PROJECT REPORT

Prediction of military conflicts in Africa using machine learning

UNT CSCE 5214

Poli Nemkova, William Baker, Brad Aiken, Nicolas Stencel

SOMALIA, Mogadishu, 2017

(credit: https://cco.ndu.edu/News/Article/1171839/continuity-and-change-in-war-and-conflict-in-africa/)

Abstract

Keeping peace is one of the oldest goals in a society. AI, as a new ubiquitous power, can offer its own contribution to possible solution. And this is what we are to examine in this this project.

Our goal is to create a model that will be able to predict military conflicts. We have four datasets with which to do this. A dataset that contains all the weapons transfers from one country to another for the past 50 years, a dataset that contains all the military exercises (war games) of each country for the past 50 years, an armed conflict location dataset, and a set of economic data with information about conflict outbreaks.

Because there is such a large amount of data, we will have to perform clustering and data visualization. In this way our model will be trained to discern which properties (features) are the most likely to lead to military conflicts. We will construct a front end website implemented with cloud technology that will display this information in a nice, elegant way. We are interested, if time allows, in having the user be able to select any country and our model being able to predict whether or not said country is likely to have any conflict in the near future.

Data Specification

The main dataset we've used in the implementation of this project has been the economic & conflict outbreak data, from [this tutorial](). We also added column 0/1 for indication participation in a current conflict.

For more specific research our project focused on the African region only. We data-engineered with pandas to create new features, e.g., change in GDP per year as well as adding Freedom score from Freedom House and number of Covid-19 deaths per country.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		Country	Year	Country Code	Debt	Account Balance	Corrupti	Foreign_Investment	GDP	Inflation	Ethnicity	Religion	unemploy	Natural Ressources	Infant Mortality
2	37	Uganda	1997	UGA	79.12	-5.036101675	26	2.791365385	2.07845297	8.169021441	0.9302	0.6332	2.099	16.85004386	96.2
3	38	Uganda	1998	UGA	75.51	-5.523161292	26	3.189155246	1.915019409	0.068804172	0.9302	0.6332	2.354	16.78556921	94.2
4	39	Uganda	1999	UGA	76.02	-4.887837094	26	2.337226332	4.939167233	5.777368987	0.9302	0.6332	2.707	12.45979269	91.7
5	40	Uganda	2000	UGA	77.58	-5.793407068	26	2.59476175	0.09886378	3.392021585	0.9302	0.6332	2.833	12.04225517	88.4
6	41	Uganda	2001	UGA	81.47	-5.232145995	26	2.593888371	1.997414455	1.865125241	0.9302	0.6332	3.156	12.41449556	84.7
7	42	Uganda	2002	UGA	89.69	-2.460116952	26	2.988527292	5.370163846	-0.287508512	0.9302	0.6332	3.5	14.19516969	80.6
8	43	Uganda	2003	UGA	90.7	-0.606561442	26	3.190820333	3.140220601	8.680476516	0.9302	0.6332	3.6	21.78688586	76.1
9	44	Uganda	2004	UGA	63.53	-0.021095009	26	3.72044058	3.459235393	3.72128744	0.9302	0.6332	2.645	15.24468246	71.5
10	45	Uganda	2005	UGA	55.22	0.54609486	26	4.213615704	3.016798329	8.448726423	0.9302	0.6332	1.9	14.15074085	67.1
11	46	Uganda	2006	UGA	36.02	-3.081937643	26	6.479820608	7.345864106	7.310676136	0.9302	0.6332	2.123	13.35730256	63.1
12	47	Uganda	2007	UGA	22.03	-4.487636842	26	6.44527613	5.048748955	6.138510833	0.9302	0.6332	2.256	15.6936079	59.3
13	48	Uganda	2008	UGA	20.31	-8.594963658	26	5.118755092	5.331585491	12.05085555	0.9302	0.6332	2.662	16.76201674	55.8
14	49	Uganda	2009	UGA	19.19	-5.768333588	26	4.631929853	3.471575151	13.01725619	0.9302	0.6332	3.6	13.09344179	53
15	50	Uganda	2010	UGA	22.94	-7.974507621	26	2.694240313	2.324401765	3.976552885	0.9302	0.6332	3.68	10.70826058	50.2
16	51	Uganda	2011	UGA	23.64	-10.34521618	26	4.432458062	5.964687795	16.56434962	0.9302	0.6332	3.595	12.65095327	47.7
17	52	Uganda	2012	UGA	24.16	-7.1626338	26	5.214905193	0.585699589	12.67903772	0.9302	0.6332	3.551	12.65883418	44.6
18	53	Uganda	2013	UGA	27.7	-7.492290077	26	4.45536594	0.290029688	4.905208757	0.9302	0.6332	1.91	12.15448262	42.7
19	54	Uganda	2014	UGA	31.21	-7.936381594	26	3.878679473	1.641212971	3.07570669	0.9302	0.6332	1.888	11.61607516	40.4
20	55	Uganda	2015	UGA	34.39	-6.166432729	26	2.721697434	1.573939518	5.589686063	0.9302	0.6332	1.87	12.55129466	38.6
21	56	Uganda	2016	UGA	42.650538	-3.442593016	26	2.592662078	1.022778127	5.706375046	0.9302	0.6332	1.792	14.62577428	36.8
22	57	Uganda	2017	UGA	46.054992	-5.736414272	26	3.087913666	0.040409572	5.209717059	0.9302	0.6332	1.698	13.85620466	35.2
23	58	Uganda	2018	UGA	46.341799	-9.339273772	26	4.869111556	2.274615022	2.61601191	0.9302	0.6332	1.742	13.85620466	33.8

