WHAT MAKES A SUCCESFUL REVOLUTION

A Project by:

1. Ashish Jha (asj49)
2. Cai-Cian Song (cas386)
3. Zhaokun Wang (zhw83)
4. Raphael Fernandes(rcf35)
5. Shanim Manzoor(shm150)

POLITICAL REVOLUTIONS

The world is always in a flux, constantly changing in cycles of growth and prosperity intermixed with strife and conflict. At any point of time, there will be a place or a group of people wanting to change their situation, be it for personal or more nobler causes. But the question that can come up is what really decides if a revolution is successful: is it the charisma of the leader, popular support, general unrest or poor policies from the ruling class.

DECISION PROBLEM:

- OBJECTIVES**:**

  • Successfully topple the ruling party /president

  DECISION Definition

  • Successfully navigate the myriad decisions and stances to take to ensure that the decision maker's objective of a successful revolution is met

  • Figure out which factors a revolutionary would need to focus on and what kind of revolution, vis a vis, the nature of the revolt itself (peaceful or not)

  DECISION OPTIONS:

  • Appeal to a group of people feeling victimized or appeal to a general majority with populism being the way to go (ethnic fractionalization)

  • Should the revolution focus on absolute change with time (radical policy)and professing an ideology or instead go the quick violent route (use of force)

  • How would the revolution get some form of military support?

  • Would external support be important?

  Conflicting Objectives:

  • Revolutionary leader might want to be the future leader to ensure his vision is carried out

  Key Sources of Uncertainty:

  • Education Level
  • GDP
  • Oil Production
  • General mood
  • Press Freedom
  • Corruption
  • Regime type
  • Youth population percentage
  • Ideology support
  • Ethnic or caste issues
  • Aggression of ruling party
  • Broken infrastructure
  • Access to resources /weapons
  • Continuous Communication

We set out to answer these questions in a broad way while looking at recent revolutions starting with a special focus on the Arab Spring as well as some of the more well-known successes and failures. Drawing from the well of data, we aim to bring into light some interesting and possibly surprising inferences into what decisions really help in playing out a successful revolution. For simplicity, we will define a successful revolution as one where the ruling class/party is toppled for at least a month.

With all this said, our focus has been on the interesting turn of events that led to the Arab Spring. This was quite the hot topic a few years back among economists and geopolitics enthusiasts as they never predicted things could happen so drastically.

ARAB SPRING:

The Arab Spring is commonly understood to have started in the spring of 2011. There's an older incident in Tunisia, where on the 17th of December 2010, a street vendor set himself on fire and became a figurehead and talking point among these repressed group of citizens in the Middle East. It served as the spark and led to the popular uprisings in Tunisia, Egypt and Libya all of which had their share of problems.
In terms of studies, a lot of research has been conducted on the results of these revolts, but few cover the reasons why some countries actually saw a change in government while others failed. One of the key takeaways is how the living conditions of the citizens as well as their social wellbeing or lack therein played a huge part.
It is here, where we need to build a hypothesis and start considering what factors are to be considered. With that, we build a preliminary dataset considering 29 countries with a few outliers put in to test how effective our model would be.
The countries considered were:

Sl. No.	**Country**
1	Algeria
2	Australia
3	Bahrain
4	Chad
5	Djibouti
6	Egypt
7	Eritrea
8	Ethiopia
9	Iran
10	Iraq
11	Israel
12	Jordan
13	Kuwait

14	Lebanon
15	Libya
16	Maldives
17	Morocco
18	Oman
19	Palestine
20	Qatar
21	Saudi Arabia
22	Sudan
23	Syria
24	Tanzania
25	Tunisia
26	Uganda
27	United Arab Emirates
28	Yemen
29	Somalia

- We considered a few examples such as Palestine, Yemen as countries with a revolution, as they are consistently in turmoil
- Australia and Maldives were added as outliers purposely to fudge the data
- The entire dataset was built using some references including:
    - Blitz,Daan (Thesis)[The Arab Spring]
    - World Bank
    - Press Freedom Index
- There are 14 variables and 1 target value called RevolutionDummy which we have defined using Dan's individual datasets for each variable.
- Country name isn't a feature but a way for us to understand the data

FEATURES:

1. Population: This variable is the total population of the country being listed.
2. Arab: From a scale of 0-1 it gives the fraction of Arabs in the populace
3. Muslim: Since we are focused on the Arab Spring, considering Muslims as a separate variable made sense. Fractional (0-1) scale has been used.
4. MedianAge: Interestingly, the younger the median age implies youth dominate the country statistics and can be mobilized a lot easier if changes are to be made.

5. Freedom: Here from a scale of 1-7 , with 7 meaning most free , the relative freedom of the citizens is considered
6. Urbanization: This variable considers how much of the populace resides in the cities
7. Population Density: This variable considers how densely the population is distributed.
8. GDPcap2010: This variable considers the GDP per capita of the country instead of just the total GDP
9. Literacy: The percentage of literate populace is compared to the total of each country
10. Ethnic Fractionalization: Studies have found that this variable is negatively correlated with development. Hence, we have considered it as well.
11. Infant Mortality Rate: Standard statistic showing how good the heath infrastructure present in the country is on a comparative scale.
12. OilExport: This considers the oil export by countries in terms of barrel count. This is a major feature of most of the countries in the middle East and hence considered.
13. PressFreedomScore: This variable considers censorship and how much freedom the press has in a country.
14. Unemployment: This is a standard statistic and can always be correlated with unrest.
15. RevolutionDummy: This variable codes as follows:
    a. 1 if Revolution
    b. 0 if not

CORRELATION MATRIX

First, we used R packages to clarify how correlated features are to each other and to our target value.

The code is pretty simple and the results are attached as well.

Code:

```
getwd()
train1=read.csv("revolution.csv",header = T)
head(train1)
install.packages("corrplot")
library(corrplot)
corrplot(train1,type = "upper", order = "hclust",
        tl.col = "black", tl.srt = 45)
```

