PREDICTING MAJOR INTERNATIONAL CONFLICTS USING DATA SCIENCE TECHNIQUES

SUMMARY

The problem:

- **Can we predict if a devastating war is likely to occur within the next five years?**

The client:

- **Organizations that operate internationally and depend on likelihood of wars.** E.g. governments, non-profits, and multinational companies.

The solution:

- Based on the model I built, **yes, we can—with accuracy of 92%, and with 85% of devastating wars predicted five years in advance.**

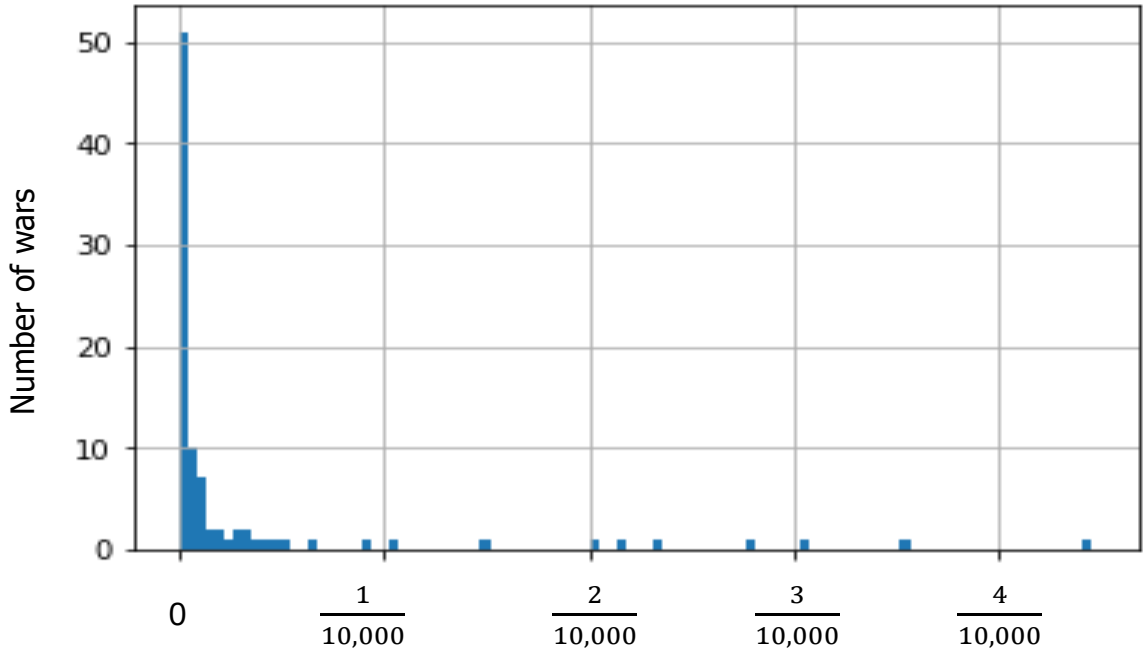
DATA SOURCES

For this iteration of the project, I used these datasets:

- **For historical world population data: Worldometers.**
- **Data on international wars and diplomatic relations: The Correlates of War. The data spans years 1823 to 2003.**

EXPLORING THE DATA

Between 1823 and 2007, the vast majority of wars resulted in less than 1 in every 10,000 people on earth killed in battle.



EXPLORING THE DATA

Defining devastating war:

When more than 1 out of every 10,000 living people die in that war.

There were only 12 such wars that happened within the time period:

1. First Russo-Turkish (1828 - 1829)
2. Mexican-American (1846 - 1847)
3. Crimean (1853 - 1856)
4. Second Schleswig-Holstein (1864)
5. Franco-Prussian (1870 - 1871)
6. Second Russo-Turkish (1877 - 1878)
7. World War I (1914 - 1918)
8. Third Sino-Japanese (1937 - 1941)
9. World War II (1939 - 1945)
10. Korean (1950 - 1953)
11. Vietnam (1965 - 1975)
12. Iran & Iraq (1980 - 1988)

PREDICTING LIKELIHOOD OF A DEVASTATING WAR

Hypothesis: data on diplomatic connections is sufficient to predict likelihood of a major war happening within the next 5 years, with results better than random chance.