	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	Ongoing Conflict	New Conflict	Country-Year	Delta_Debt	Delta_Account_Balance	Delta_Corruption	Delta_Investment	Delta_GDP	Delta_Inflation	Delta_Ethnicity	Delta_Religion	Delta_Unemployment	Delta_Natural_Resources	Delta_Infant_Mortality
2	1	0	Uganda-1997	3.42	-1.537102347	0	0.789573807	-3.780178828	0.977374837	0	0	0.271000028	-1.732424256	-1.4
3	1	0	Uganda-1998	-3.61	-0.487059617	0	0.39778986	-0.163433561	-8.100217269	0	0	0.255000114	-0.064474658	-2
4	1	0	Uganda-1999	0.51	0.635324199	0	-0.851928914	3.024147824	5.708564815	0	0	0.352999926	-4.325776512	-2.5
5	1	0	Uganda-2000	1.56	-0.905569974	0	0.257535418	-4.840303453	-2.385347402	0	0	0.125999928	-0.417537522	-3.3
6	1	0	Uganda-2001	3.89	0.561261073	0	-0.000873379	1.898550675	-1.526896344	0	0	0.322999954	0.372240387	-3.7
7	1	0	Uganda-2002	8.22	2.772029042	0	0.394638922	3.372749391	-2.152633753	0	0	0.344000101	1.780674125	-4.1
8	1	0	Uganda-2003	1.01	1.853555511	0	0.20229304	-2.229943245	8.967985027	0	0	0.099999905	7.591716176	-4.5
9	1	0	Uganda-2004	-27.17	0.585466432	0	0.529620247	0.319014792	-4.959189076	0	0	-0.954999924	-6.542203401	-4.6
10	1	0	Uganda-2005	-8.31	0.567189869	0	0.493175123	-0.442437065	4.727438983	0	0	-0.745000005	-1.093941614	-4.4
11	1	0	Uganda-2006	-19.2	-3.628032503	0	2.266204905	4.329065777	-1.138050287	0	0	0.22299993	-0.793438288	-4
12	1	0	Uganda-2007	-13.99	-1.405699199	0	-0.034544478	-2.297115151	-1.172165303	0	0	0.133000135	2.336305346	-3.8
13	1	0	Uganda-2008	-1.72	-4.107326816	0	-1.326521038	0.282836536	5.912344715	0	0	0.405999899	1.068408833	-3.5
14	1	0	Uganda-2009	-1.12	2.82663007	0	-0.486825239	-1.86001034	0.966400641	0	0	0.937999964	-3.668574946	-2.8
15	1	0	Uganda-2010	3.75	-2.206174032	0	-1.937689539	-1.147173386	-9.040703304	0	0	0.080000162	-2.385181213	-2.8
16	1	0	Uganda-2011	0.7	-2.370708555	0	1.738217748	3.64028603	12.58779673	0	0	-0.085000038	1.942692695	-2.5
17	0	0	Uganda-2012	0.52	3.182582376	0	0.782447132	-5.378988206	-3.885311895	0	0	-0.04399991	0.007880904	-3.1
18	1	1	Uganda-2013	3.54	-0.329656277	0	-0.759539253	-0.295669901	-7.773828966	0	0	-1.641000152	-0.504351559	-1.9
19	1	0	Uganda-2014	3.51	-0.444091517	0	-0.576686467	1.351183284	-1.829502067	0	0	-0.021999955	-0.538407455	-2.3
20	1	0	Uganda-2015	3.18	1.769948864	0	-1.156982039	-0.067273454	2.513979373	0	0	-0.018000007	0.935219496	-1.8
21	1	0	Uganda-2016	8.26053844	2.723839714	0	-0.129035356	-0.55116139	0.116688984	0	0	-0.077999949	2.074479623	-1.8
22	1	0	Uganda-2017	3.40445393	-2.293821256	0	0.495251588	-0.982368556	-0.496657987	0	0	-0.094000101	-0.769569626	-1.6
23	1	0	Uganda-2018	0.28680682	-3.6028595	0	1.78119789	2.234205451	-2.59370515	0	0	0.04400003	0	-1.4

Data cleaning & engineering

Struggle 1:

With the initial excitement about the amount of the available data came the struggle of picking the right features what would work best for our model within a realistic amount of time.

Over the course of this project we had three very different versions of our initial dataset. In fact, about 80% of the time of this project we spent on polishing dataset.

