



HUMAN FREEDOM INDEX

INTRODUCTION

Problem Statement

Primary Goal: Predicting Human freedom score based on Personal Freedom score and Economic freedom score for all countries.

Secondary Goal: To analyse India's position for consecutive years from 2008 - 2016 and suggest which attributes to focus on to improve its position for upcoming years.

DESCRIPTION

Human Freedom Index presents the state of human freedom in the world based on a broad measure that encompasses personal, civil, and economic freedom. HFI presents a broad measure of human freedom, understood as the absence of coercive constraint and uses 79 distinct indicators of personal and economic freedom. The following areas as follows:

Rule of Law, Security and Safety, Movement, Religion, Association, Assembly, Civil Society, Expression, Relationships, Size of Government, Legal System and Property Rights, Access to Sound Money, Freedom to Trade Internationally, and Regulation of Credit, Labour and Businesses

DATASET

The dataset contains 1458 observations and 123 variables.

The following columns in the dataset explains.

- **pf_score** - Personal Freedom score which says about other 57 pf columns.
- **ef_score** - Economic freedom score which says about other 55 ef columns
- Based on **pf_score** and **ef_score** averages, **hf_score**- Human freedom score is calculated.
- Based on **hf_score**, **hf_rank** - Human freedom rank and **hf_quartile** - Human freedom quartile is calculated.

	year	ISO_code	countries	region	ef_regulation_business_compliance	ef_regulation_business	ef_regulation	ef_score	ef_rank	hf_score	hf_rank	hf_quartile
0	2016	ALB	Albania	Eastern Europe	7.074366	6.705863	6.906901	7.54	34.0	7.568140	48.0	2.0
1	2016	DZA	Algeria	Middle East & North Africa	7.029528	5.676956	5.268992	4.99	159.0	5.135886	155.0	4.0
2	2016	AGO	Angola	Sub-Saharan Africa	6.782923	4.930271	5.518500	5.17	155.0	5.640662	142.0	4.0
3	2016	ARG	Argentina	Latin America & the Caribbean	6.508295	5.535831	5.369019	4.84	160.0	6.469848	107.0	3.0
4	2016	ARM	Armenia	Caucasus & Central Asia	6.491481	6.797530	7.378069	7.57	29.0	7.241402	57.0	2.0

5 rows × 123 columns

Fig.(i)

Fig.(i). shows the dataset

PREPROCESSING THE DATA

Complexity involved:

Since the data has 123 columns, the complexity was

- Finding relationship between the variables.
- Eliminating multicollinear features.

Cleaning the data:

- Treated null values using KNN imputer
- Dropped the average columns in the dataset and performing null value imputations for columns having less than 80% and the features reduced to 84 columns.

```
hf_score 0.0
pf_expression_influence 0.0
pf_association_political_establish 0.0
pf_association_political_operate 0.0
pf_association_prof_establish 0.0
...
ef_money_inflation 0.0
ef_money_currency 0.0
ef_trade_tariffs_revenue 0.0
ef_trade_tariffs_mean 0.0
pf_rol_procedural 0.0
Length: 84, dtype: float64
```

Fig.(ii)