PREDICTING LIKELIHOOD OF A DEVASTATING WAR

Defining the world's diplomatic connectivity for each year in the dataset:

Part 1: Calculate the world's overall connectivity ratio.

Divide total *actual* number of connections by total *possible* number of connections:

$$\frac{number\ of\ diplomatic\ connections}{number\ of\ countries * (number\ of\ countries\ - 1)}$$

PREDICTING LIKELIHOOD OF A DEVASTATING WAR

Defining the world's diplomatic connectivity continued:

Part 2: Calculate the most war-prone countries' connectivity ratio.

Divide total *actual* number of connections by total *possible* number of connections:

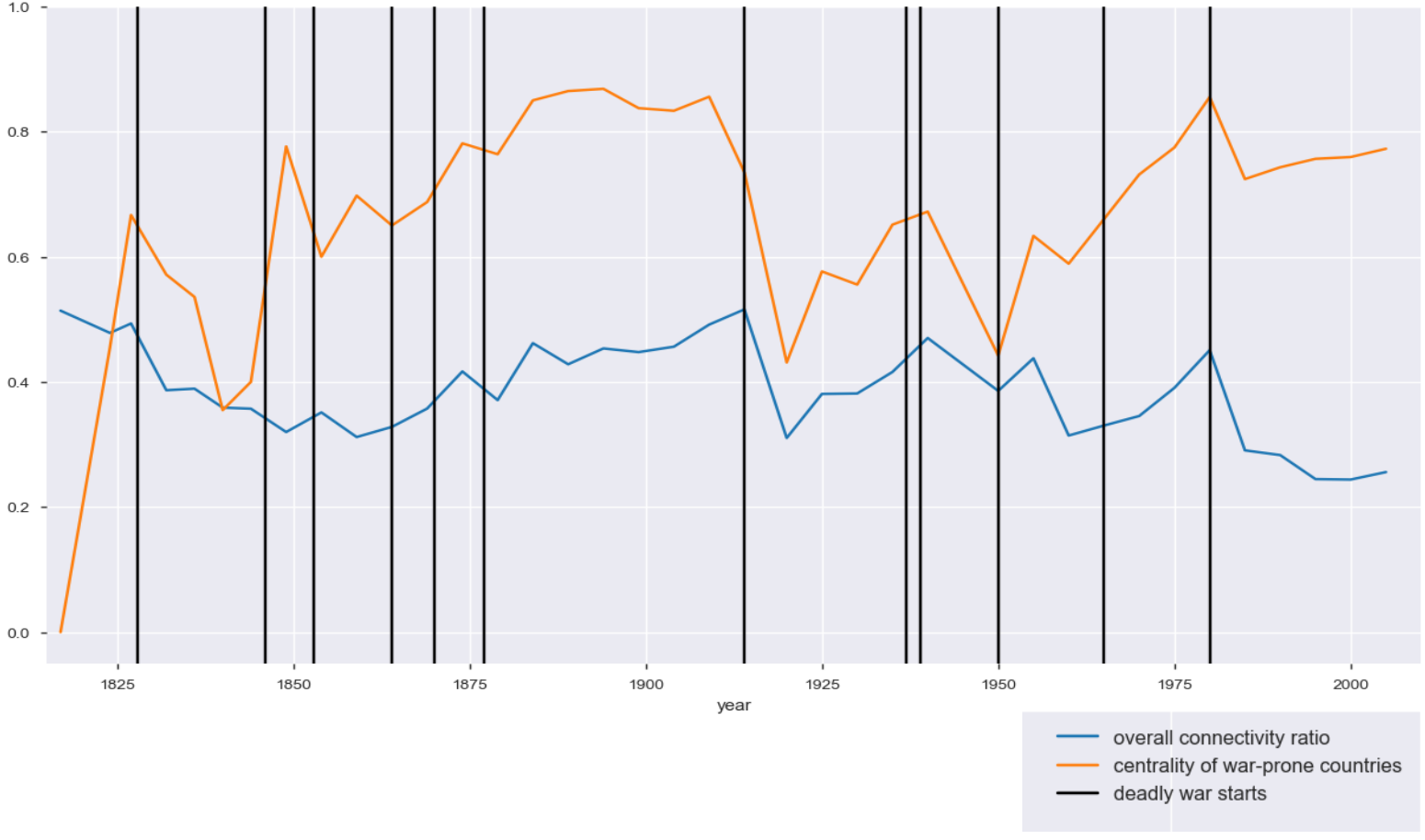
$$\frac{number\ of\ diplomatic\ connections\ warprone\ countries\ have}{number\ of\ warprone\ countries * (number\ of\ countries\ - 1)}$$