Design & Milestones

- **Data preprocessing & feature engineering:**
  Jupyter Python notebook,
  Google Colab,
  Pandas,
  Numpy,
  old but good hands typing.
- **Data visualization & feature importance, prediction:**
  Random forest model,
  Correlation (Pearson)
  Principal Component Analysis (PCA) for dimensionality reduction
  Sklearn
- **Front end:**
  Flask

Repository

Our GitHub repository for this project:

https://github.com/PoliNemkova/armed-conflicts-prediction

Random Forest & Correlation

As a way to find the most relevant features we performed Random Forest (RF) feature importance (Fig.1)

as well as Pearson's correlation (Fig.2).

```
Variable: Ongoing Conflict         Importance: 0.21
Variable: GDP                      Importance: 0.07
Variable: Infant Mortality         Importance: 0.06
Variable: Religion                 Importance: 0.05
Variable: unemployment             Importance: 0.05
Variable: Debt                     Importance: 0.04
Variable: Account Balance          Importance: 0.04
Variable: Corruption               Importance: 0.04
Variable: Foreign_Investment       Importance: 0.04
Variable: Natural Ressources       Importance: 0.04
Variable: Delta_Debt               Importance: 0.04
Variable: Delta_Investment         Importance: 0.04
Variable: Delta_Unemployment       Importance: 0.04
Variable: Delta_Natural_Resources Importance: 0.04
Variable: Year                     Importance: 0.03
Variable: Inflation                Importance: 0.03
Variable: Delta_Account_Balance Importance: 0.03
Variable: Delta_GDP                Importance: 0.03
Variable: Delta_Inflation          Importance: 0.03
Variable: Delta_Infant_Mortality Importance: 0.03
Variable: Ethnicity                Importance: 0.02
Variable: Delta_Corruption         Importance: 0.0
Variable: Delta_Ethnicity          Importance: 0.0
Variable: Delta_Religion           Importance: 0.0
```

Fig.1 RF feature importance

	Code	Year	Debt	Account Bal	Corruption	Foreign_Inve	GDP	Inflation	Ethnicity	Religion	unemployme	Natural Ress	Infant Morta	Ongoing Cor	New Conflict	Delta_Debt	Delta_Accou
Code	1	-0.00134	-0.027778	-0.175392	0.429517	0.008	0.049484	-0.075645	-0.997521	-0.505322	0.312009	-0.26161	-0.382885	-0.172115	-0.100728	0.02078	0.013521
Year	-0.00134	1	-0.41285	-0.061297	-0.005458	0.02412	-0.058552	-0.112315	0.004729	-0.013052	-0.143687	-0.009522	-0.418442	0.026745	0.054401	0.073012	0.003875
Debt	-0.027778	-0.41285	1	0.036426	-0.072731	0.031101	-0.096529	0.118118	0.030464	-0.052457	-0.01896	-0.048825	0.212738	-0.012853	-0.018243	-0.028224	0.001787
Account Bal	-0.175392	-0.061297	0.036426	1	0.149672	-0.350414	-0.150491	-0.007323	0.17159	0.209832	0.122363	-0.25696	-0.132447	0.086395	0.029293	0.012562	0.124023
Corruption	0.429517	-0.005458	-0.072731	0.149672	1	0.007093	0.009564	-0.130118	-0.429519	-0.147577	0.221586	-0.573814	-0.50152	-0.30847	-0.128216	0.063579	0.008154
Foreign_Inve	0.008	0.02412	0.031101	-0.350414	0.007093	1	0.138662	0.14125	0.004803	-0.082257	-0.041229	0.204107	0.026905	-0.080696	-0.055318	-0.136027	-0.139776
GDP	0.049484	-0.058552	-0.096529	-0.150491	0.009564	0.138662	1	0.002058	-0.050207	-0.045007	-0.002262	0.176515	0.025074	-0.031315	-0.06548	-0.16121	0.112951
Inflation	-0.075645	-0.112315	0.118118	-0.007323	-0.130118	0.14125	0.002058	1	0.071695	0.136085	0.132137	0.130213	0.233649	0.163365	0.037401	-0.103961	-0.028923
Ethnicity	-0.997521	0.004729	0.030464	0.17159	-0.429519	0.004803	-0.050207	0.071695	1	0.501246	-0.321116	0.258759	0.377311	0.174388	0.099663	-0.026843	-0.013158
Religion	-0.505322	-0.013052	-0.052457	0.209832	-0.147577	-0.082257	-0.045007	0.136085	0.501246	1	-0.04662	-0.012241	0.341705	0.023051	0.017358	-0.018516	0.01624
unemployme	0.312009	-0.143687	-0.01896	0.122363	0.221586	-0.041229	-0.002262	0.132137	-0.321116	-0.04662	1	-0.025156	-0.204796	-0.125358	-0.046011	0.026197	0.014056
Natural Ress	-0.26161	-0.009522	-0.048825	-0.25696	-0.573814	0.204107	0.176515	0.130213	0.258759	-0.012241	-0.025156	1	0.237398	0.197047	0.111628	-0.086922	0.018678
Infant Morta	-0.382885	-0.418442	0.212738	-0.132447	-0.50152	0.026905	0.025074	0.233649	0.377311	0.341705	-0.204796	0.237398	1	0.146234	0.098171	-0.092589	0.001333
Ongoing Cor	-0.172115	0.026745	-0.012853	0.086395	-0.30847	-0.080696	-0.031315	0.163365	0.174388	0.023051	-0.125358	0.197047	0.146234	1	0.471411	0.041825	-0.034886
New Conflict	-0.100728	0.054401	-0.018243	0.029293	-0.128216	-0.055318	-0.06548	0.037401	0.099663	0.017358	-0.046011	0.111628	0.098171	0.471411	1	0.046255	-0.027963
Delta_Debt	0.02078	0.073012	-0.028224	0.012562	0.063579	-0.136027	-0.16121	-0.103961	-0.026843	-0.018516	0.026197	-0.086922	-0.092589	0.041825	0.046255	1	-0.055718
Delta_Accou	0.013521	0.003875	0.001787	0.124023	0.008154	-0.139776	0.112951	-0.028923	-0.013158	0.01624	0.014056	0.018678	0.001333	-0.034886	-0.027963	-0.055718	1
Delta_Corruption																	
Delta_Invest	-0.025359	-0.007925	0.05126	0.079323	0.018087	0.353966	-0.285561	0.022866	0.025895	0.025137	0.008216	-0.102413	-0.017753	0.008767	-0.021031	-0.007591	-0.220217
Delta_GDP	-0.005736	0.002356	-0.004221	0.065759	0.015123	-0.061005	0.618599	-0.01952	0.005745	0.008101	-0.003294	0.006156	-0.009004	-0.042328	-0.078537	-0.124933	0.179787
Delta_Inflatic	0.022887	0.059446	-0.03227	0.003555	0.040199	0.000555	-0.008266	-0.348511	-0.021449	-0.018928	-0.052562	-0.05314	-0.088313	-0.054331	0.004327	0.04478	0.265615
Delta_Ethnicity																	
Delta_Religion																	
Delta_Unem	-0.035202	0.023456	0.011478	-0.084783	-0.014112	-0.006295	-0.040825	0.071146	0.036059	-0.008178	-0.063516	-0.06648	0.062686	0.018186	0.04397	0.027319	0.020884
Delta_Natura	-0.010774	-0.031188	0.012094	0.100831	0.022753	0.047385	0.112984	0.054903	0.01189	0.001297	0.025643	0.116933	0.005996	-0.042981	-0.024753	-0.127213	0.326176
Delta_Infant	0.276171	0.131421	-0.131494	0.084967	0.313013	-0.020672	-0.041076	-0.12531	-0.280677	-0.212359	0.316154	-0.196798	-0.444221	-0.063835	-0.016097	0.175979	0.006378

Fig. 2. Correlation and p-values (nothing was identified as significant)

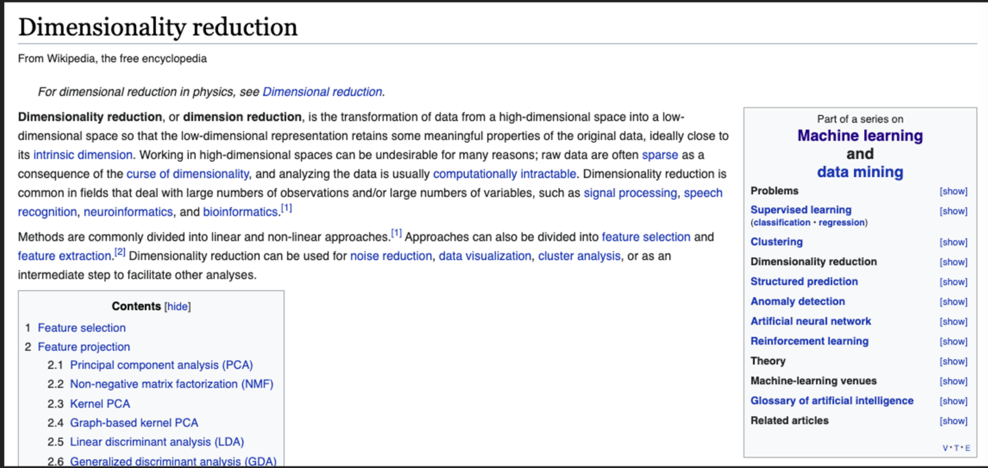
Dimensionality Reduction

As a part of dimensionality reduction we used Principal Component Analysis (PCA) to reduce dimensionality and be able to plot the dataset on 2- and 3-dimensional plot.

The dataset for this part included Arfican region in 2018 year.

First, you can see the initial data plotting (next slide).

Than you can see the plots after applying PSA.



Dimensionality reduction

From Wikipedia, the free encyclopedia

For dimensional reduction in physics, see Dimensional reduction.

Dimensionality reduction, or **dimension reduction**, is the transformation of data from a high-dimensional space into a low-dimensional space so that the low-dimensional representation retains some meaningful properties of the original data, ideally close to its intrinsic dimension. Working in high-dimensional spaces can be undesirable for many reasons; raw data are often sparse as a consequence of the curse of dimensionality, and analyzing the data is usually computationally intractable. Dimensionality reduction is common in fields that deal with large numbers of observations and/or large numbers of variables, such as signal processing, speech recognition, neuroinformatics, and bioinformatics.[1]

Methods are commonly divided into linear and non-linear approaches.[1] Approaches can also be divided into feature selection and feature extraction.[2] Dimensionality reduction can be used for noise reduction, data visualization, cluster analysis, or as an intermediate step to facilitate other analyses.

Contents [hide]

Part of a series on
Machine learning and data mining

Problems [show]
Supervised learning (classification • regression) [show]
Clustering [show]
Dimensionality reduction [show]
Structured prediction [show]
Anomaly detection [show]
Artificial neural network [show]
Reinforcement learning [show]
Theory [show]
Machine-learning venues [show]
Glossary of artificial intelligence [show]
Related articles [show]
V • T • E

Initial data plotting *(legend is available on the next slide)*

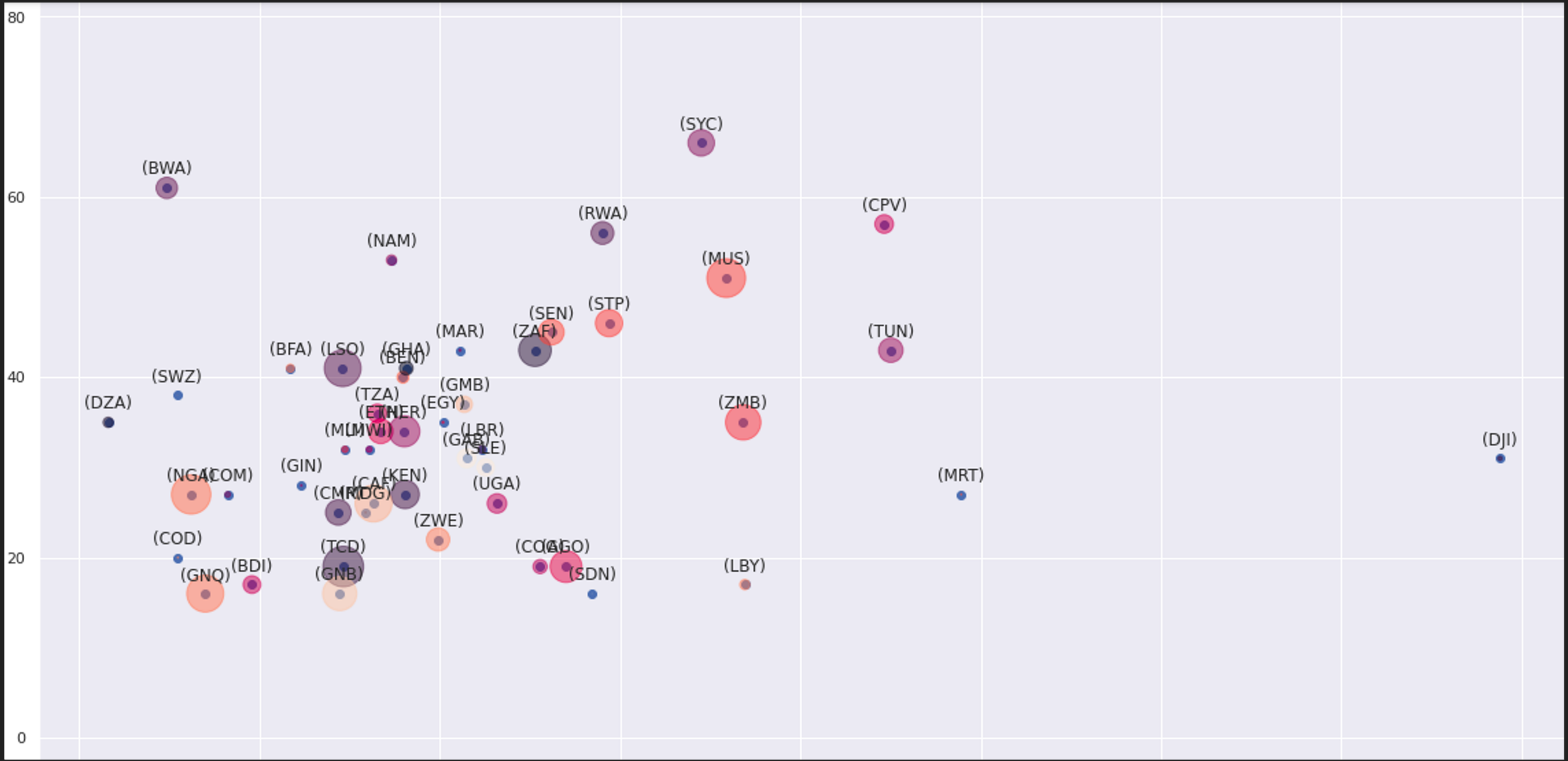


Fig.3 Countries abbreviation legend

1	Uganda	UGA	26	Benin	BEN
2	Liberia	LBR	27	Angola	AGO
3	Madagascar	MDG	28	Gambia, The	GMB
4	Congo, Dem.	COD	29	Zambia	ZMB
5	Congo, Rep.	COG	30	Gabon	GAB
6	Cameroon	CMR	31	South Africa	ZAF
7	Chad	TCD	32	Guinea	GIN