Fig.(ii) shows null value treatment

PREPROCESSING THE DATA

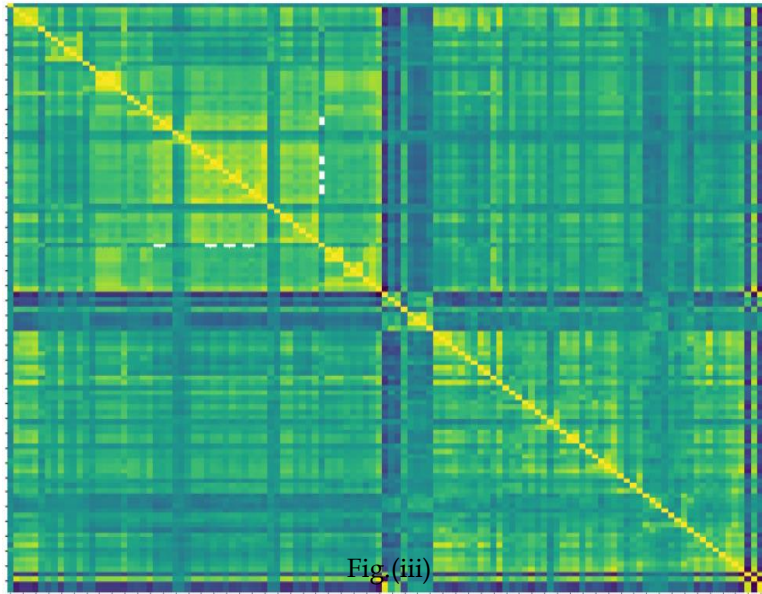


Fig.(iii)

Fig.(iii) Correlation matrix for the dataset after null value treatment.

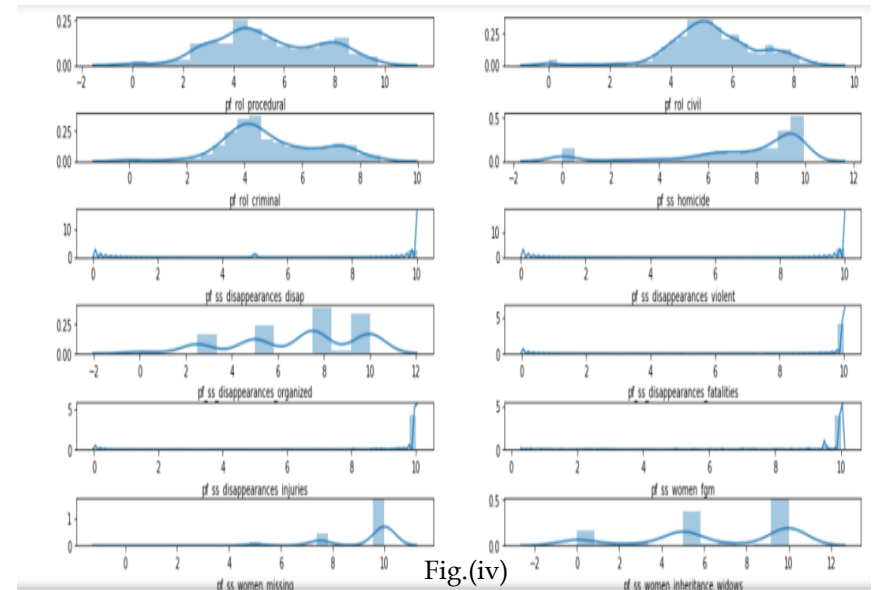


Fig.(iv)

Fig.(iv) Some variables in the dataset are skewed

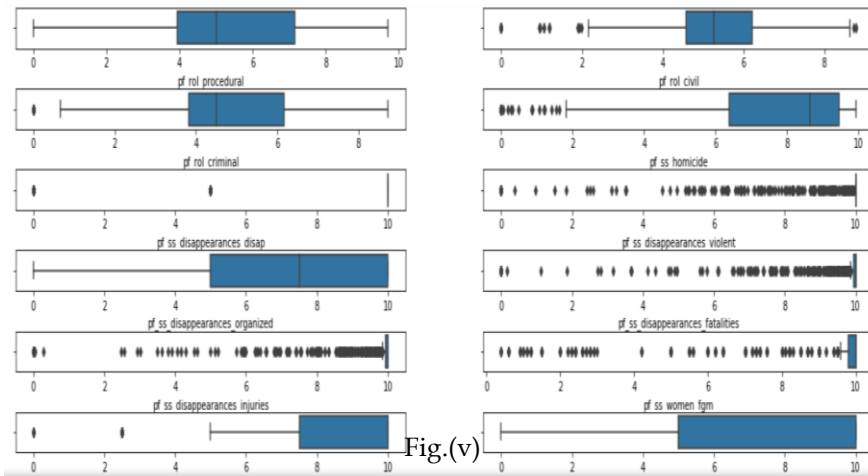


Fig.(v)

Fig.(v) Outliers are present in the data

FEATURE ENGINEERING

There are some extreme values present in the data when compared to other normal values present in the data. In feature engineering segment using winsorization (Outlier Capping) to get rid of outliers.

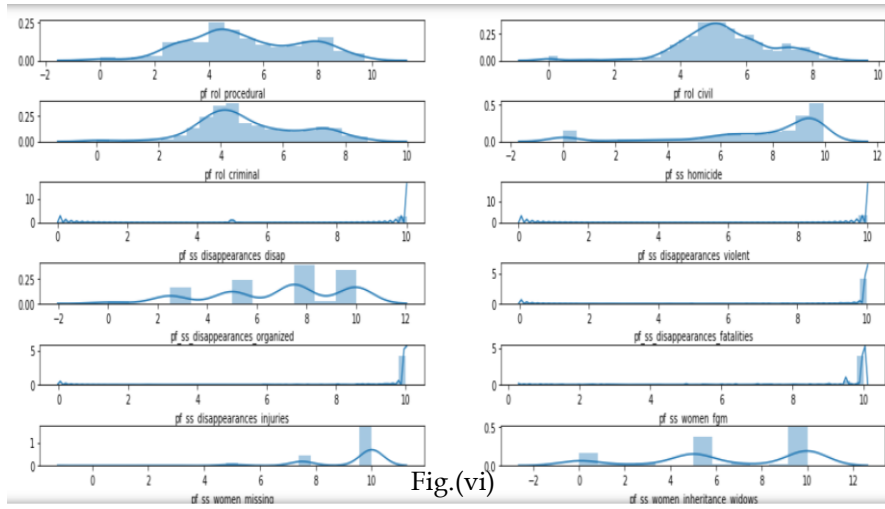


Fig.(vi) Treating skewness through power transformer

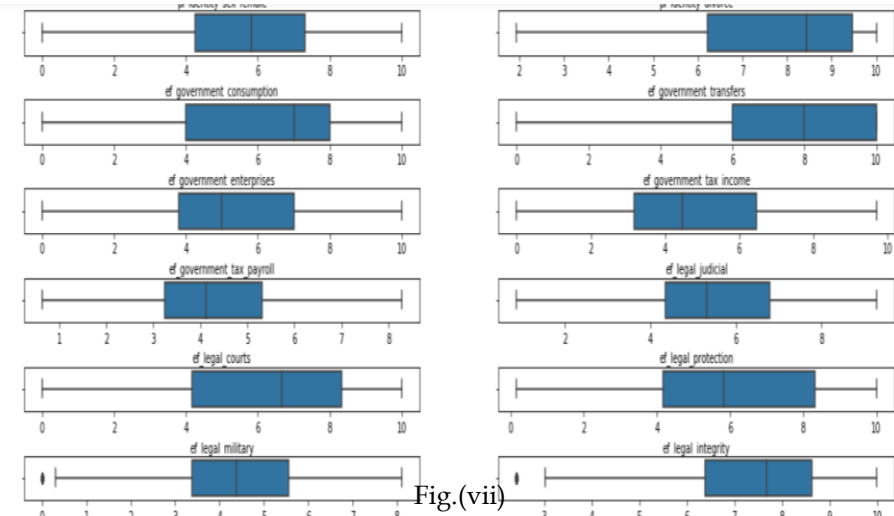


Fig.(vii) Power Transform had no effect on outlier treatment. So, to eliminate outliers using winsorization.

FEATURE ENGINEERING

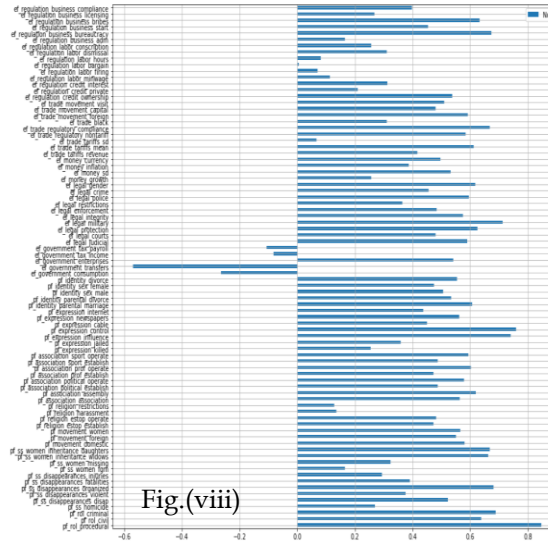


Fig.(viii) How much each variable is correlated with the target variable.

All variable values are between -1 to 1

MODEL PARAMETERS

Fitting the 72 feature columns in OLS and evaluating the model through OLS model by checking varying parameters and checking R-square and RMSE values.

1. Multicollinearity:

	VIF	feature
40	4.949525	ef_legal_integrity
26	4.890172	pf_expression_control
18	4.714106	pf_association_political_establish
17	4.601265	pf_association_assembly
39	4.236683	ef_legal_military
33	4.217498	pf_identity_divorce
21	4.146740	pf_association_prof_operate
35	4.100066	ef_government_transfers
31	4.092546	pf_identity_sex_male
43	4.056723	ef_legal_crime
13	4.032311	pf_religion_estop_establish
50	3.892490	ef_trade_tariffs_mean

Fig.(ix)

Fig.(ix) After eliminating multicollinear columns through VIF the features are reduced to 72 columns

MODEL PARAMETERS

2. Normality :

The value for normality is -0.16

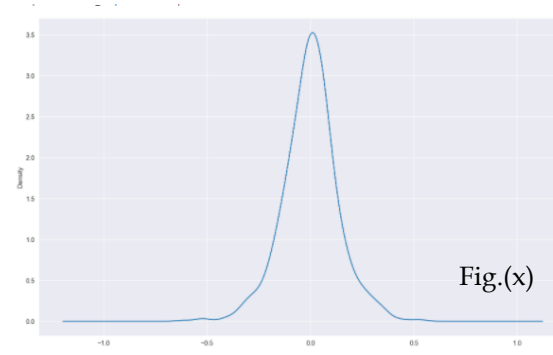


Fig.(x) The resid plot shows the data is Normal.

3. Linearity:

For Statistical test of Linear Rainbow the p- value obtained is 0.89.

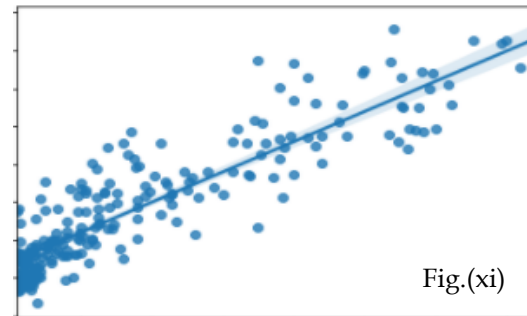


Fig.(xi) The scatterplot shows randomness with zero residue axis.

4. Auto Correlation :

Checked with respect to Durbin Watson value. Accepted range is between -1.5 to +2.5. The value obtained is 1.8.

5. Homoscedasticity:

Two Statistical tests are done to verify homoscedasticity

- i.e., Goldfeld Quandt test or Breusch-Pagen test. For Goldfeld Quandt test, Null hypothesis: The model is Homoscedastic The value obtained is 0.99. Hence rejecting Null hypothesis shows the model is Homoscedastic.

PERFORMANCE OF MODEL

Dropped insignificant features from highest order of iterations	R2	Adj.R2	RMSE Test	RMSE Train	Error
Base Model (72 features)	0.981	0.979	0.165	0.135	3%
1 st iteration (ef_regulation_labor_bargain)	0.981	0.979	0.165	0.135	3%
2 nd iteration (pf_ss_disappearances_injuries)	0.981	0.979	0.164	0.135	2.9%
3 rd iteration (ef_regulation_credit_interest)	0.981	0.979	0.164	0.135	2.9%
Final iteration (dropping remaining 23 insignificant columns)	0.980	0.979	0.167	0.137	2.9%
Final Model (Top 30 significant features)	0.982	0.981	0.272	0.293	2%