Sl. No.	Country	Populatio	Arab	Muslim	MedianAge	Freedom	Urbanizat	Populatio	GDPcap20	Literacy	EthnicFrac	Infantmortality	OilExport	PressFree	Unemploy	RevolutionDummy
1	Algeria	38087012	0.84	0.99	26	5.5	0.73	16	7112	0.699	0.3394	22.57	0	47.33	10.2	0
2	Australia	22003000	0.00008	0.026	36.9	7	0.8518	2.9	51936.89	0.99	0.149	4.1	1607100	5.38	5.1	0
3	Bahrain	1281332	0.514	0.812	30	5.5	0.887	1925	27129	0.946	0.5021	9.93	0	51.38	15	1
4	Chad	11089000	0.55	0.518	17.8	6.5	0.218	9	1841	0.345	0.862	91.93	1106100	33.17	22.6	0
5	Djibouti	888716	0.05	0.94	22	5.5	0.7718	40	2527	0.679	0.7962	51.77	152.6	30.5	59	0
6	Egypt	85294388	0.996	0.9	24.4	5.5	0.435	82	6344	0.72	0.1836	23.3	1395000	43.33	12.5	1
7	Eritrea	4391000	0.5	0.48	18.9	7	0.213	47	683	0.678	0.6524	39.38	0	105	12.1	0
8	Ethiopia	82095000	0.339	0.339	17.7	6	0.17	79	1019	0.427	0.7235	58.28	1378000	49.38	20.4	0
9	Iran	79853900	0.02	0.98	27	6	0.691	48	12789	0.77	0.6684	40.02	0	94.56	15.5	0
10	Iraq	31858481	0.775	0.97	19.1	6	0.664	77	3895	0.782	0.3689	38.86	6880000	45.58	16	0
11	Israel	7707042	0.205	0.169	30.1	1.5	0.91	343	34000	0.926	0.526	3.7	0	31	8.48	0
12	Jordan	6482081	0.98	0.92	22.5	5.5	0.82	71	5907	0.933	0.5926	15.26	705.1	37	12.3	1
13	Kuwait	2695316	0.8	0.85	28.4	4.5	0.983	162	41701	0.933	0.6604	7.68	152.4	23.75	2.2	0
14	Lebanon	4131583	0.95	0.597	28.5	4	0.872	414	15523	0.874	0.1314	14.81	2600000	20.5	6.2	0
15	Libya	6002347	0.97	0.97	25.6	7	0.772	4	6017	0.894	0.792	12.26	2142000	63.5	30	1
16	Maldives	315885	0	0.984	25	4.5	0.412	1110	8,092	0.984	0.2	25.5	2377000	16	11.7	0
17	Morocco	32649130	0.66	0.99	26.2	4.5	0.57	73	5080	0.561	0.4841	25.49	7798	47.4	8.8	0
18	Oman	3154134	0.9	1	25.1	5	0.734	10	27567	0.814	0.4373	14.46	0	40.25	15	0
19	Palestine	4068779	0.9	0.92	18.2	7	0.74	778.2	2338.72	95%	0.8	19.9	0	134	23.4	1
20	Qatar	2042444	0.4	0.775	31.6	5.5	0.98	171	98948	0.963	0.7456	6.6	85	38	0.5	0
21	Saudi Ara	26939583	0.9	1	26.1	6.5	0.823	13	24411	0.863	0.18	15.08	1097000	61.5	10.7	0
22	Sudan	34847910	0.7	0.97	18.7	7	0.332	18	273	0.611	0.7147	54.23	0	85.33	20	1
23	Syria	22457336	0.903	0.9	21.9	6.5	0.561	122	5041	0.796	0.5399	14.63	0	91.5	18	1
24	Tanzania	46098591	0.0015	0.45	17.9	3	0.267	50	1,511	0.678	0.7353	45.1	175.2	13	10.7	0
25	Tunisia	10835873	0.98	0.98	29	6	0.663	66	9389	0.748	0.0394	24.07	97.27	72.5	18.8	1
26	Uganda	33042000	0.01	0.151	15	4.5	0.156	148	1,332	0.732	0.9302	62.47	7337	25.5	4.2	1
27	United Arab Emirates	5473972	0.4	0.96	28	5.5	0.846	97	47729	0.779	0.6252	11.25	0	23.75	2.4	0
28	Yemen	25408288	0.95	0.99	18.2	5.5	0.323	48	2307	0.639	0	57.5	0	82.13	35	1
29	Somalia	12050000	0.05	0.99	18.2	7	0.377	16	600	0.378	0.8117	51.93	125.7	66	54.26	1

Figure 1: Snapshot of Dataset (ArabSpring)

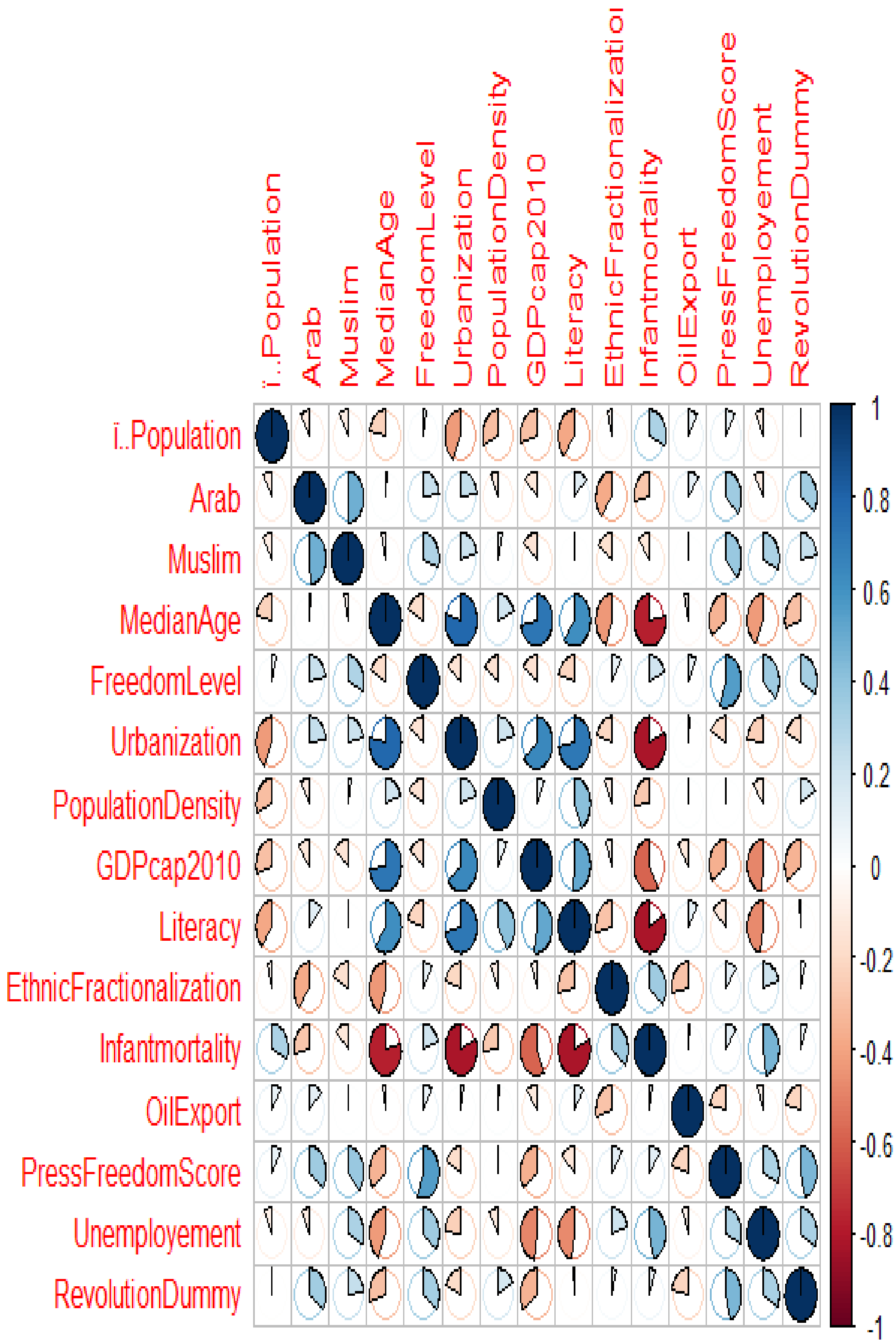


Figure 2: Correlation Matrix

- From the correlation matrix we can see some interesting correlations as Arab, Unemployment and PressFreedom seem important as they have positive correlations
- GDPperCapita2010 is negatively correlated which implies that lower the value ,higher the chances of a political revolution

Let’s now load this into Genie and see what Bayesian models can be formed .

GENIE

TAN Model (Tree Augmented Naïve Bayes)

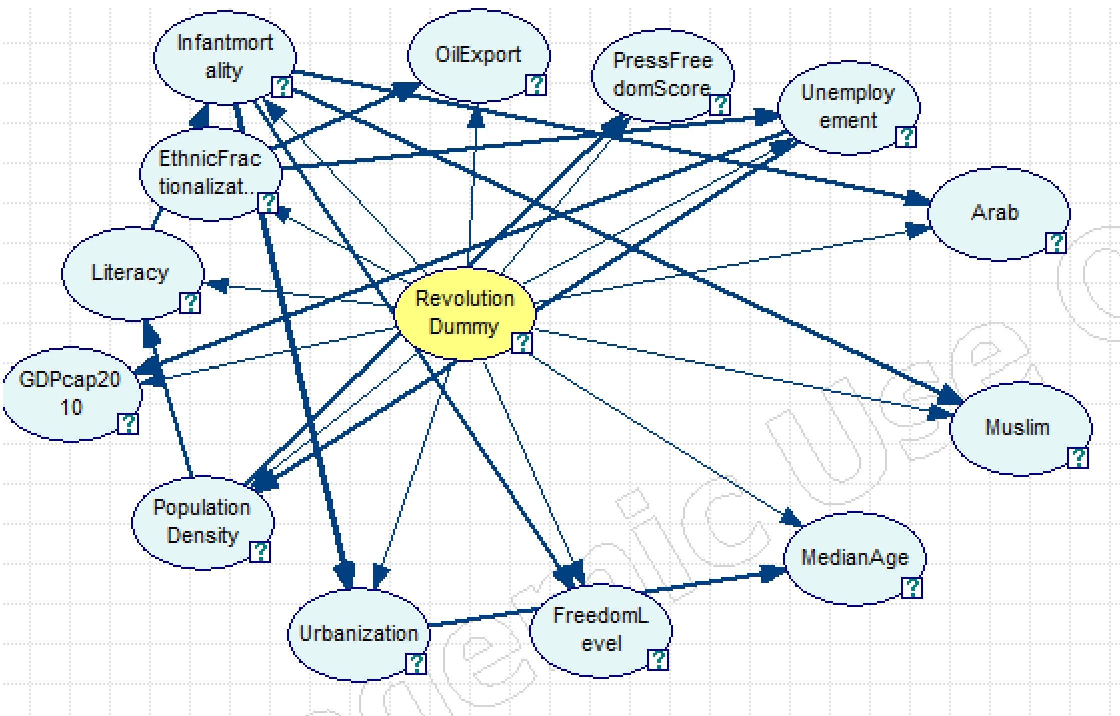
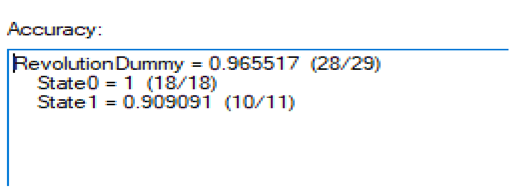