8	Kenya	KEN	33	Burkina Faso	BFA
9	Nigeria	NGA	34	Tanzania	TZA
10	Central Africa	CAF	35	Ethiopia	ETH
11	Cote d'Ivoire	CIV	36	Sudan	SDN
12	Sierra Leone	SLE	37	Togo	TGO
13	Guinea-Bissa	GNB	38	Senegal	SEN
14	Djibouti	DJI	39	Mali	MLI
15	Libya	LBY	40	Malawi	MWI
16	Algeria	DZA	41	Ghana	GHA
17	Rwanda	RWA	42	Niger	NER
18	Burundi	BDI	43	Namibia	NAM
19	Lesotho	LSO	44	Mauritania	MRT
20	Seychelles	SYC	45	Morocco	MAR
21	Egypt, Arab R	EGY	46	Mauritius	MUS
22	Eswatini	SWZ	47	Cabo Verde	CPV
23	Tunisia	TUN	48	Botswana	BWA
24	Comoros	COM	49	Zimbabwe	ZWE
25	Sao Tome an	STP	50	Equatorial Gu	GNQ

Fig.6 2D scatter plot after PSA with 2 components

Graph is interactive, and the color range represent Freedom score (feature)

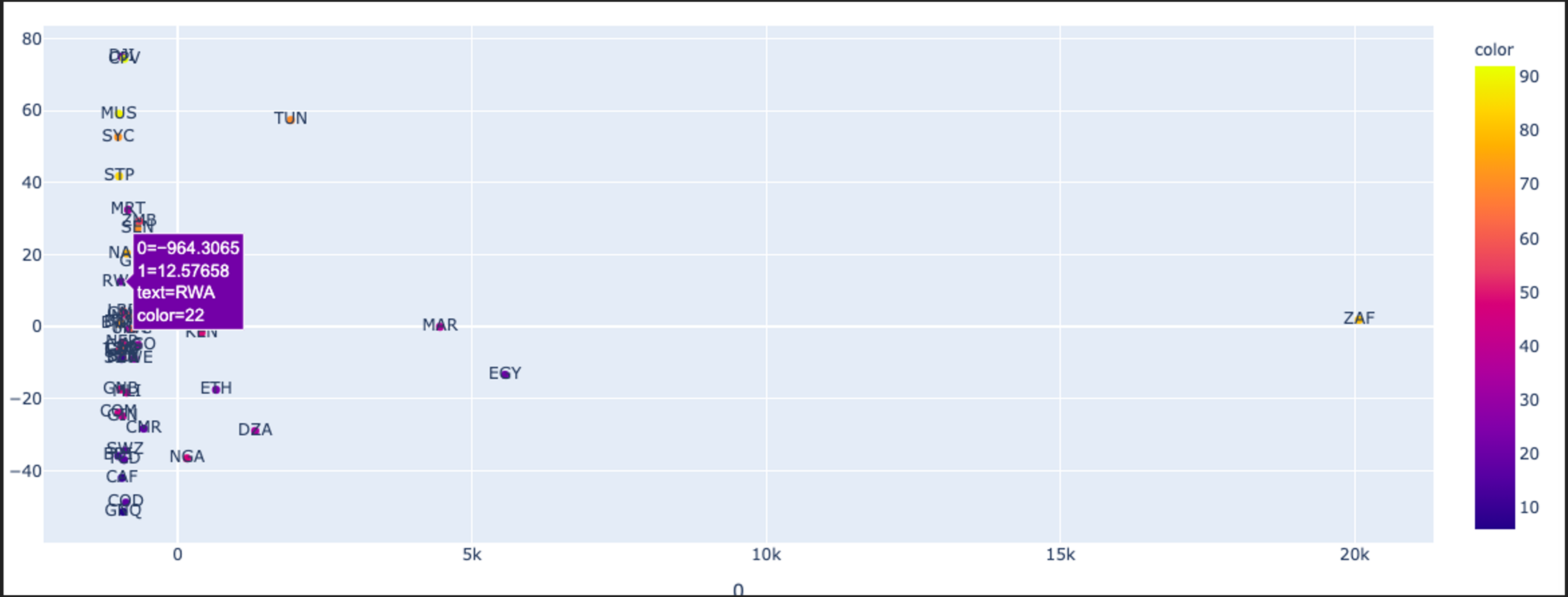
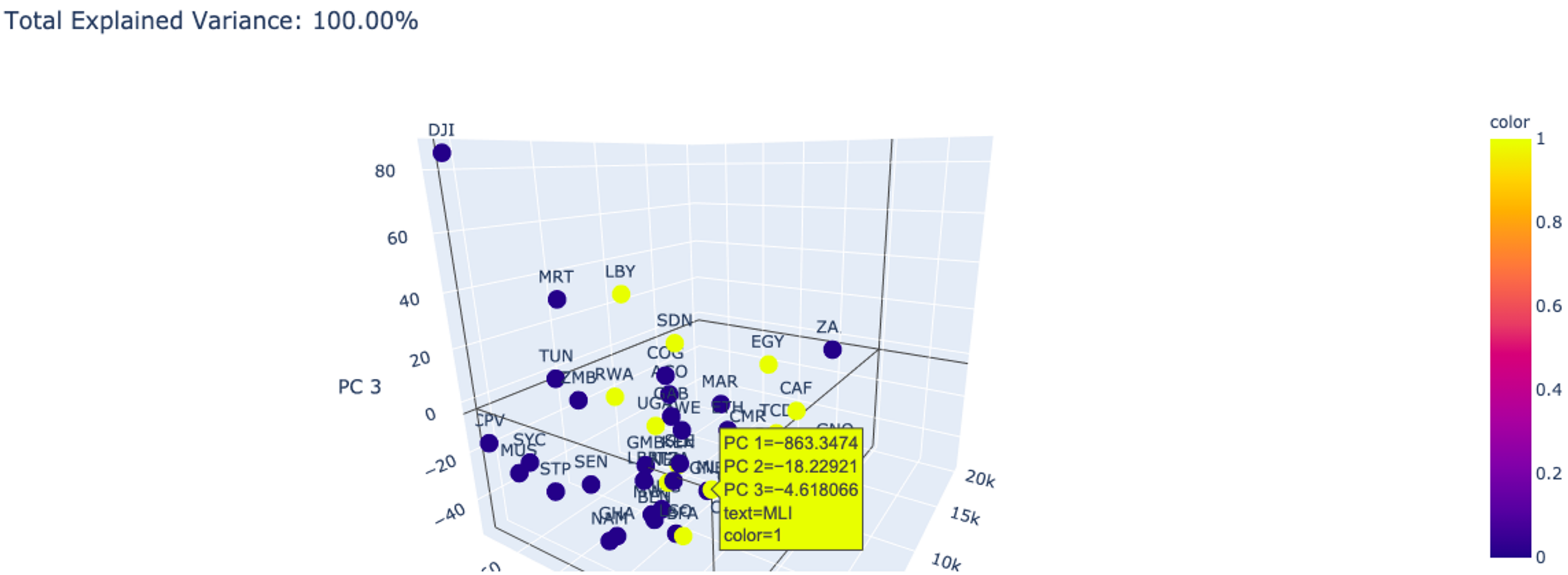


Fig.7 3D scatter plot after PSA with 3 components

Graph is interactive, Color range: yellow - in conflict, blue - peaceful.

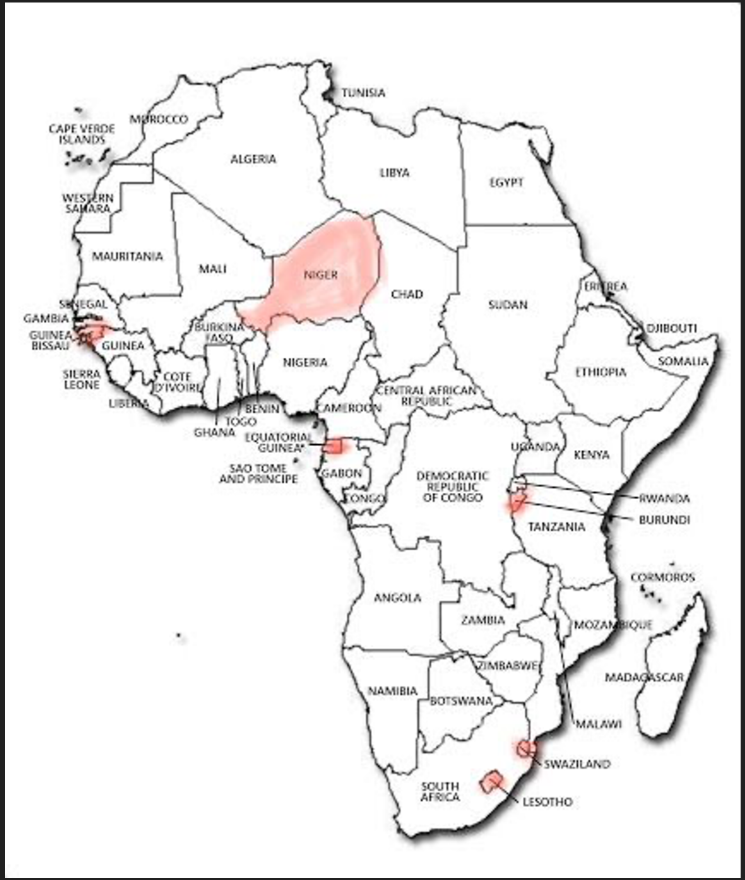


Suggestions from PSA graphs

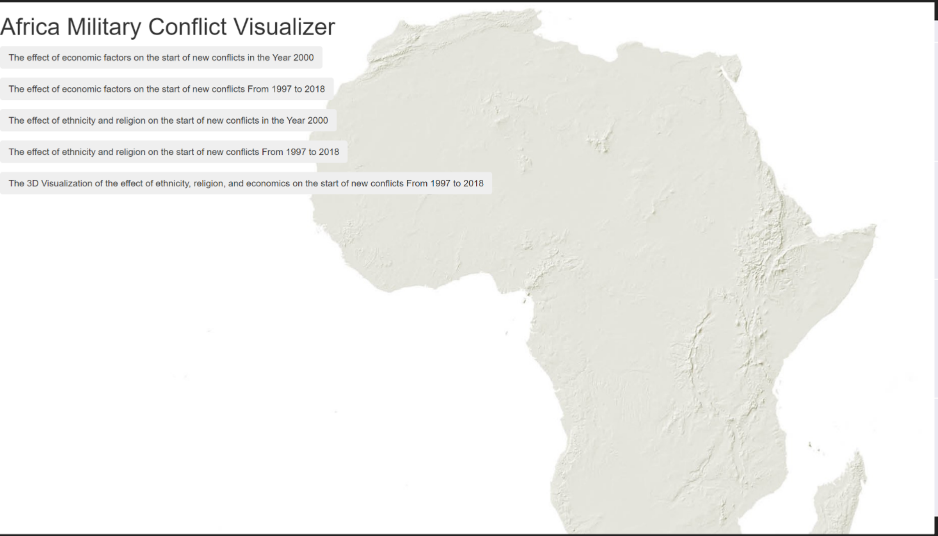
When examining 3D interactive graph we found that some points that represent peaceful states are closer to other which are in conflict. Hence, we suggest that those peaceful stated may stop being so the in coming years.

This list includes:

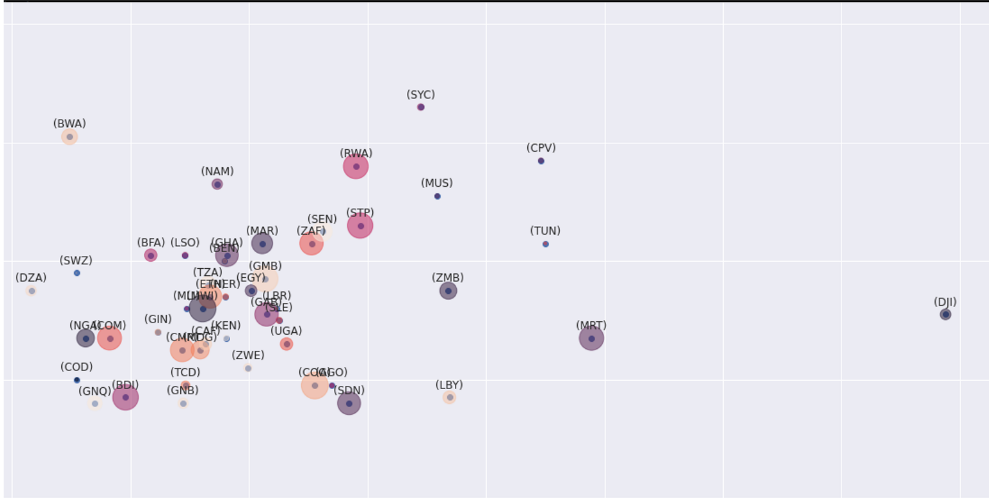
- Niger
- Guinea-Bissau
- Lesotho
- Burundi
- Equatorial Guinea
- Eswatini



Demo



Africa Military Conflict Visualizer
The effect of economic factors on the start of new conflicts in the Year 2000
The effect of economic factors on the start of new conflicts From 1997 to 2018
The effect of ethnicity and religion on the start of new conflicts in the Year 2000
The effect of ethnicity and religion on the start of new conflicts From 1997 to 2018
The 3D Visualization of the effect of ethnicity, religion, and economics on the start of new conflicts From 1997 to 2018



(SYC)
(BWA)
(RWA)
(CPV)
(NAM)
(MUS)
(SEN)
(STP)
(MAR)
(ZAF)
(TUN)
(BFA)
(LSO)
(GHA)
(SWZ)
(DZA)
(TZA)
(EGY)
(GMB)
(ZMB)
(DJI)
(LBR)
(SLE)
(MRT)
(GIN)
(KEN)
(UGA)
(ZWE)
(COD)
(TCD)
(LBY)
(GNQ)
(BDI)
(GNB)
(SDN)

Future Steps Suggestions

The project has an enormous potential. The goal of the project includes a lot of creativity - since there is no direct way to predict military conflicts. With booming amount of available data and models to analyze it, the war prediction maybe be tackled from multiple approaches.

<u>Some potential avenues of developing the project:</u>

- Accounting for even more factors and predicting the likelihood of armed conflict within next 5 years. Amount of available data is underestimated.
- Separating civil (endogene) and international (exogene) conflicts. Hence, possibly improving prediction.
- Originally we saw this project as a conflict predictor model with user interface. We wanted a front-end which employing an LSTM or other RNN model allows the user to pick a region and see the likelihood of a conflict breaking out. Time was not on our side with this goal.
- We also still have a lot of unused data that would have to be engineered into a usable format. With including data about military exercises & arms transfers the model should become more robust.

Thank you for your interest in our work!

We interested in conflicts in order to keep piece.

The image on the front slide is Pablo Picasso's "Dove of peace"

In 1949, after the end of World War II, Picasso was invited to create an image representing peace. Later this "Dove of Peace" would be chosen to represent the first International Peace Conference in Paris in 1949.