Predicting Protest Outcomes

A study of the Mass Mobilization dataset alongside World Governance Indicators

Terri John,
Data Scientist

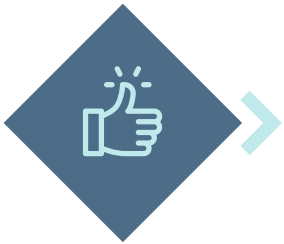
Problem Statement and Background

- Data Driven Decisions for the Department of State
  - Will a protest end in violence?
  - Should an alert be sent?
- What are the short and long term effects of protests?

The Data

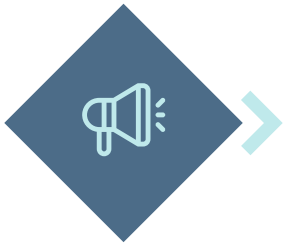
Mass Mobilization

Protests and their attributes



World Governance Indicators

From the WorldBank Databank



Freedom in the World Score

3 possible scores:
Free
Not Free
Partially Free

Methodology

1

Data Cleaning and Merging

Flaws in the data

2

Feature Engineering

Participants

Duration

3

Exploratory Data Analysis

Country Specific Case Studies

4

Modeling

Multiclass Classification

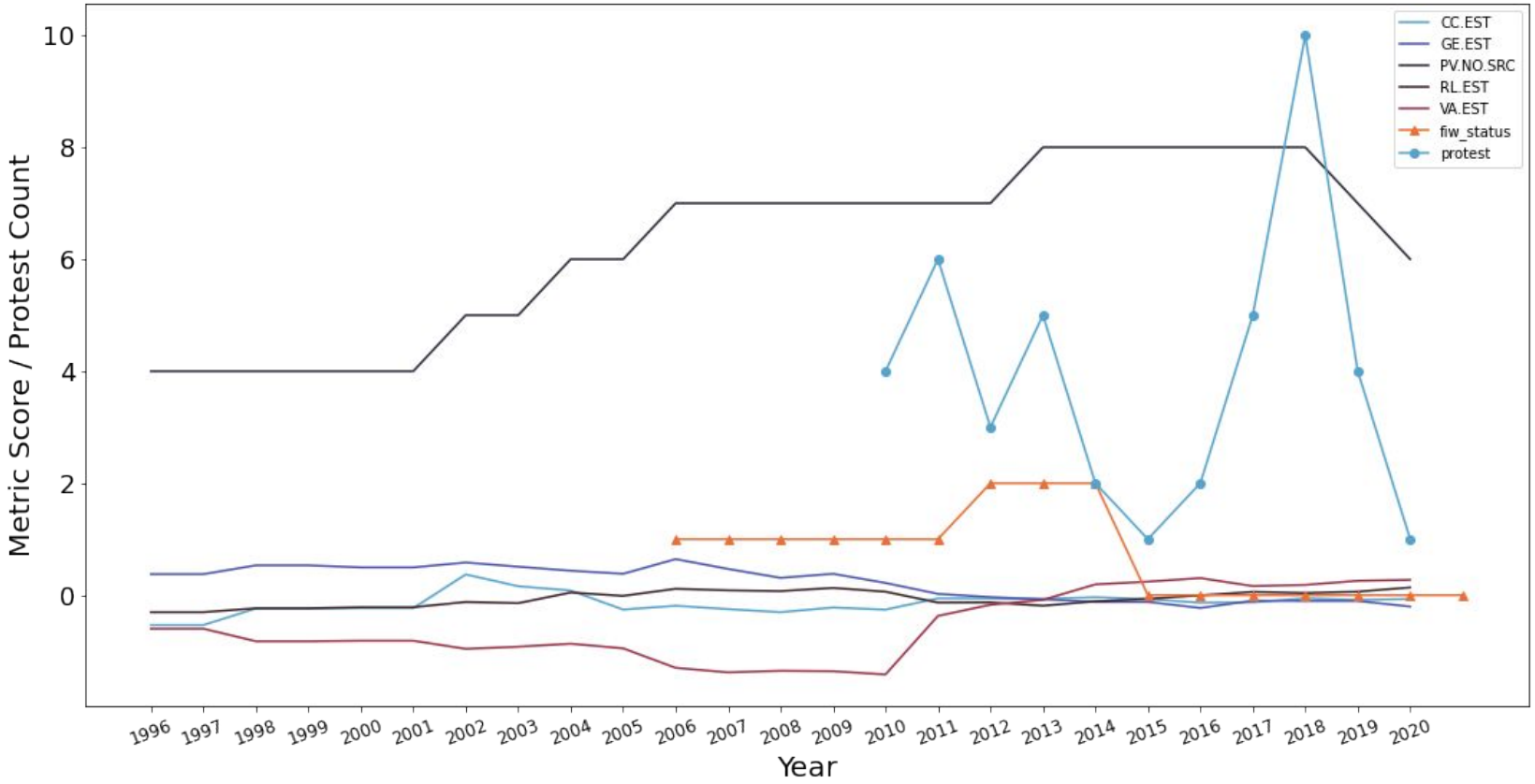
Binary Classification

Case Studies

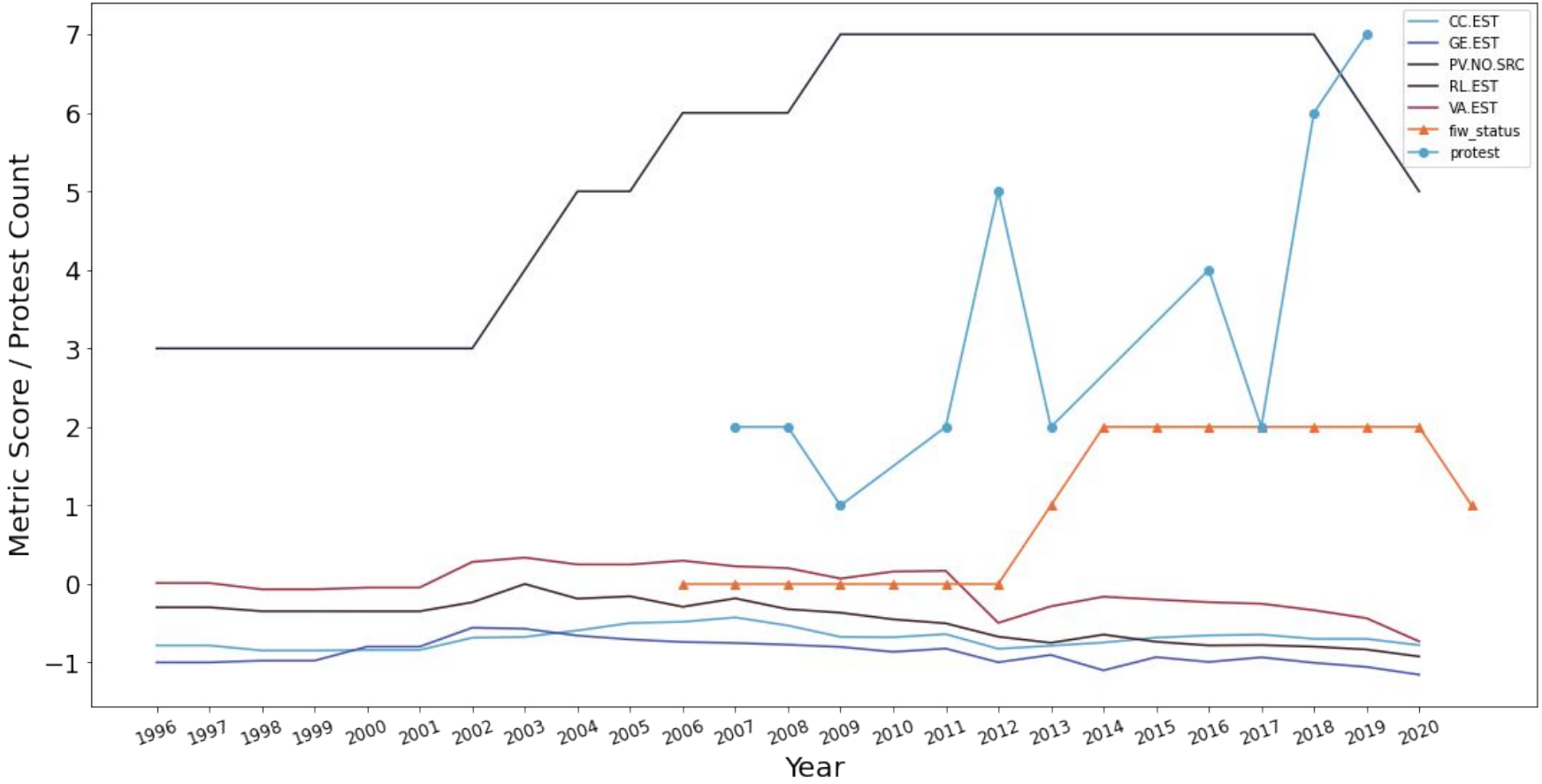
Tunisia

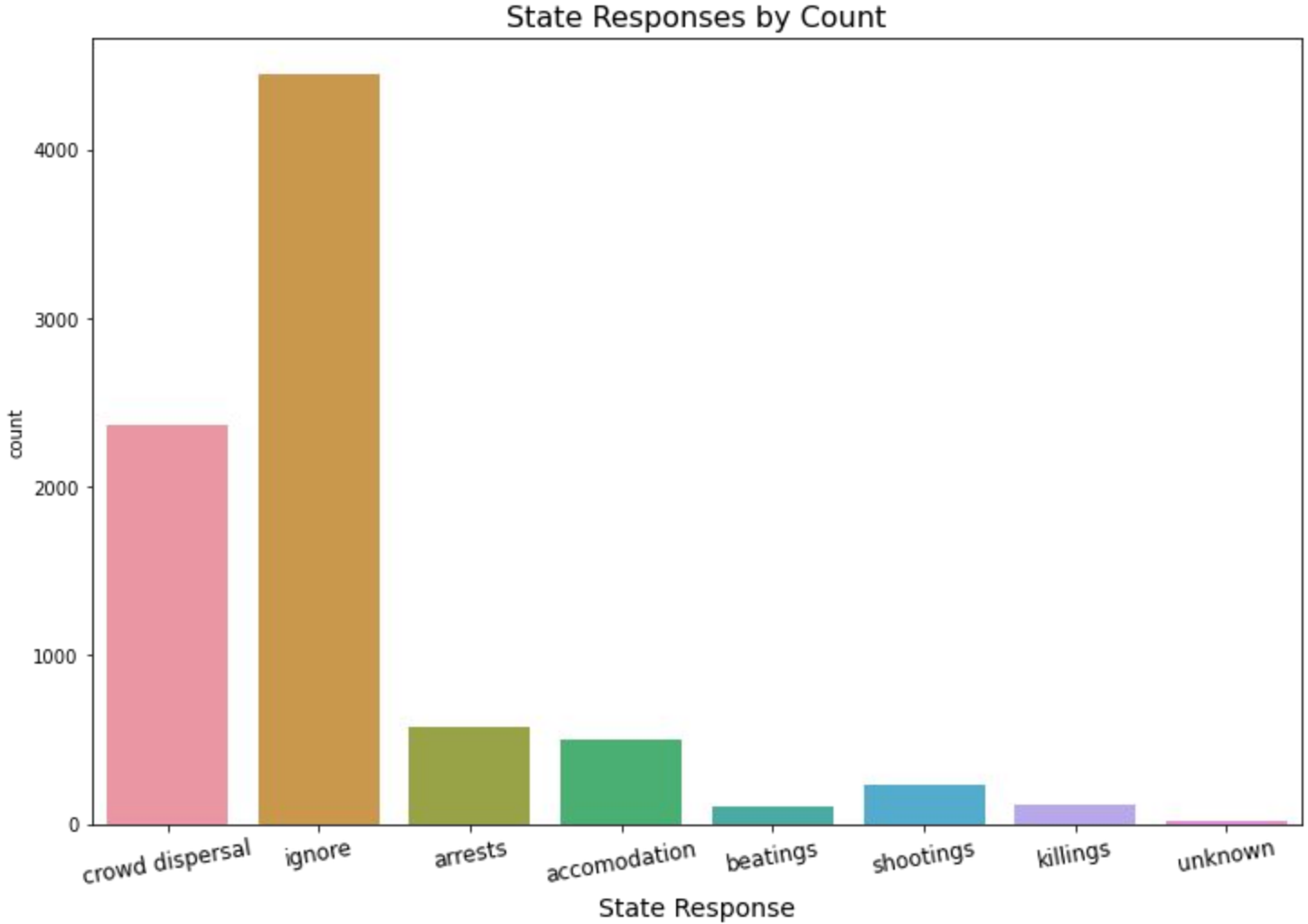
Mali

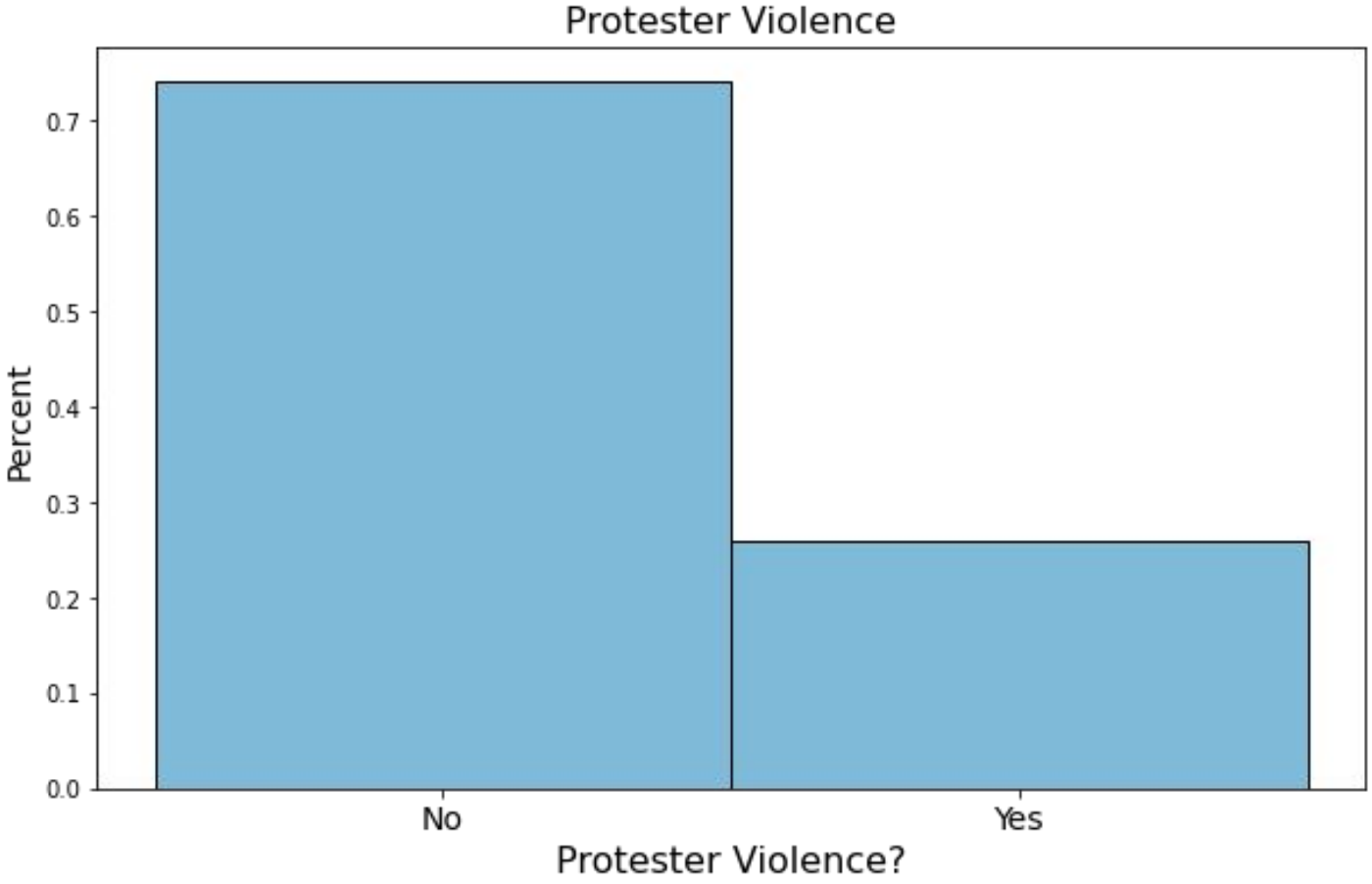
World Governance Indicators, Freedom Score, and Protest Count from 1996-2021: Tunisia