Figure 3: TAN Model

- Influence levels are also shown here



Accuracy:

RevolutionDummy = 0.965517 (28/29)
State0 = 1 (18/18)
State1 = 0.909091 (10/11)

- The accuracy came upto 0.96 but the dataset is extremely limited in scope though educational

Bayesian Search

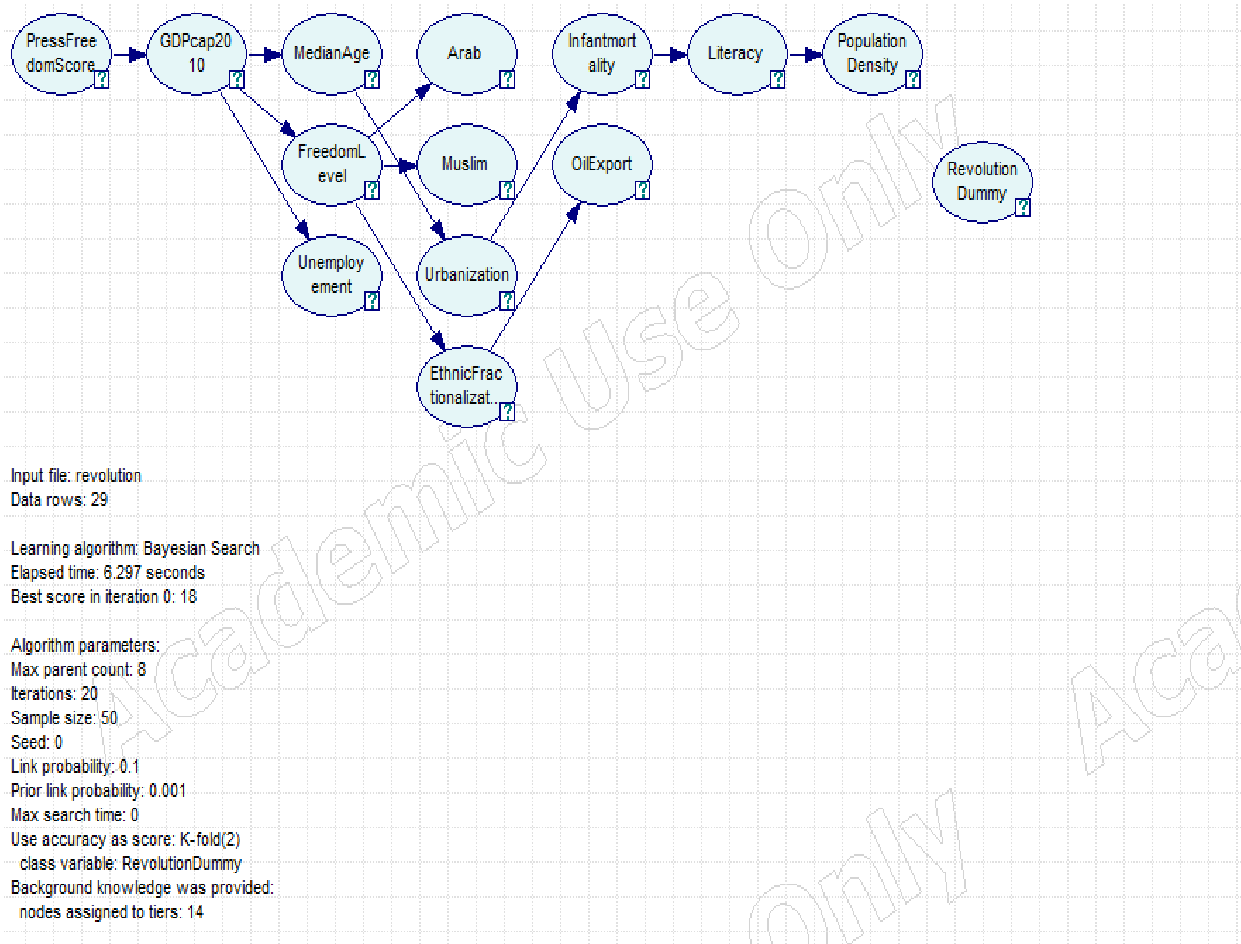


Figure 4: Bayesian Search

We built a Bayesian Search model as seen above as well though this didn't connect to our target value. Background knowledge wasn't specified.

Colgan Dataset

JD Colgan has an excellent dataset called Revolutionary leaders and revolutions. In his paper Measuring revolutions, he talks about peace management and how revolutions are started. The research is extremely meticulous, and we have used the dataset for our models.

ID	Institution or practice	Example	Variable Name
1	Executive power and selection	• Major change to a formal constitution • *De facto* change to leader selection (e.g., abolishment of monarchy)	*chg_executivepower*
2	Political ideology	• Adoption of communism or fascism as official ideology of the state or its single-party leadership	*chg_politicalideology*
3	Official state name	• Change from USSR to Russian Federation	*chg_nameofcountry*
4	Property ownership	• Major changes in property ownership, such as land reform or nationalization of key industries • Changes in economy type (market vs. collectivized ownership)	*chg_propertyowernship*
5	Gender and Ethnic Status	• Implementation of major restrictions on women's dress, employment, inheritance and/or property ownership • Changing the institutionalized status or political rights of major ethnic groups • Granting women the right to vote	*chg_womenandethnicstatus*
6	State-religion relationship	• Constitutional adoption of a single religion as the official state religion, to the detriment of other religions • Adoption of a religion in the official state name (e.g., "Islamic Republic")	*chg_religioningovernment*
7	Leadership of revolutionary council while in power	• Leader creates and chairs a National Revolutionary Council	*chg_revolutionarycommittee*

Figure 5: Variable explanation [Revolutionary Leaders Codebook]

- The dataset was in .dta format , which is used in Stata
- We converted it into CSV using **Stata SE15**

- Data cleaning was done removing values such as year and dates as they aren't useful in the model as of now
- Using Genie, we then replaced missing values with median values in each column.
- There were 268 such rows in total

Bayesian Search:

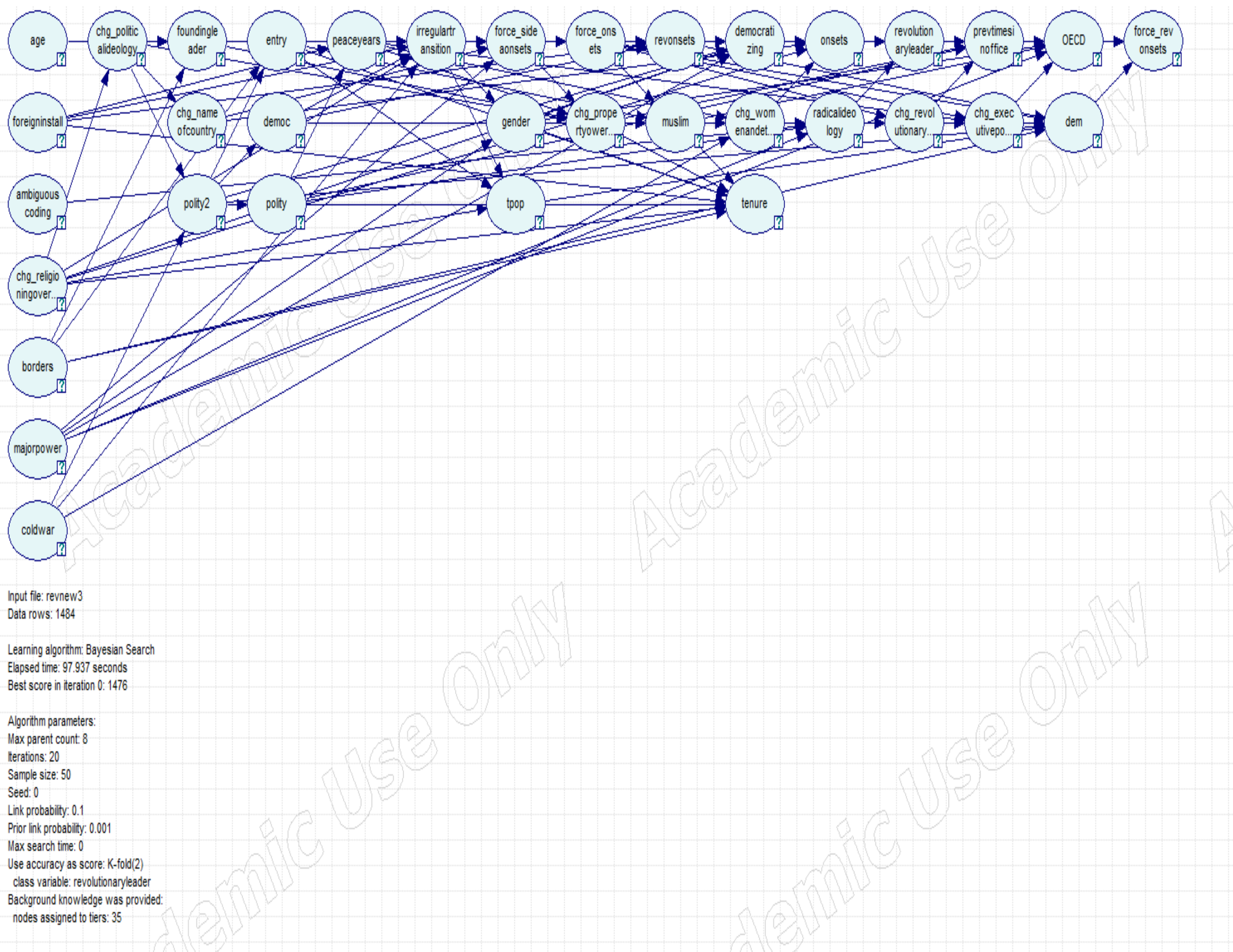


Figure 6: Bayesian Search Model

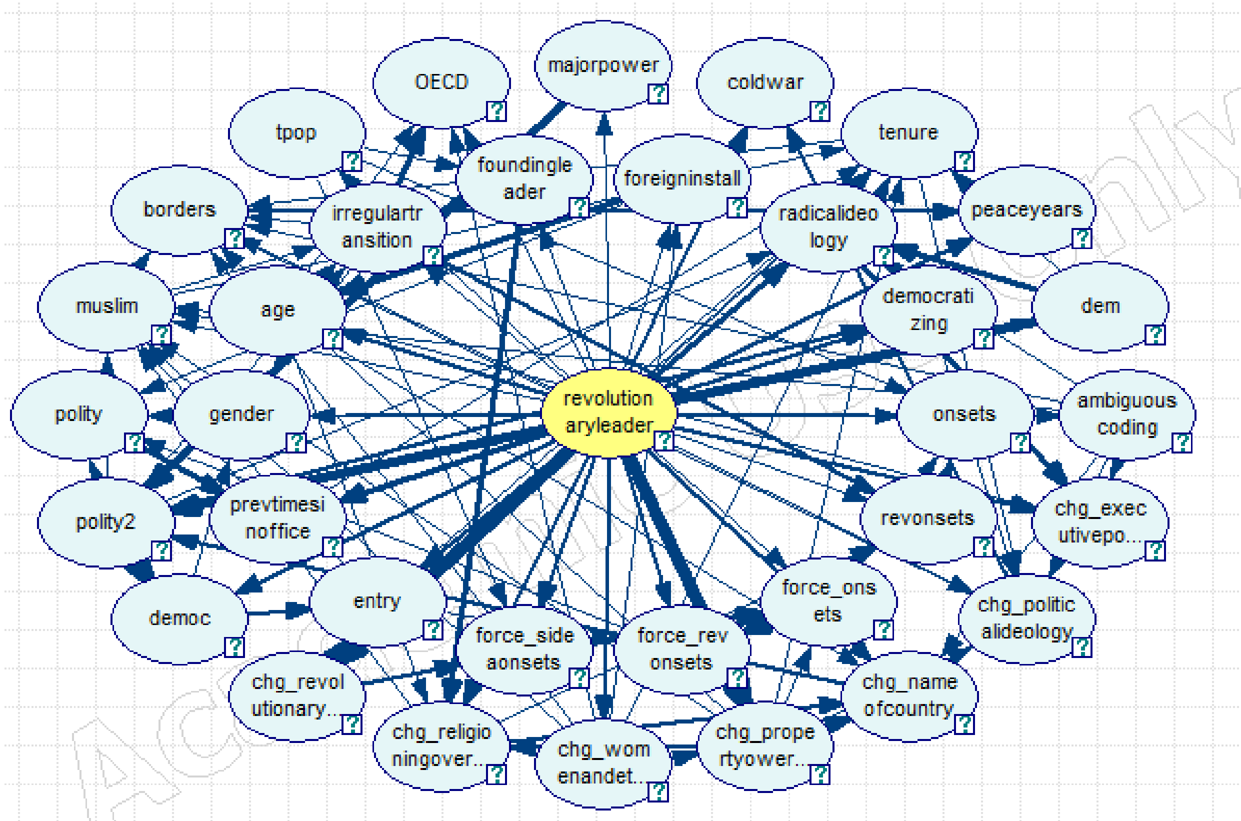


Figure 7: Naïve Bayes Model

- Here we considered Revolutionary leader as our target variable and built the models accordingly.
- To further delve into some basic information about the dataset, the original aim was to classify leaders depending on their ideology and impact on international politics.
- The timeframe considered is 1945-2004
- The dataset considers variables such as change in name of country,women's rights ,property ownership,ideology as categorical variables
- Radical policies were also considered
- A few of the revolutionary cases considered include:
    - Venezuela 1998-2004
    - Myanmar 1988
    - Romania(1989-1991)
    - Eastern European countries in early 1990s (Cold war ,USSR dissolution)
    - Greece 1967
    - Ghana 1981
    - Chile 1978
    - Pakistan 1977 etc.

variable name	storage type	display format	value label	variable label
stabb	str3	%9s		COW state abbreviation
ccode	float	%8.0g		COW state code
year	float	%8.0g		
onsets	float	%9.0g		MID onsets
revonsets	float	%9.0g		MID onsets where state is coded as revisionist by COW
sideaonsets	float	%9.0g		MID onsets where state is coded as State A by COW
force_onsets	float	%9.0g		MID onsets where force was used
force_revonsets	float	%9.0g		MID onsets where force was used and state is coded as revisionist
force_sideaon~s	float	%9.0g		MID onsets where force was used and state is coded as State A
obsid	str10	%10s		
ccname	str28	%28s		
leader	str31	%31s		
startdate	str10	%10s		
enddate	str10	%10s		
entry	byte	%8.0g		
prevtimesinof~e	byte	%8.0g		
posttenurefate	float	%8.0g		
gender	byte	%8.0g		
age0	byte	%8.0g		Leader's age when entering office
startobs	str9	%9s		
endobs	str9	%9s		
age	byte	%8.0g		Leader's age in current year
numld	byte	%8.0g		Number of leaders coded by Archigos in this calendar year
usedforce	byte	%8.0g		Use of Force
irregulartran~n	int	%8.0g		IrregularTransition
foundingleader	byte	%8.0g		FoundingLeader
foreigninstall	byte	%8.0g		ForeignInstall
radicalideology	byte	%8.0g		RadicalIdeology
democratizing	byte	%8.0g		Democratizing
revolutionary~r	int	%8.0g		RevolutionaryLeader
ambiguouscoding	byte	%8.0g		Ambiguous coding
chg_executive~r	int	%8.0g		Executive power
chg_political~y	float	%8.0g		Political ideology change
chg_nameofcou~y	int	%8.0g		Name of country change
chg_propertyo~p	float	%8.0g		Property owernship change
chg_womenande~s	float	%8.0g		Women and Ethnic status change
chg_religioni~t	float	%8.0g		Religion-in-government change
chg_revolutio~e	float	%8.0g		Revolutionary committee
totalcategori~d	float	%8.0g		Total domestic institutions changed
revage	byte	%8.0g		Years since the revolution
radrestrict	byte	%8.0g		'Unambiguous' revolutionary/radical leaders
start_year	int	%8.0g		Year in which state comes into being as a polity
code	str4	%9s		World Bank state code
democ	byte	%8.0g		DEMOC
autoc	byte	%8.0g		AUTOC
polity2	byte	%8.0g		POLITY2
durable	int	%8.0g		DURABLE
polity	float	%9.0g		
muslim	float	%9.0g		Muslim%
flgdpen	float	%9.0g		From Fearon and Laitin 2003, GDP per capita
flethfrac	float	%9.0g		From Fearon and Laitin 2003, ethnic fractionalization
region	str4	%9s		World Bank coding of geographic region
leg_british	byte	%8.0g		
leg_french	byte	%8.0g		
leg_socialist	byte	%8.0g		
leg_german	byte	%8.0g		
leg_scandivan~n	byte	%8.0g		
borders	byte	%8.0g		
tpop	long	%12.0g		Total population, COW dataset
cinc	float	%9.0g		National Military Capabilities, COW dataset
nonrevonsets	float	%9.0g		MID onsets where state is not coded as revisionist by COW
OECD	float	%9.0g		
majorpower	float	%9.0g		
state_num	float	%9.0g		group(ccode)
regnum	float	%9.0g		group(region)
coldwar	float	%9.0g		
uniqtag	float	%9.0g		unique rank of (age0) by code leader age0
tenure	float	%9.0g		
peaceyears	float	%9.0g		
entcat	float	%9.0g		RECODE of polity
entrc	float	%9.0g		Regime change coding based on Enterline 1998
entrcage	float	%9.0g		
entrc15	float	%9.0g		
revstart	float	%9.0g		Dichotomous indicator for year of revolution
durrc	float	%9.0g		Regime change coding based on Polity's durable variable
durrcage	float	%9.0g		
durrc15	float	%9.0g		
dem	float	%9.0g		Democracy indicator
seveninstitut~d	float	%9.0g		Coding of seven domestic insitutional change areas exists

Figure 8 : Description of Variables(Stata)

For our next step, we needed a few more variables in the economic sense similar to the Arab Spring dataset we just built, so we set out to use other sources to incorporate more features to our dataset

GAPMINDER

Gapminder is an organization started by Hans Rosling and his colleagues to collect statistics to better understand the different socioeconomic criteria and developments around the world. The datasets are verified and accurate but may have some countries missing compared to what we would need. As they are a nonprofit with the goal of sustainable inclusive development, they possess reliable data.

Method:

1. For the first step, we collected all the different datasets for different metrics such as " unemployment_rate_ 15plus" , "Child_Mortality_rate_per1000births" using an R script to get all of them into a single csv file
2. From this CSV file , our original dataset considers the years 1812-2011 as well as upto 2060 or so in some cases where projections were made
3. The entire dataset is around 141 mb
4. From here we need only certain features which we selected as below :

FEATURE NAMES
Aged.15.24.Employment.Rate.Percent
Aged.15.24.Unemployment.Rate.Percent
Aged.15.64.Labour.Force.Participation.Rate.Percent
Aged.15Plus.Employment.Rate.Percent
Aged.15Plus.Labour.Force.Participation.Rate.Percent
Armed.Forces.Personnel.Percent.Of.Labor.Force
Arms.Imports.Us.Inflation.Adjusted
Child.Mortality.0.5.Year.Olds.Dying.Per.1000.Born
Corruption.Perception.Index.Cpi
Crude.Birth.Rate.Births.Per.1000.Population
Crude.Death.Rate.Deaths.Per.1000.Population
Gdp.Per.Capita.Yearly.Growth
Hdi.Human.Development.Index
Inflation.Annual.Percent
Life.Expectancy.Female
Life.Expectancy.Male
Life.Expectancy.Years
Literacy.Rate.Adult.Total.Percent.Of.People.Ages.15.And.Above
Literacy.Rate.Youth.Female.Percent.Of.Females.Ages.15.24

Male.Long.Term.Unemployment.Rate.Percent
Median.Age.Years
Military.Expenditure.Percent.Of.Gdp
Muslim
Oil.Consumption.Per.Cap (capita)
Population.Density.Per.Square.Km
Population.Growth.Annual.Percent
Urban.Population.Percent.Of.Total

5. Next we used some basic Database concepts where we did the **union** of the Colgan Dataset and the filtered Gapminder dataset
6. The union (GEO/STABB) for the country code like AFG,USA etc with an AND condition using year
7. We removed most of the features related to timestamps as we weren't clear on how to process them

COMBINED DATASET

MISSING VALUES:

- This new dataset is now combined but we have a lot of NA values
- So next, we used R again, to sort column by column and replace NA values for those columns which are numeric using as.numeric
- The replacement was done using median values, to make sure categorial values weren't affected

SELECTED FEATURES (Brief Description)

"Age0 ": age of leader entering office

"Age": Leader's age current (relative)

"Aged.15.24.Employment.Rate.Percent": employment rate of youth between the ages of 15-24

"Aged.15.24.Unemployment.Rate.Percent": ": unemployment rate of youth between the ages of 15-24

"Aged.15.64.Labour.Force.Participation.Rate.Percent": Proportion of people between the ages of 15-64 in the labor force

"Aged.15Plus.Employment.Rate.Percent": Employment rate overall for any individual above the age of 15
"Aged.15Plus.Labour.Force.Participation.Rate.Percent": percentage of individuals above the age of 15 taking part in the labor force.
"Ambiguouscoding": This is where uncertainty arises as to whether the individual was a revolutionary or not
"Armed.Forces.Personnel.Percent.Of.Labor.Force": Proportion of armed forces compared to overall labor force
"Arms.Imports.Us.Inflation.Adjusted": value of arms imported over the year overall adjusted for inflation
"Autoc": Autocratic or not
"Borders": number of countries bordering the country
"Chg.Executivepower": This variable defines if there's a major change in executive power or constitution intended /executed by the leader
"Chg.Nameofcountry": As the name suggest, the value becomes 1 if the name of the country has been changed
"Chg.Politicalideology": If the leader intends to adopt a specific ideology such as communism or fascism
"Chg.Propertyowernship": If the leader intends to make major changes in land laws
"Chg.Religioningovernment": This defines if the leader intends to change the official state religion
"Chg.Revolutionarycommittee": This variable deals with whether the leader creates a radical revolutionary council to leader (for example:Iran)
"Chg.Womenandethnicstatus": This deals with whether the leader plans to change woman's rights or some of the ethnic groups' representation
"Child.Mortality.0.5.Year.Olds.Dying.Per.1000.Born": This variable deals with child mortality below the age of 5
"Cinc": This variable uses the COW dataset to encode strength of individual military power of the state
"Coldwar": This variable deals with if the country was involved in the cold ware
"Crude.Birth.Rate.Births.Per.1000.Population": This variable gives the birth rate per 1000
"Crude.Death.Rate.Deaths.Per.1000.Population": This variable gives the death rate per 1000
"Dem": Indicator of democracy

"Democ": DEMOC
"Democratizing": Decides if new government was significantly more democratizing
"Durable" : measures durability of regime
"Durrc15": related to durability coding
"Durrc":related to durability coding
"Entcat": related to polity
"Entry": How the leader entered office
"Flethfrac": Measures ethnic fractionalization
"Flgdpen": uses Fearon and Latin 2003 to define GDP per capita
"Force.Onsets": Force was used in Military issues
"Force.Revonsets": Force was used and state is considered revisionist
"Force.Sideaonsets": Force used and the state/country is considered aggressor
"Foreigninstall": Leader installed by outside party
"Foundingleader": If the individual is a founding leader
"Gdp.Per.Capita.Yearly.Growth": Measures GDP per capita year-on year growth
"Hdi.Human.Development.Index": HDI is a very popular measure
"Inflation.Annual.Percent": Inflation rate percentage
"Irregulartransition": If transition of power was not normal, this is significant to our project
"Leg.British": Has legislative system similar to British system
"Leg.French": Has legislative system similar to the French
"Leg.German": Has legislative system similar to the Germans
"Leg.Scandivanian": Has legislative system similar to the Scandinavians
"Leg.Socialist": Has legislative system similar to the socialist
"Life.Expectancy.Female": Average life expectancy of females in the country
"Life.Expectancy.Male": Average life expectancy of males in the country
"Life.Expectancy.Years": Average life expectancy overall

"Literacy.Rate.Adult.Total.Percent.Of.People.Ages.15.And.Above": Literacy rate of individuals above the age of 15
"Literacy.Rate.Youth.Female.Percent.Of.Females.Ages.15.24": Literacy rate of female youth between the ages of 15-24
"Majorpower": Clout of the concerned state
"Male.Long.Term.Unemployment.Rate.Percent": Unemployment rate of individuals in the situation for a long period of time
"Median.Age.Years": Median age of individuals
"Military.Expenditure.Percent.Of.Gdp": % of GDP used for military expenditure
"Muslim": Proportion of Muslims in the country
"Nonrevonsets": If the revolution happened with the COW dataset not defining state as revisionist
"Oil.Consumption.Per.Cap": Oil consumption per capita in the country
"Onsets": Overall MID onsets
"Peaceyears": Number of years of peace in the country
"Polity2": ": Form of governance (used mainly if transitioning from monarchy)
"Polity": Form of governance (used mainly if transitioning from monarchy)
"Population.Density.Per.Square.Km": Self-explanatory
"Population.Growth.Annual.Percent": Self-explanatory
"Posttenurefate": Fate of the country at the end of the regime
"Prevtimesinoffice": Number of times individual has been in office
"Radicalideology": defines if radical ideology has been used
"Radrestrict": Unambiguosly revolutionary leaders
"Regnum": region coding
"Revolutionaryleader": If the leader is a true revolutionary or not (difficult to define)
"Revstart": Indicator of start of revolution
"State.Num": country code
"Tenure": tenure in years

"Totalcategorieschanged": Number of changes of domestic institutions
"Tpop": Total population
"Uniqtag": unique rank of leader by age
"Urban.Population.Percent.Of.Total": percentage of population living in urban areas
"Usedforce": "Force used by entity

NEXT STEPS

- Once done, we are ready to create a model using Genie
- The dataset looks like below:

Oecd	Age0	Age	Aged.15.2	Aged.15.2	Aged.15.6	Aged.15Pl	Aged.15Pl	Ambiguou	Armed.Fo	Arms.Imp	Autoc	Borders	Ccname	Ccode	Chg.Execu	Chg.Name	Chg.Politi	Chg.Prope	Chg.Religi	Chg.Revol	Chg.Wom	Child.Mor	Cinc	Code	Coldwar	Corruptio	Crude.Bir
0	50	50	39.1	12.95	66.9	56.2	61.6	0	1.43112	3.62E+08	7	4	Afghanist	700	1	0	0	1	0	0	0	255.5	0.001475	AFG	1	NA	50.67
0	39	44	50.4	12.95	60.1	56.7	58.8	0	1.37887	1.44E+09	8	7	Afghanist	700	1	0	0	1	0	0	0	174.2	0.001333	AFG	0	NA	48.9
0	53	53	50.2	12.95	60	56.5	58.7	0	1.28698	1.26E+08	-77	6	Afghanist	700	1	1	1	0	1	0	0	167.8	0.001397	AFG	0	NA	48.83
0	53	54	50.4	12.95	59.9	56.6	58.7	0	1.19532	1.26E+08	-77	6	Afghanist	700	1	1	1	0	1	0	0	162	0.001459	AFG	0	NA	48.84
0	53	55	50	12.95	59.8	56.2	58.6	0	1.12042	1.26E+08	-77	6	Afghanist	700	1	1	1	0	1	0	0	156.8	0.001525	AFG	0	NA	48.9
0	53	56	50	12.95	59.7	56.2	58.5	0	9.01913	1.26E+08	-77	6	Afghanist	700	1	1	1	0	1	0	0	152.3	0.001232	AFG	0	NA	48.98
0	37	37	50	12.95	59.7	56.1	58.5	0	9.74832	1.26E+08	7	6	Afghanist	700	0	1	0	0	1	0	1	148.6	0.001272	AFG	0	NA	49.04
0	37	38	50.1	12.95	59.6	56.2	58.4	0	9.50544	1.26E+08	7	6	Afghanist	700	0	1	0	0	1	0	1	145.5	0.001286	AFG	0	NA	49.04
0	37	39	50.1	12.95	59.6	56.2	58.4	0	8.67861	1.26E+08	7	6	Afghanist	700	0	1	0	0	1	0	1	142.6	0.001273	AFG	0	NA	48.93
0	37	40	50.1	12.95	59.6	56.1	58.4	0	8.47497	1.26E+08	7	6	Afghanist	700	0	1	0	0	1	0	1	139.9	0.004156	AFG	0	NA	48.7
0	37	41	50.2	12.95	59.6	56.1	58.5	0	8.29228	1.26E+08	7	6	Afghanist	700	0	1	0	0	1	0	1	137	0.004176	AFG	0	NA	48.33
0	44	44	50.8	12.95	59.7	56.5	58.5	0	1.43112	1.26E+08	-66	6	Afghanist	700	1	0	0	1	0	0	0	133.8	0.00133	AFG	0	NA	47.84
0	44	45	50.6	12.95	59.8	56.4	58.6	0	2.2667	3.40E+07	-66	6	Afghanist	700	1	0	0	1	0	0	0	130.3	0.001811	AFG	0	NA	47.23
0	44	46	47.8	12.95	59.9	54.4	58.8	0	2.33452	1.26E+08	-66	6	Afghanist	700	1	0	0	1	0	0	0	126.8	0.001811	AFG	0	NA	46.54

Figure 9: Head of part of dataset combined

GENIE:

- First, since all the missing values were treated , we discretize the entire set
- We used hierarchial discretization with 2 levels for each column and had a total of 90 features including the target variable irregular transition

KEY TAKEAWAYS:

- The network structure is very complicated even after increasing probability requirements and using cross validation.
- Due to so many missing values, we've used median at too many columns. This increases chances of our model converging to median values for most cases.

