Statistical Analysis of the Human Freedom Index

Collin Carpenter

DS329: Data Mining

Professor Shafqat Ali Shad

18 May 2019

The dataset I used was the Human Development index from Kaggle.com, but the actual dataset was created by the Cato Institute. The purpose of the index is to create metrics to create a reasonable picture of the freedom of the world. Using measures like access to sound money, to trade internationally, and the prevalence of bribes for economic freedom. While measures such as rule of law, freedom of speech, religious freedom, and the freedom of association for personal freedom. The data itself is 179,334 data points on 1458 rows using 58 metrics for personal freedom, and 54 for economic. There were a fair number of NaN values for several countries due to the fact that there are certain nations such as Iraq, Syria, and Afghanistan in the years 2008 to 2016 whose political climate do not lend themselves to quality data collection.

Many of the metrics did not have high correlation values with the targets so I chose to drop them from my find dataset before modeling. Leaving me with rule of law, movement, religion, association, assembly, and civil society, identity and relationship, security and safety for personal freedoms. Legal system and property rights, access to sound money, freedom to trade internationally, regulation of credit, labor, and business for economic. Splitting the data into training and test sets was done using a simple 80-20 split on 54/58 metrics for the x, and the same for the target variables of score, and ranks, as well as the target I created which was the sum of the two scores.

The model that I chose to use as my final was randomforest since it gave me my highest accuracy score of 73%. I did have to label encode my response variables for the model to work, however. Many of the models and scores that I have used in prior dataset worked on binary response variables, so I had to improvise and create models that fit my data. That it also part of the reason that the accuracy of 73% was relatively low compared to other predictive models I have made in the past. This model is the best one especially compared to some other models such as the decision tree that gave 69%, and naïve bayes which gave a terrible .002%.

In conclusion, my data analysis was insightful and revealing. Going into this analysis, I was working under the assumption that the freest nations and regions in the world would most likely be the nations of the western economic and philosophical tradition. When, the one nation that was ranked number one in both economic and personal freedom was the special administrative region of the People’s Republic of China; Hong Kong. This result was fairly surprising to me since I know that Hong Kong served as a colonial possession of the British for centuries, and then passed back into the hands of the mainland Chinese in 1997. Also, looking at values such as the mean ratios of scores and ranks gives us the revealing values that show western nations such as the United States rank in the top 11 in terms of inequality between the two values. Before taking on the freedom index problem, I originally intended on doing a time series analysis of the Four Asian Tigers of Hong Kong, Taiwan, South Korea, and Singapore because of their incredible economic growth from the 1960s on. So I expected to see a correlation between those four nations, but the one country that stood out the most of the four was Singapore which had a mean ratio of human freedom to economic freedom of 15.00, while the next highest nation, the United Arab Emirates, had a value of only 4.231. These findings are indicative of a need for further and deeper analysis, perhaps one the focuses on these four nations.

**Metrics:**

Year

ISO\_code

countries

region

pf\_rol\_procedural = Procedural justice

pf\_rol\_civil = Civil justice

pf\_rol\_criminal = Criminal justice

pf\_rol = Rule of law

pf\_ss\_homicide = Homicide

pf\_ss\_disappearances\_disap = Disappearances

pf\_ss\_disappearances\_violent = Violent conflicts

pf\_ss\_disappearances\_organized = Organized conflicts

pf\_ss\_disappearances\_fatalities = Terrorism fatalities

pf\_ss\_disappearances\_injuries = Terrorism injuries

pf\_ss\_disappearances = Disappearances, conflict, and terrorism

pf\_ss\_women\_fgm = Female genital mutilation

pf\_ss\_women\_missing = Missing women

pf\_ss\_women\_inheritance\_widows = Inheritance rights for widows

pf\_ss\_women\_inheritance\_daughters = Inheritance rights for daughters

pf\_ss\_women\_inheritance = Inheritance

pf\_ss\_women = Women's security

pf\_ss = Security and safety

pf\_movement\_domestic = Freedom of domestic movement

pf\_movement\_foreign = Freedom of foreign movement

pf\_movement\_women = Women's movement

pf\_movement = Freedom of movement

pf\_religion\_estop\_establish = Freedom to establish religious organizations

pf\_religion\_estop\_operate = Freedom to operate religious organizations

pf\_religion\_estop = Freedom to establish and operate religious organizations

pf\_religion\_harassment = Harassment and physical hostilities

pf\_religion\_restrictions = Legal and regulatory restrictions

pf\_religion = Religious freedom

pf\_association\_association = Freedom of association

pf\_association\_assembly = Freedom of assembly

pf\_association\_political\_establish = Freedom to establish political parties

pf\_association\_political\_operate = Freedom to operate political parties

pf\_association\_political = Freedom to establish and operate political parties

pf\_association\_prof\_establish = Freedom to establish professional organizations

pf\_association\_prof\_operate = Freedom to operate professional organizations

pf\_association\_prof = Freedom to establish and operate professional organizations