World Governance Indicators, Freedom Score, and Protest Count from 1996-2021: Mali




State Responses by Count
count
4000
3000
2000
1000
0
crowd dispersal
ignore
arrests
accomodation
beatings
shootings
killings
unknown
State Response


Protester Violence
Percent
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
No
Yes
Protester Violence?

A Tale of Two Models

Predicting State Response and Protester Violence

	Null Model	Train Accuracy	Test Accuracy	F1 Score
Binary Classification: Protester Violence KNN	74%	**82%**	**82%**	0.72
Multiclass Classification: State Response Stacked Model	53%	**70.3%**	**67%**	0.66

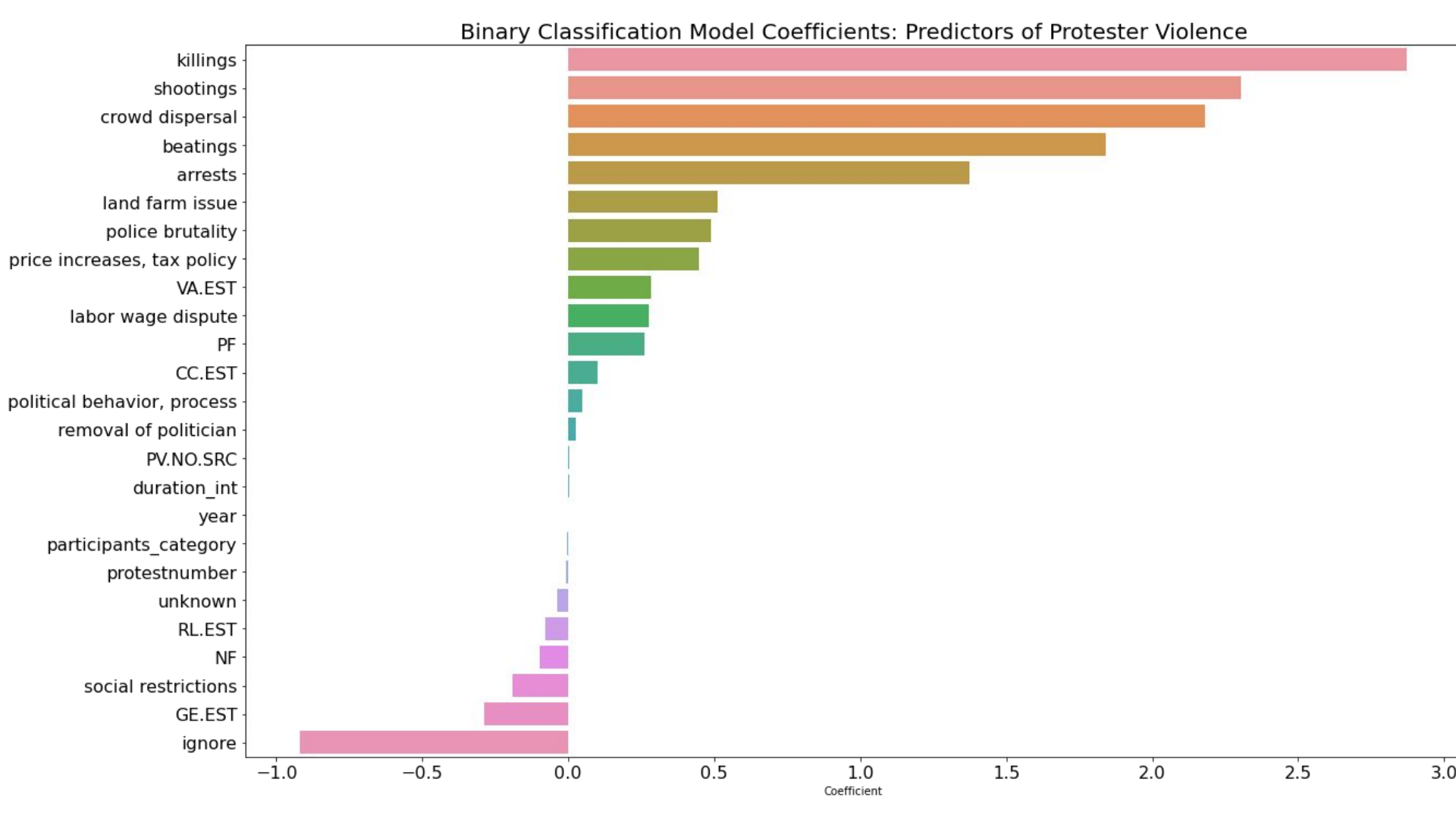
Binary Classification:

Predicting Protester Violence

Model	Hyperparameters	Test Score	Train Score	F1 Score
LogisticRegression	Default	80.6%	80.8%	.80
KNearestNeighbors	Default	82.9%	82.4%	0.72
DecisionTree	ccp_alpha=.0015	80.8%	80.5%	0.81
Bagging Model	Gridsearch max_features=2, max_samples=4, n_estimators=11	72.8%	72.0%	0.73
StackingClassifier		82.5%	79.6%	0.81

Multiclass Classification:
Predicting State Response

Model	Hyperparameters	Test Score	Train Score	F1 Score
DecisionTree	ccp_alpha=0.0012 min_samples_leaf=3 min_samples_split=3	68.4%	68.5%	0.62
BaggingModel	Gridsearch max_features=6, max_samples=6, n_estimators=10	52.3%	53.3%	0.71
SVC	C=.0001	53.3%	53.2%	0.37
StackingClassifier	KNN, Tree, RF	70%	67%	0.65


Binary Classification Model Coefficients: Predictors of Protester Violence
killings
shootings
crowd dispersal
beatings
arrests
land farm issue
police brutality
price increases, tax policy
VA.EST
labor wage dispute
PF
CC.EST
political behavior, process
removal of politician
PV.NO.SRC
duration_int
year
participants_category
protestnumber
unknown
RL.EST
NF
social restrictions
GE.EST
ignore
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Coefficient

Streamlit

Conclusion

These models can be used in conjunction with subject matter expertise to make data driven decisions regarding protest alerts overseas.

Recommendations and Next Steps

- Bring in more metrics to include in the modeling process.
- Perform time series modeling.
- Improve the data cleaning process to minimize loss of data.
- Deliberate ethics.

Sources:

Mass Mobilization Project Dataverse: https://dataverse.harvard.edu/file.xhtml?persistentId=doi:10.7910/DVN/HTTWYL/TJJZNG&version=5.0

Freedom House: https://freedomhouse.org/countries/freedom-world/scores

WorldBank Databank: https://databank.worldbank.org/home.aspx

Wikipedia: Tunisian Protests: https://en.wikipedia.org/wiki/2018_Tunisian_protests

Thanks!

Do you have any questions?

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution