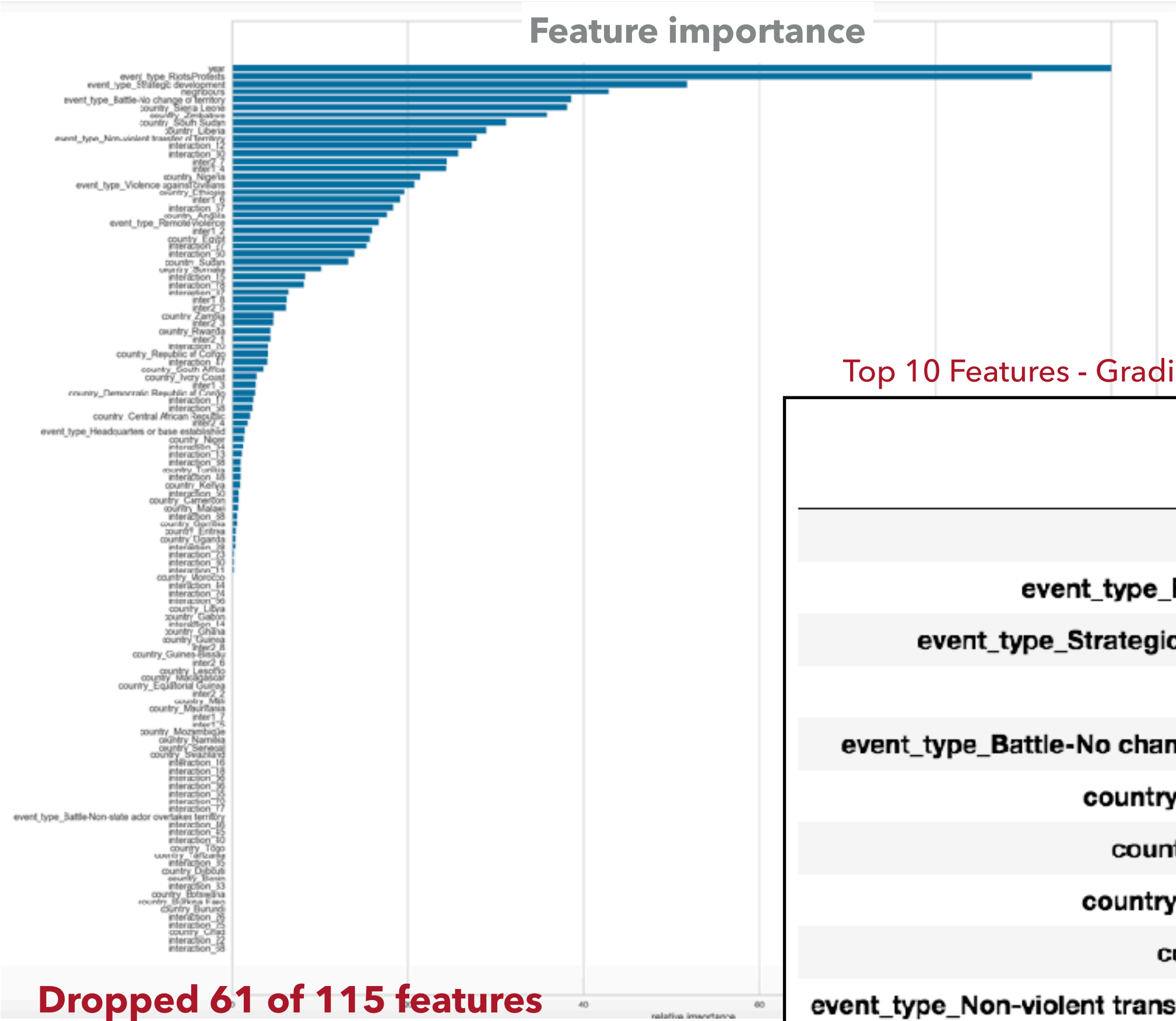


FEATURE ENGINEERING

Q: Does the number of neighbours a country has affect the likelihood of having a violent event?



FEATURE SELECTION



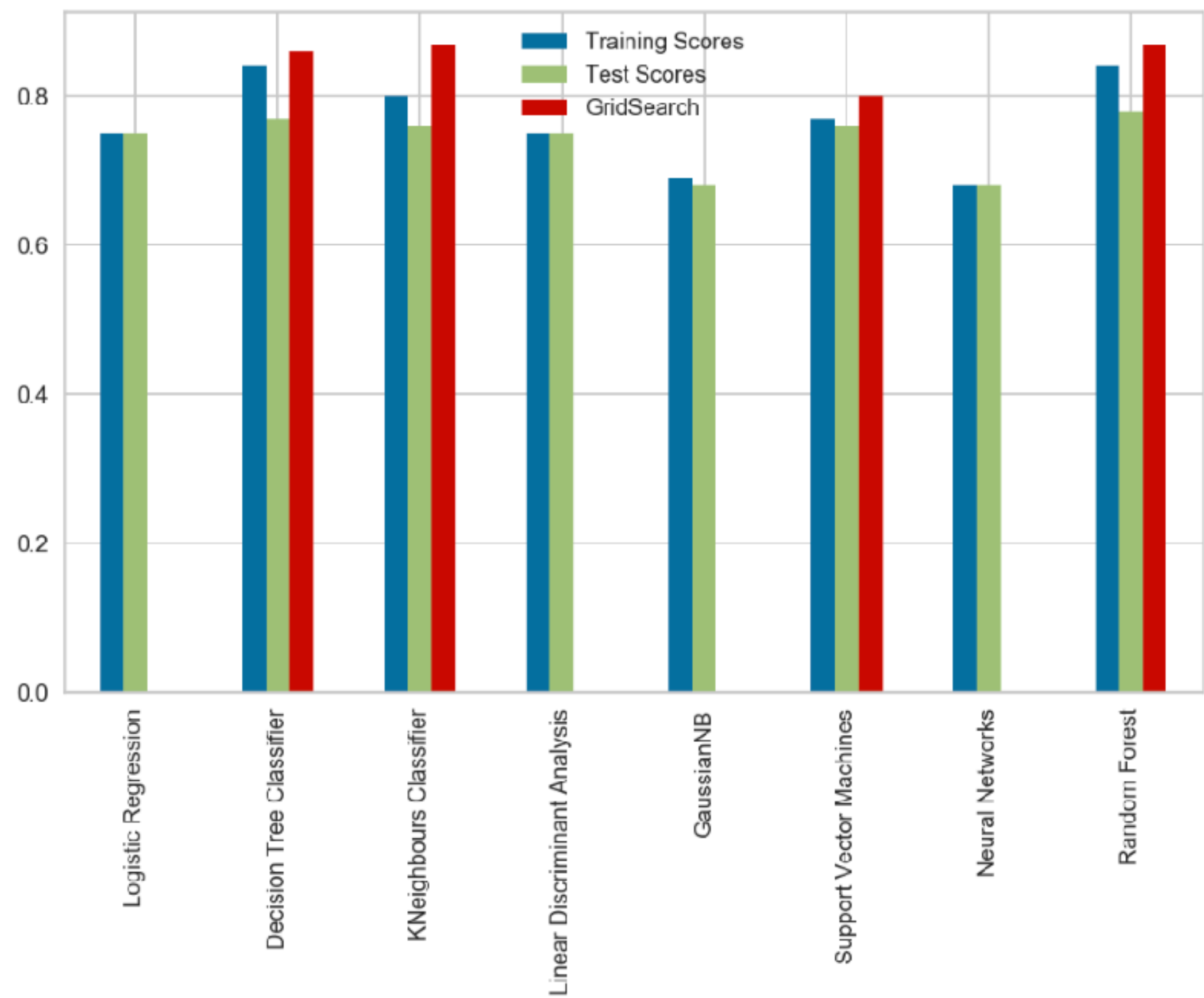
WHAT MODELS?

Baseline Accuracy: 0.68

Accuracy Scores

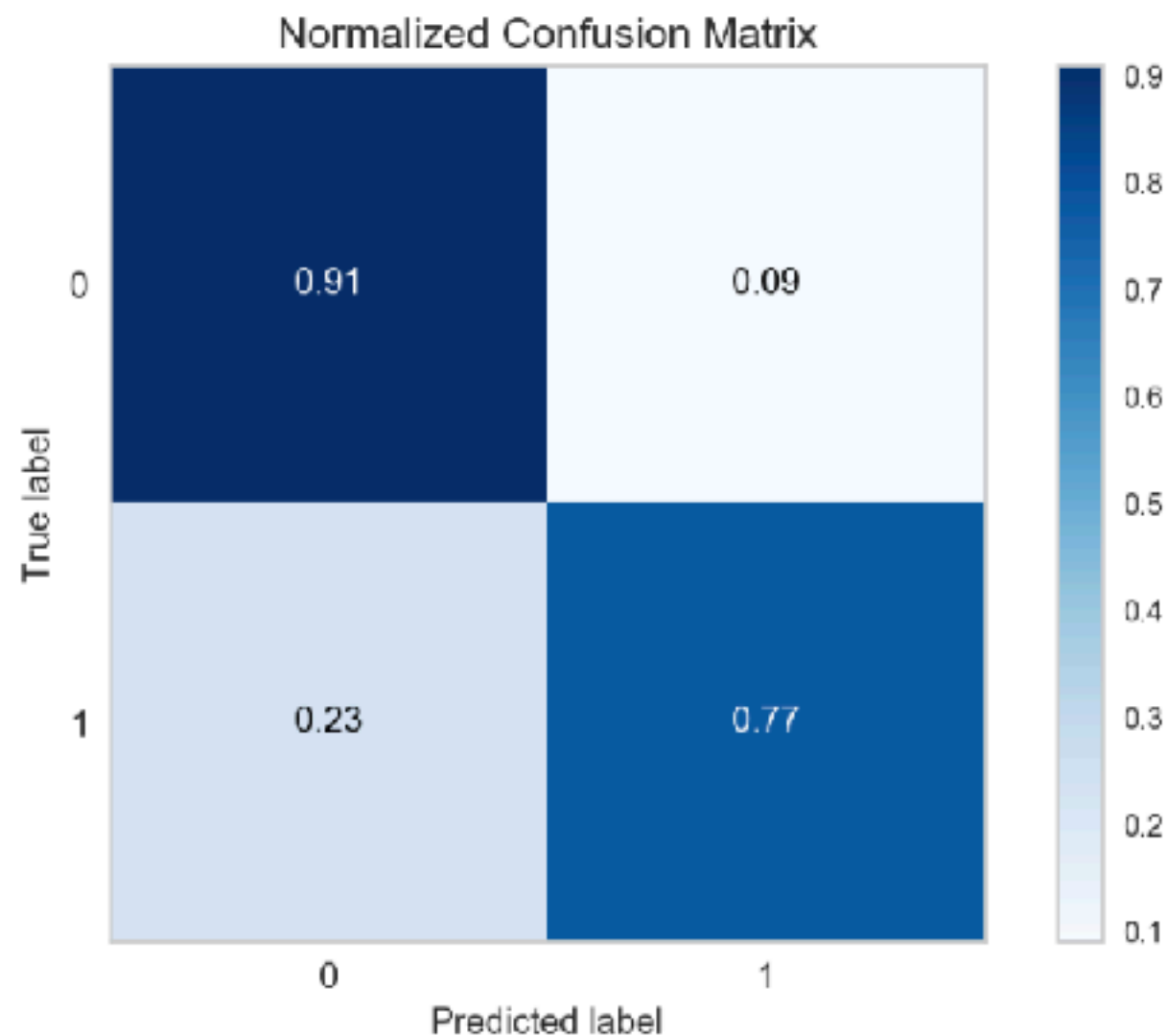
	Train	Test
Logistic Regression	0.75	0.75
Decision Tree	0.84	0.77
KNearest Neighbour	0.80	0.76
Linear D Analysis	0.75	0.75
GaussianNB	0.69	0.69
SVN	0.77	0.76
Neural Network	0.68	0.68
Random Forest	0.84	0.78

POST-GRIDSEARCH ACCURACY SCORES



MODELLING RESULTS – RANDOM FOREST

	precision	recall	f1-score	support
0	0.90	0.91	0.90	5660
1	0.80	0.77	0.79	2631
avg / total	0.87	0.87	0.87	8291

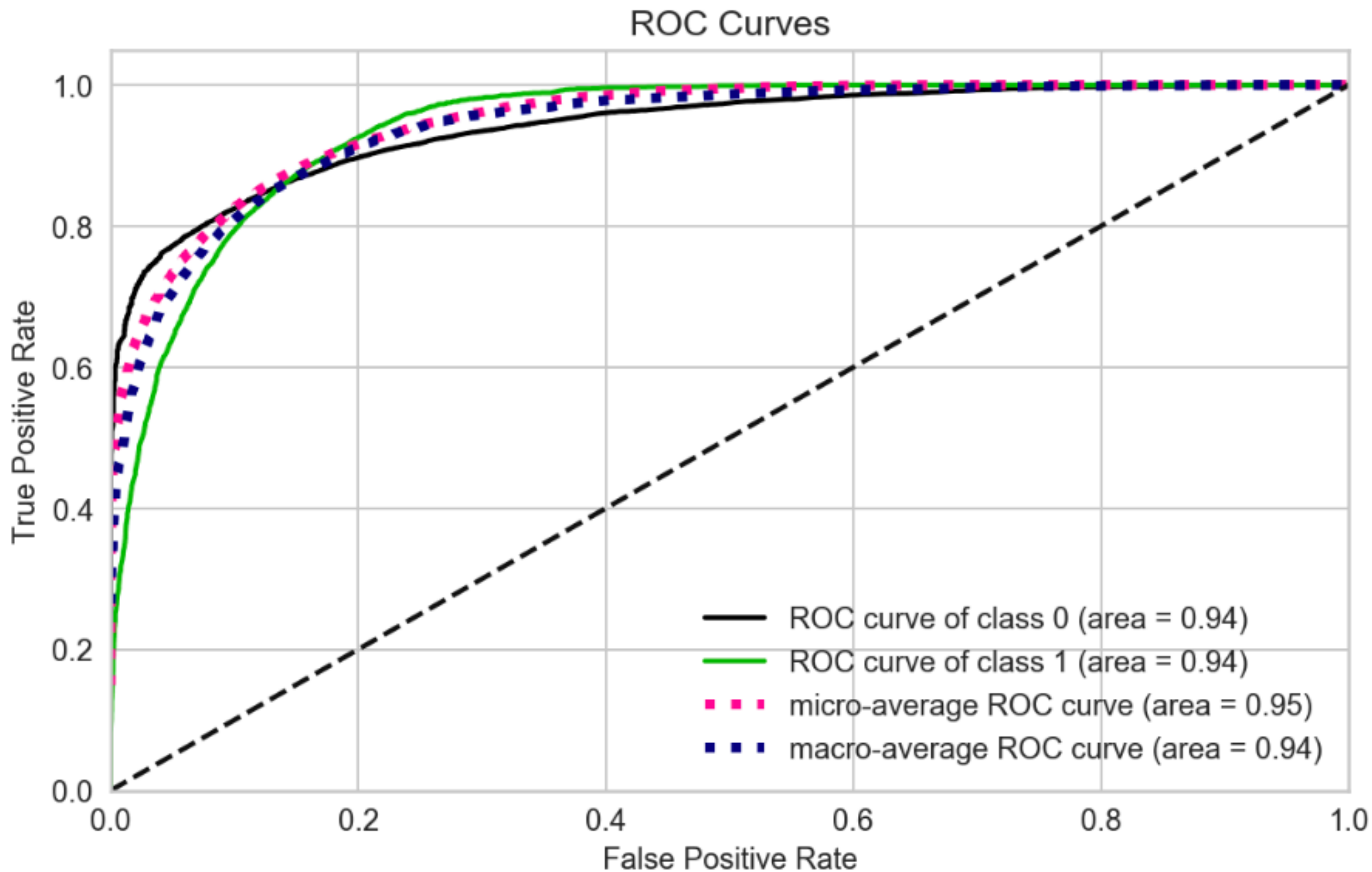


Q1

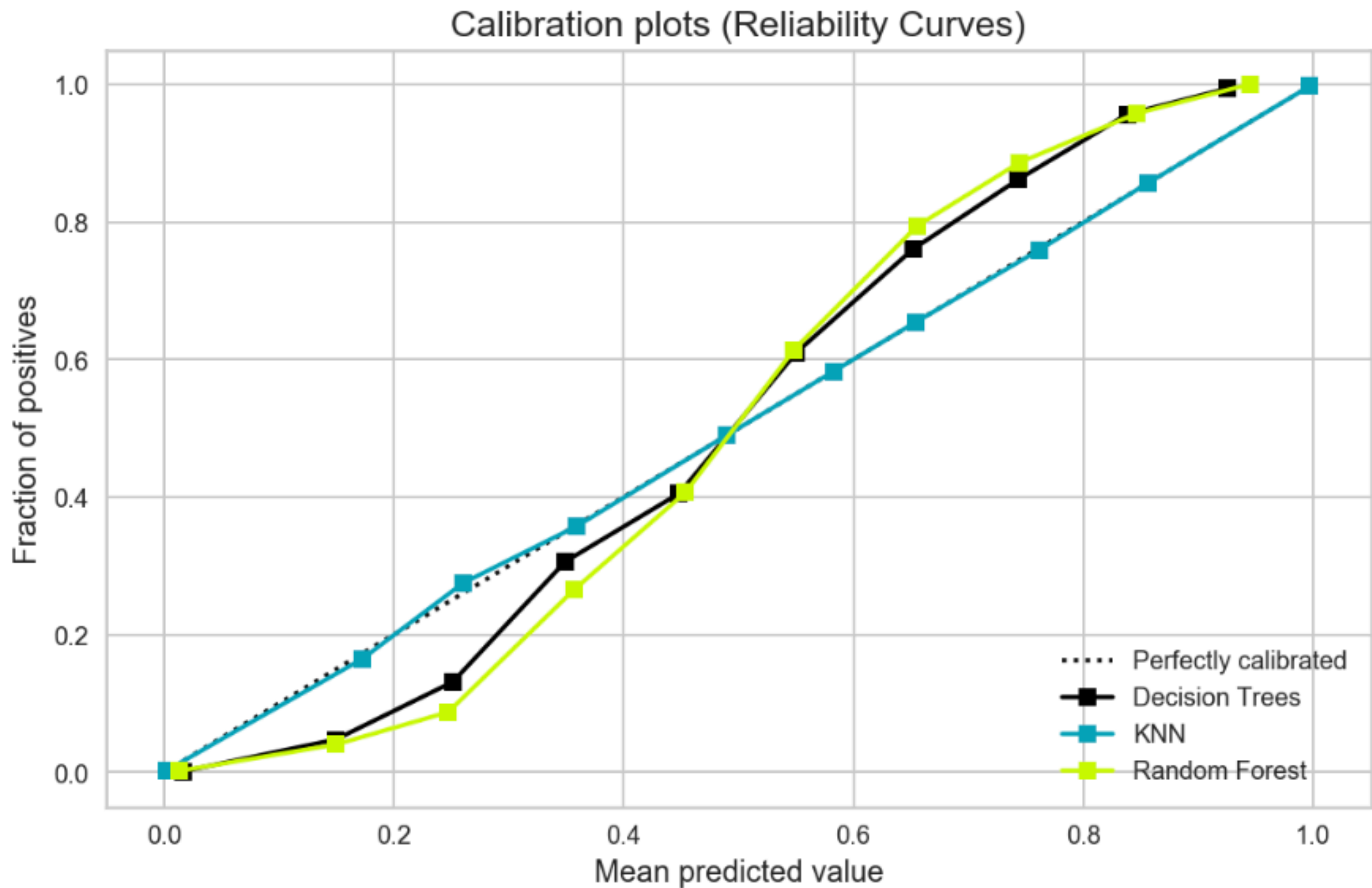
■ Predicting Project Success

- ~87% accuracy
- Low misclassification rate

MODELING



MODELING



RESULTS FROM MODEL

▶ **What are the features that contribute the most?**

- ▶ Specific Interactions:
 - ▶ Of the top 10 40% involve GF (M, RG,C)
 - ▶ Highest FI interaction between RG and C
- ▶ Certain event types (VAC, RV)
- ▶ Certain countries (CAF)
- ▶ Neighbours not as important as I expected