MODEL:

- After using 2-fold cross validation and increasing link probability values, we got the model using **Bayesian Search**
- The model performs well on the target variable and has 63 parameters in total
- After running, the model's view in the genie interface is a complicated network structure
-

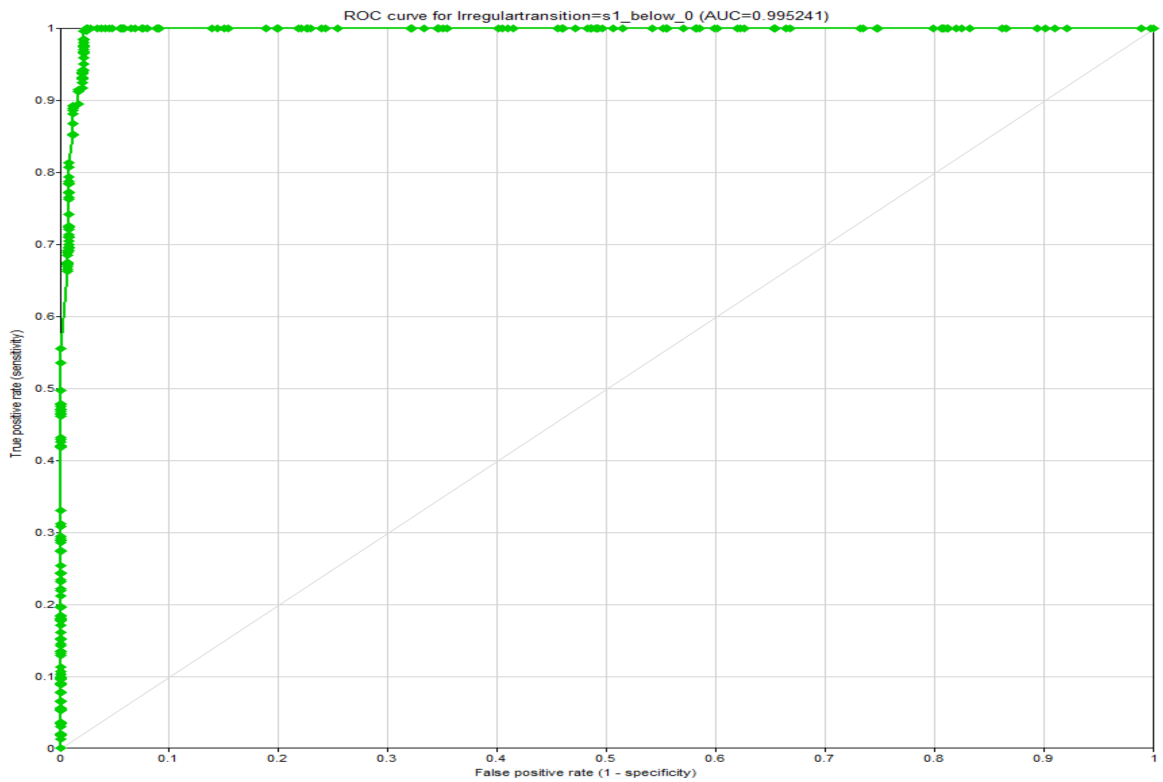


Figure 10: ROC Curve

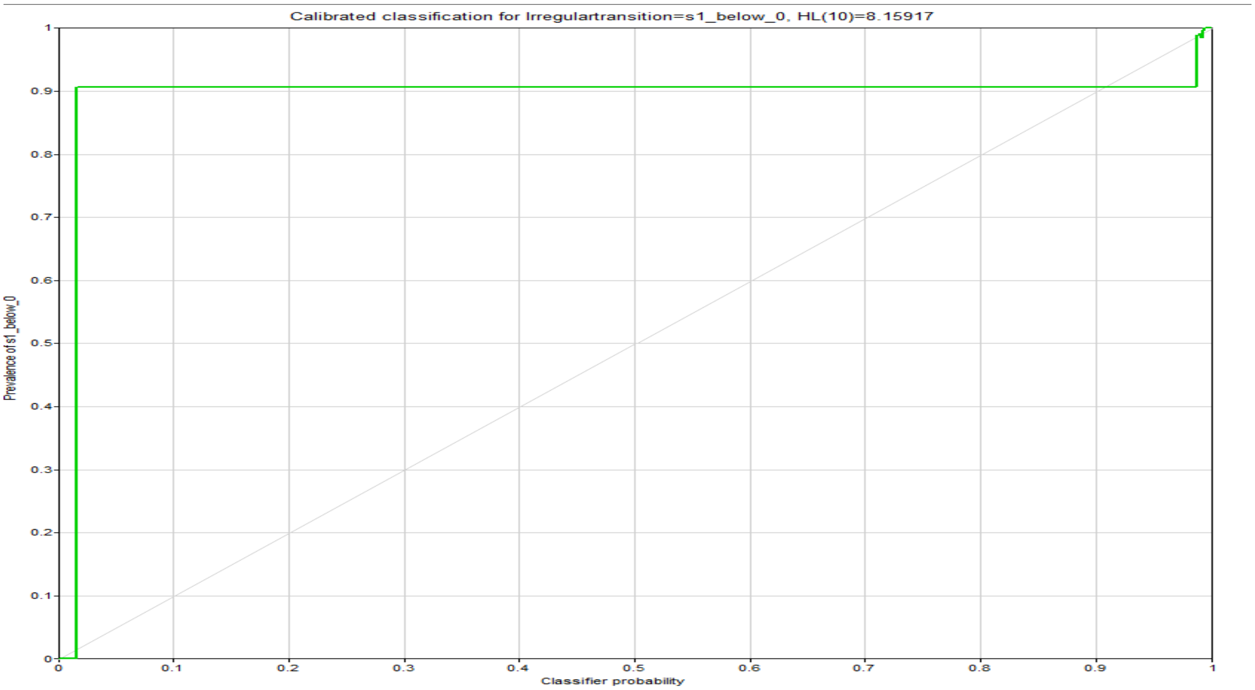


Figure 11: Calibration Curve

Class node: Irregulartransition

	s1_below_0	s2_0_up
s1_below_0	**2416**	0
s2_0_up	27	**1074**

Figure 12 : Confusion Matrix

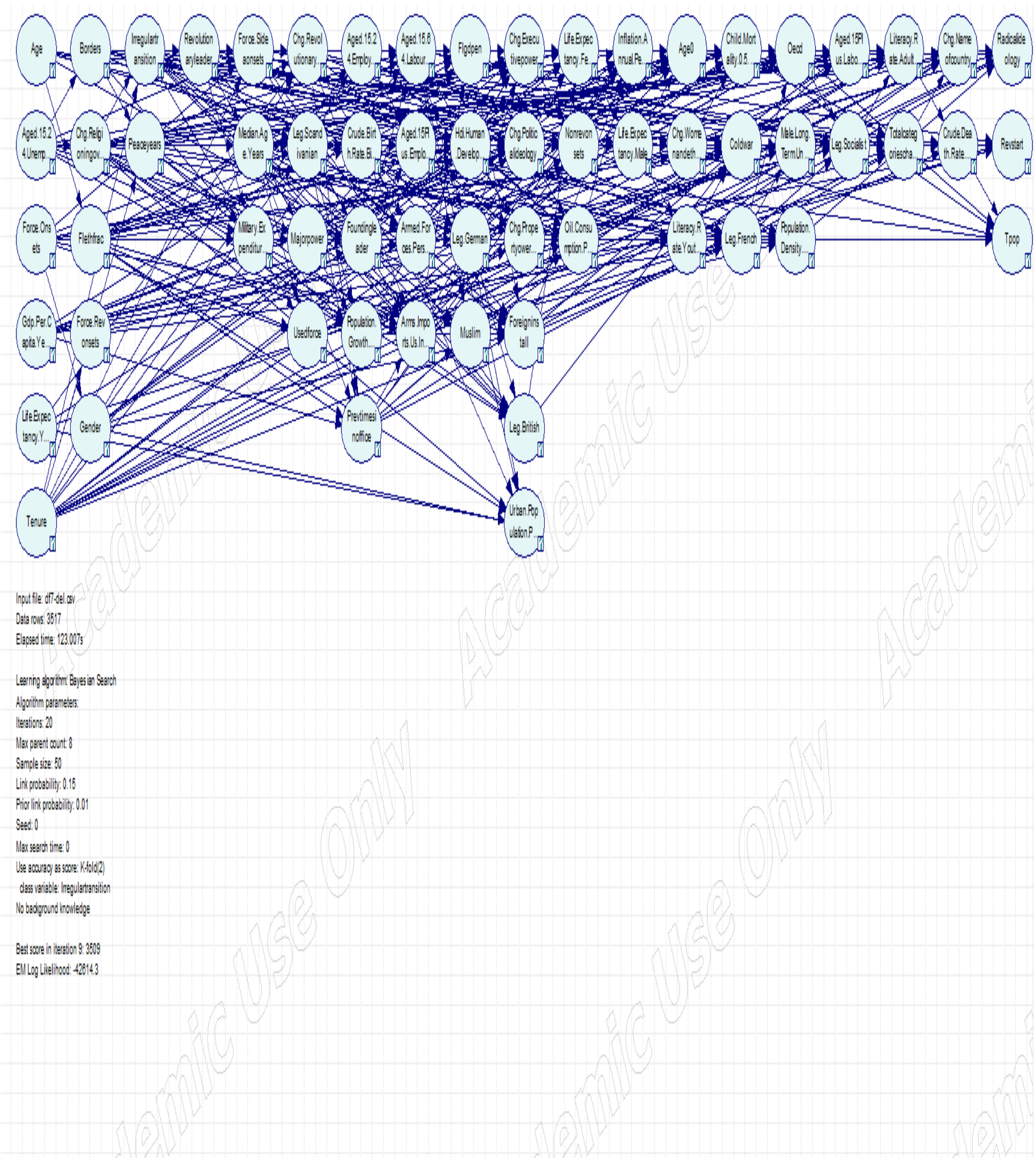


Figure 13: Bayesian Search Network (Low clarity, xdsl file will look better)

User Interface:

- Using Pysmile we have integrated every node with its 2 states and as options

```
if sys.argv[1] > 17:
    net.set_evidence("Force_Onsets","s2_17_up")
else:
    net.set_evidence("Force_Onsets","s1_below_17")

if sys.argv[2] == "Male":
    net.set_evidence("Gender","State0")
else:
    net.set_evidence("Gender","State1")

if sys.argv[3] < 25:
    net.set_evidence("Aged_15_24_Unemployment_Rate_Percent","s1_below_25")
else:
    net.set_evidence("Aged_15_24_Unemployment_Rate_Percent","s2_25_up")
```

Figure 10: Sample code with different states being considered

- The xdsl file is integrated and was successfully run from the terminal when running the python file
- Here let's consider the first 3 lines:
    - The first argument is "Force_Onsets" which has 2 states 0 and 1 which depend on the threshold value 17
    - Depending on this , we can code the values accordingly to the correct band ,in this case being 17 and above or below 17.

FEATURE SELECTION

- We used different correlation coefficients to compare different features to select those which would be most valuable to the form
- A brief snapshot of how it would look:

Feature	Pearson	Chi-2	RFE	Logistics	Random F	distance c	mic	Total
Usedforce_s2_0_up	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	7
Radicalideology_s1_below_(	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	7
Chg.Propertyowernship_s2_	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	7
Chg.Propertyowernship_s1_	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	7
Chg.Executivepower_s2_0_u	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	7
Chg.Executivepower_s1_bel	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	7
Usedforce_s1_below_0	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	6