pf\_association\_sport\_establish = Freedom to establish educational, sporting, and cultural organizations

pf\_association\_sport\_operate = Freedom to operate educational, sporting, and cultural organizations

pf\_association\_sport = Freedom to establish and operate educational, sporting, and cultural organizations

pf\_association = Freedom to associate and assemble with peaceful individuals or organizations

pf\_expression\_killed = Press killed

pf\_expression\_jailed = Press jailed

pf\_expression\_influence = Laws and regulations that influence media content

pf\_expression\_control = Political pressures and controls on media content

pf\_expression\_cable = Access to cable/satellite

pf\_expression\_newspapers = Access to foreign newspapers

pf\_expression\_internet = State control over Internet access

pf\_expression = Freedom of expression

pf\_identity\_legal = Legal gender

pf\_identity\_parental\_marriage = Parental rights in marriage

pf\_identity\_parental\_divorce = Parental rights after divorce

pf\_identity\_parental = Parental rights

pf\_identity\_sex\_male = Male-to-male relationships

pf\_identity\_sex\_female = Female-to-female relationships

pf\_identity\_sex = Same-sex ralitionships

pf\_identity\_divorce = Divorce

pf\_identity = Identity and relationships

pf\_scorePersonal = Freedom (score)

pf\_rankPersonal = Freedom (rank)

ef\_government\_consumption = Government consumption

ef\_government\_transfers = Transfers and subsidies

ef\_government\_enterprises = Government enterprises and investments

ef\_government\_tax\_income = Top marginal income tax rate

ef\_government\_tax\_payroll = Top marginal income and payroll tax rate

ef\_government\_tax = Top marginal tax rate

ef\_government = Size of government

ef\_legal\_judicial = Judicial independence

ef\_legal\_courts = Impartial courts

ef\_legal\_protection = Protection of property rights

ef\_legal\_military = Military interference in rule of law and politics

ef\_legal\_integrity = Integrity of the legal system

ef\_legal\_enforcement = Legal enforcement of contracts

ef\_legal\_restrictions = Regulatory restrictions on the sale of real property

ef\_legal\_police = Reliability of police

ef\_legal\_crime = Business costs of crime

ef\_legal\_gender = Gender adjustment

ef\_legal = Legal system and property rights

ef\_money\_growth = Money growth

ef\_money\_sd = Standard deviation of inflation

ef\_money\_inflation = Inflation: most recent year

ef\_money\_currency = Freedom to own foreign currency bank account

ef\_money = Sound money

ef\_trade\_tariffs\_revenue = Revenue from trade taxes (% of trade sector)

ef\_trade\_tariffs\_mean = Mean tariff rate

ef\_trade\_tariffs\_sd = Standard deviation of tariffs rates

ef\_trade\_tariffs = Tariffs

ef\_trade\_regulatory\_nontariff = Nontariff trade barriers

ef\_trade\_regulatory\_compliance = Compliance costs of importing and exporting

ef\_trade\_regulatory = Regulatory trade barriers

ef\_trade\_black = Black-market exchange rates

ef\_trade\_movement\_foreign = Foreign ownership/investment restrictions

ef\_trade\_movement\_capital = Capital controls

ef\_trade\_movement\_visit = Freedom of foreigners to visit

ef\_trade\_movement = Controls of the movement of capital and people

ef\_trade = Freedom to trade internationally

ef\_regulation\_credit\_ownership = Ownership of banks

ef\_regulation\_credit\_private = Private sector credit

ef\_regulation\_credit\_interest = Interest rate controls

ef\_regulation\_credit = Credit market regulations

ef\_regulation\_labor\_minwage = Hiring regulations and minimum wage

ef\_regulation\_labor\_firing = Hiring and firing regulations

ef\_regulation\_labor\_bargain = Centralized collective bargaining

ef\_regulation\_labor\_hours = Hours regulations

ef\_regulation\_labor\_dismissal = Mandated cost of worker dismissal

ef\_regulation\_labor\_conscription = Conscription

ef\_regulation\_labor = Labor market regulations

ef\_regulation\_business\_adm = Administrative requirements

ef\_regulation\_business\_bureaucracy = Bureaucracy costs

ef\_regulation\_business\_start = Starting a business

ef\_regulation\_business\_bribes = Extra payments/bribes/favoritism

ef\_regulation\_business\_licensing = Licensing restrictions

ef\_regulation\_business\_compliance = Cost of tax compliance

ef\_regulation\_business = Business regulations

ef\_regulation = Regulation

ef\_score = Economic Freedom (score)

ef\_rank = Economic Freedom (rank)

hf\_score = Human Freedom (score)

hf\_rank = Human Freedom (rank)

hf\_quartile = Human Freedom (quartile)

**Potent Visualizations:**



A screenshot of a cell phone

Description automatically generatedA picture containing text

Description automatically generated

A close up of a logo

Description automatically generatedA picture containing screenshot

Description automatically generated

A close up of a logo

Description automatically generated

A screenshot of a social media post

Description automatically generated