FORWARD PLAN:

- ▶ TSA
- ▶ Individual countries
- ▶ Add socio-economic indicators

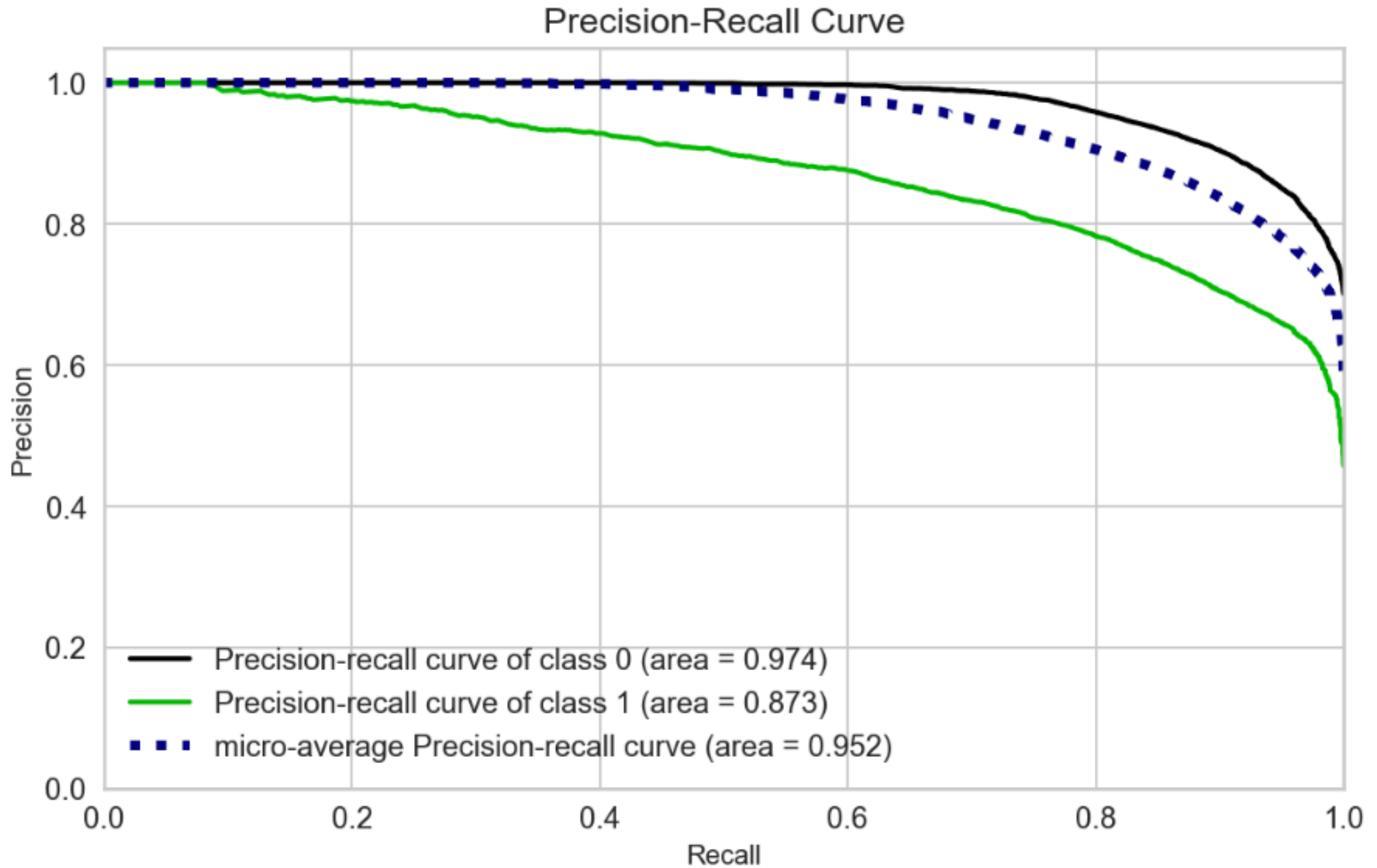
CHALLENGES:

- ▶ Data size and computing power

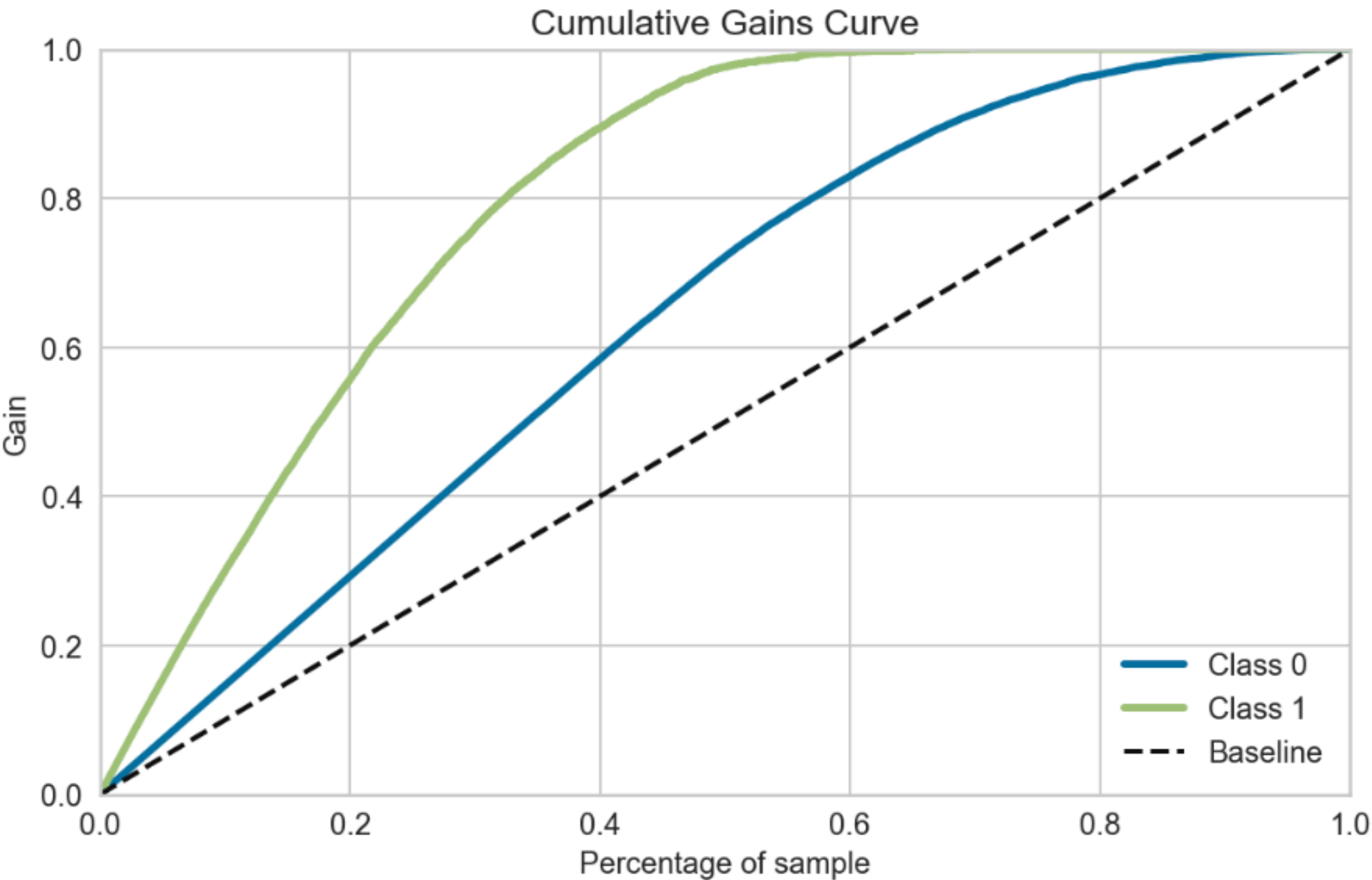
Questions?

BACKUP SLIDES

MODELING



MODELING



MODELING