A NOTE ON WARPRONE COUNTRIES FOR EACH YEAR

This was a simple calculation of how many wars *up to that year* each country participated in.

And taking top 10% of the most warprone countries up to that year.

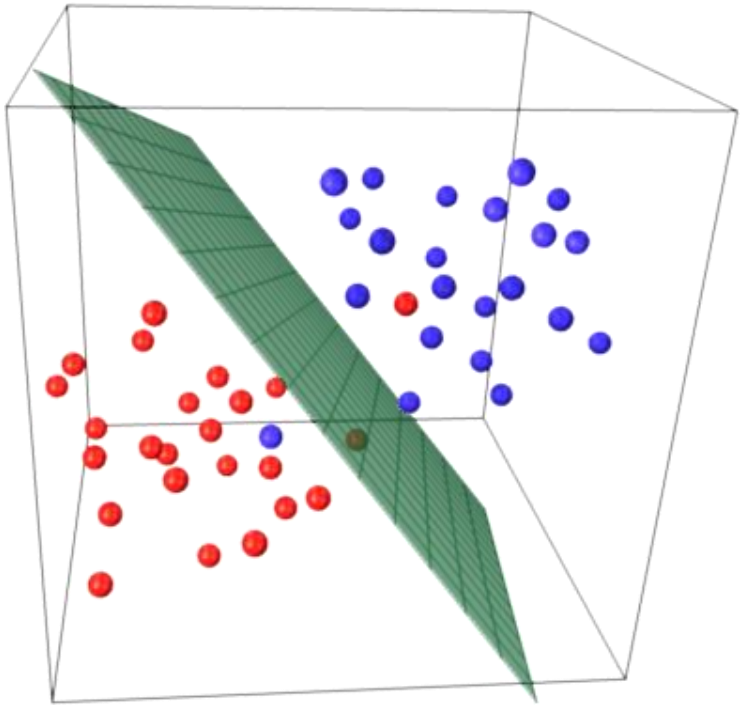
MAPPING DEADLY WARS TO DIPLOMATIC CONNECTIVITY OF THE WORLD



PREDICTING LIKELIHOOD OF A DEVASTATING WAR

Using an Support Vector Classifier (SVC) model to predict the likelihood of war.

WHAT DOES THE SVC MODEL DO?



Simplified view

Support Vector Classifier attempts to build a ‘hyperplane’ in the space of our data, to separate the dots that represent years with high war likelihood (red dots) from the rest of the years.

When we put a new dot in the space, we will know if it is a red or a blue one, based on which side of the hyperplane the dot falls into.

ASSUMPTIONS FOR OUR SVC MODEL

- Each year’s data on diplomatic relations is independent of other year’s data (based on the assumption that each country reconsiders its diplomatic ties at least once a year).
- In the years for which we don't have diplomatic exchange data, diplomatic exchange built up / broke down linearly.
- Data we are using for classification is not going to span more than ~100K years (in our case data spun less than 100 years).
- To satisfy SVC’s assumption that data is identically distributed, I scaled the data to have zero mean and unit variance.

AFTER BUILDING OUT, TESTING AND TUNING THE MODEL, HERE ARE THE FINAL RESULTS

Testing the model on holdout dataset:

- ✓ Accuracy was 92% across all holdout years
- ✓ The model correctly red-flagged 85% of all years 5 years or less away from a major war

Recommendations:

Our best model's prediction is better than random chance and can be used in tandem with other international affairs analysis to predict major wars. See more details on how to run the model in the code here.