INFERENCE

Fig.(xii) The top 30 significant features for the final model

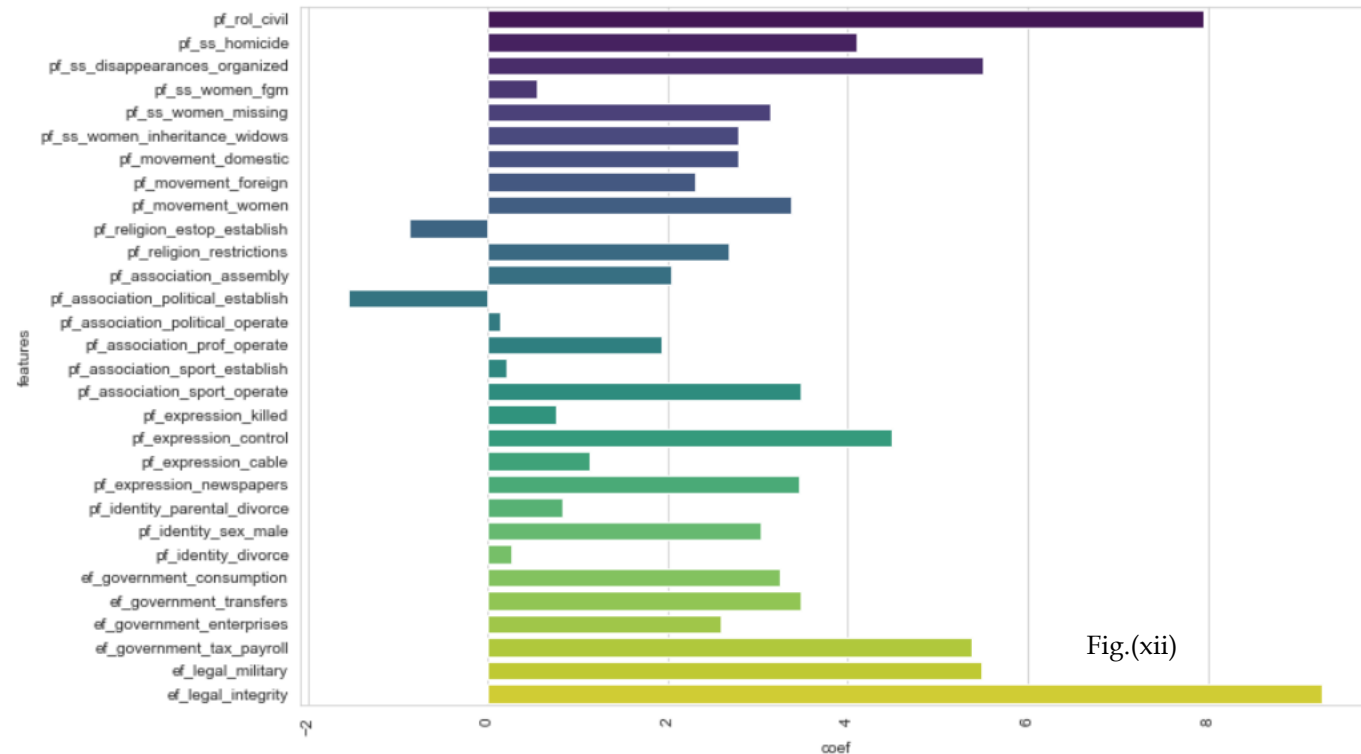


Fig.(xii)

INFERENCE

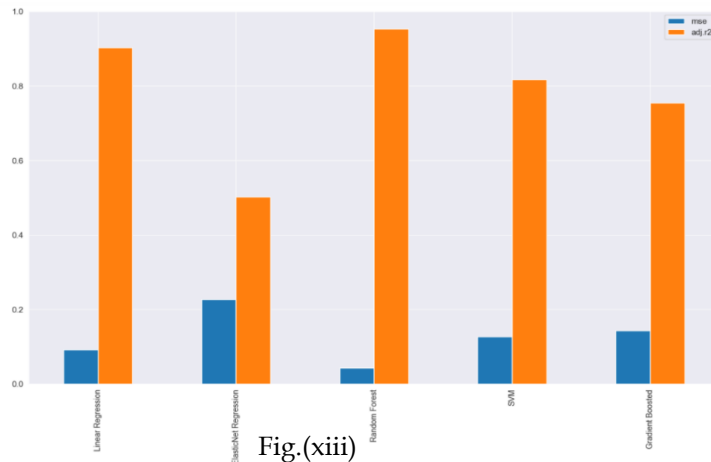


Fig.(xiii) Mean Square Error and Adjusted R-square values for different regression models

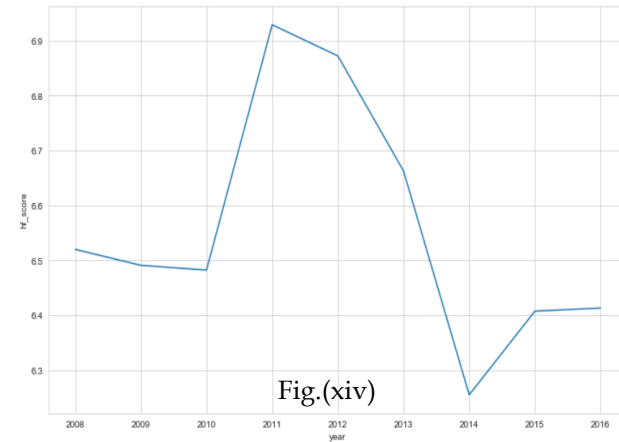


Fig.(xiv) Line plot for India for hf_score between 2008 - 2016

- The top 30 significant features are the most important features in which the countries should focus on improving the hf_score.

➤ w.r.t. India

Fig.(xv) Decreasing order of significant features

	features	coef
29	ef_legal_integrity	9.268
0	pf_rol_civil	7.948
2	pf_ss_disappearances_organized	5.502
28	ef_legal_military	5.489
27	ef_government_tax_payroll	5.378
18	pf_expression_control	4.488
1	pf_ss_homicide	4.098
16	pf_association_sport_operate	3.480
25	ef_government_transfers	3.477
20	pf_expression_newspapers	3.461

Fig.(xv)

are the top significant features present in the model which are contributing less for India where the average scores of top two significant features are 4.26 and 5 respectively which are comparatively very low than other countries scores.

So, India's should focus on these attributes to improve its position for upcoming years. 10

FUTURE SCOPE

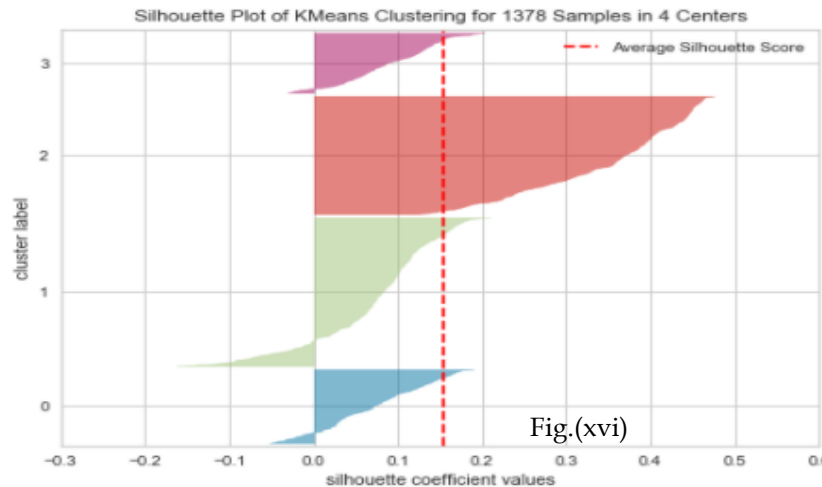


Fig.(xvi) Silhouette Score for k-means clustering for n cluster = 4

- By using four different labels obtained from silhouette score, the data can be further explained using Linear Discriminant Analysis, Multiclass models.
- Different models can be built with respect to each country.
- Rather than capping which is apt for OLS, the model can be built using Generalized linear model, Linear discriminant analysis with no outlier capping

REFERENCES

- <https://www.cato.org/sites/cato.org/files/2020-12/human-freedom-index-2020.pdf>
- <https://www.kaggle.com/gsutters/the-human-freedom-index/code>
- https://en.wikipedia.org/wiki/List_of_freedom_indices
- <https://dsekharulz.wixsite.com/group3>