Revolutionaryleader_s2_0_u	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	6

Figure 14: Feature Correlation Measures

WEB INTERFACE:

- Next we used HTML forms using Bootstrap to select 5-6 major features and hard coded the remaining the most occurring ones for the purpose of simplicity
- This was done because having 63 dropdowns/toggle switches would be too long and time consuming for the user and we couldn't design an effective solution
- We used a python script to get all the most common values for the features into a dictionary
- Using this dictionary, we hardcoded 57-58 variables while the remaining were given for user inputs using dropdowns

REQUIREMENTS (To run)

To install server packages:

1 - go to root folder and in the terminal type: npm install

To install front-end packages:

1 - go to root/src and in the terminal type: npm install

To execute in the system you need to:

1 - Open a terminal and in the root folder type: node server. This will instantiate the server in the localhost:8080

2 - In a different terminal go to root/src and type: ng serve. This will instantiate Angular 2 in the localhost:4200

The page will be instantiated on localhost:4200.

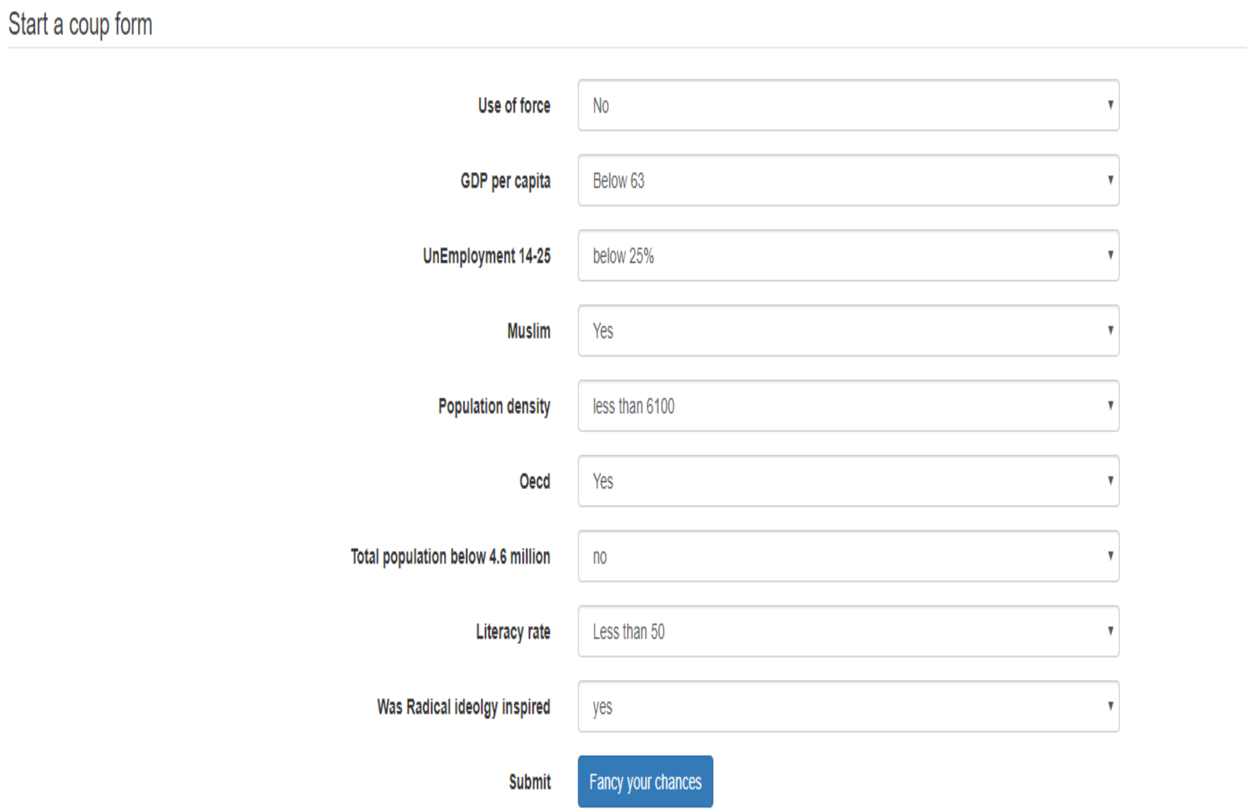


Figure 14: Start a Coup Form

- Usually the probabilities are below 50% but some inputs can lead to high amounts if some triggers are not possible

CONCLUSIONS AND FUTURE EXPANSION:

- The results suggest some features are important such as being Muslim, using force, being radical in ideology
- The sample size is for a 60-year period, so relevance will change for our data
- The median filling for NA values is not ideal at all especially for such an important topic, more complete datasets would be valuable which is why nearly none of the new induced values to the Colgan dataset show much relevance
- Another step to consider is how nonviolent revolutions would be relevant and what factors would influence it
- A sequential decision model is also something we considered implementing but the incompleteness of the data wouldn't allow it
- The model is highly complicated but some of the features have really poor p values, feature engineering is something we would need to further consider

REFERENCES:

1. Colgan 2012 CMPS
2. Blitz,Daan: Arab Spring : A parsimonious explanation of recent contentious politics
3. World Bank
4. Wikipedia (statistics)
5. Press Freedom Index
6. Gapminder
7. Measuring Revolution-Jeff Colgan First Published September 11, 2012