Statistical Classification of Democracy Index Scores of Countries of the World

A K M Raquibul Bashar[1*], Chris P. Tsokos[2]

University of South Florida, Tampa, FL 33620, USA

Abstract		Original Research Article

167 countries around the world cover almost all the total population of the world. These countries are clustered into four different clusters as per their democracy index scores (DIS) that was introduced by the Economist Intelligence Unit (EIU). The EIU introduced this DIS first time in the year 2006. But after running some statistical tests, it is found that some countries are falling into some clusters where they are not supposed to be. So, in this study we have revisited the EIU data and found that there are some statistically significant differences among some of the countries included in the adjacent clusters. It was found that by implementing proper statistical clustering and classification techniques the number countries in each cluster was changed significantly. It is also found that the new arrangement of countries in each of the clusters are statistically robust both parametrically and non-parametrically. Another interesting finding of this study was to determine appropriate countries for each of the cluster members so that any government or non-government organizations (e.g.: IMF, WHO, ADB, WB) who are deciding body to make a monetary sanction to any specific country to improve the democracy, to vitalize the infrastructure or other development related projects can have a statistical stand point to back up their decision with certain level of certainty.
Keywords: K-means clustering, Multinomial Logistic Regression, Scree Plot, Democracy index.

INTRODUCTION

Freedom is often used as a substitute for Democracy, a form of government where the power lies with the people and was started by the ancient Greeks (Athenians) around the eighth-century B.C. It established constitutional rules to protect economic (property owners), social and political gains [1-3]. A form of government in which the supreme power is locked in the hands of the people collectively. The underlying principle of Democracy is equality and freedom for all its people and acknowledges the decision of the majority and the protection of the minority—equal voting rights regardless of stature, gender, race, etc. Democracy can be thought of as a set of rules and principles that have been implemented to protect the freedom of the people [4]. The three elements that constitute Democracy:

- **Democratic Values**- encouraging individual values, loyalty and justice and moral reciprocity
- **Structure**- institutions of decision making, justice and managerial follow through
- **Practices**- behavior that determines core values, behaviors, and actions

Of course, there is no unique definition that can describe the philosophy and beauty that constitutes Democracy. Likewise, there is no unique method to measure Democracy. For example, Democracy in the United States government is significant in our foreign policy; however, there are differences within our government about what constitutes Democracy in another country. President Bill Clinton in his inaugural address stated that "Our Democracy must be not only the envy of the world, but the engine of our OWN renewal." We also should note that Louis Horowitz in his book, "The Struggle for Democracy," states that "The world's only superpower is rhetorically and militarily promoting a political system that remains undefined…." [5]. His theme speaks volumes of what is happening today in solving cultural, regional, political conflicts in our global society. I believe that almost all differences among countries can be solved through the establishment of a powerful educational system known as Hellenic Paideia [6, 7]. This brings us to the question of how we measure the Democracy of a given country in the world. There are three methods that I am familiar with:

- Freedom House (U. S.)

- The Policy Project (U. S)
- The Economist Intelligence Unit (UK)

Personally, we believe the Democracy Index of the intelligence Unit (EIU) has collected the best data of 167 to classify them as being Full Democracy, Flawed Democracy or Hybrid Regime Democracy, or No Democracy. However, their method of analysis is simply empirical (descriptive) and not very accurate. For example, they do not take into consideration the significant interaction (interrelated) effects of the data they collected, among other difficulties. We will introduce two statistically driven models with their appropriate algorithms that will show that the Intelligence Classification Index is approximately 42% accurate. Misclassification of a certain country can be seriously economically and politically damaging.

In the present time, Democracy is one of the concerns for those countries wants to practice democratic culture in all the sectors within the infrastructure. For this purpose, any country has to be identified as one of the member of four clusters defined by Economist Intelligence Units (EIU) [8]. Democracy Index Scores (DIS) is used to categorize 167 countries of the world as per their respective scores and after certain threshold falling beyond and above the threshold, that country is categorized. In this case if the country is tagged wrongfully then the improvement of democracy quality is very important. For this purpose, correct clustering or classification should be done correctly. Mihaiela Ristei and M. Centellas has done democracy cluster classification with DCC index [9]. There are some other scholars used machine learning classification techniques such as machine learning [10]. Rahman *et al.* [11] has done some trend on the improvement of democracy over the certain period of time through Freidman's test.

In this study we have done the clustering and classification of the DIS scores. Although, we have done a parametric analysis on this DIS [12], we have done clustering with the classical statistical methods named as K-means [13] method and multinomial logistic regression method [14] to show the statistical differences among the number of cluster members. It turns out that, these classical statistical methods have performed better than other classification and clustering methods mentioned in this study.

THE ECONOMIST INTELLIGENCE GROUP'S DATA AND METHODOLOGY

The Economist Intelligence Units of democracy [8] , on a 0 to 10 scale, is based on the ratings for 60 indicators grouped in five categories mentioned above. Each category has a rating on a 0 to 10 scale and overall index of democracy is the simple average of the five category indexes. The category indexes are based on the indicator scores in the category converted to 0 to 10 scales. Adjustments to the category

scores are made if countries do not score a 1 in the following critical areas of democracy:

- Whether national elections are free and fair
- The security of the voters
- The influence of foreign powers on government
- The capability of the civil service to implement policies

If the scores for the first three questions are 0 (or 0.5), one point (0.5 point) is deducted from the index in the relevant category (either the electoral process and pluralism or the functioning of government). If the score for 4 is 0, one point is deducted from the functioning of government category index. The index values are used to place countries within one of four types of regimes:

- Full democracies--- scores of 8- 10
- Flawed democracies--- scores of 6 to 7.9
- Hybrid regimes--- scores of 4 to 5.9
- Authoritarian regimes—scores below 4

Threshold points for regime types depend on overall scores that are rounded to one decimal point. Based on the scores defining different types of regimes for any country to be fell in the definitions. For scoring system the EIU has used a combination of a dichotomous and a three- point scoring system for the 60 indicators [8]. According to their claim, a dichotomous 1-0 scoring system (1 for yes and 0 for no) has some drawbacks, but it has several distinct advantages over more refined scoring scales (such as 1-5 or 1-7). Also they say, for many indicators, the possibility of a 0.5 score is introduced, to capture grey areas where a simple yes or no is problematic with guidelines as to when that should be used. Thus for many indicators there is a three- point scoring system, which represents a compromise between simple dichotomous scoring and the use of finer scales. The also declare that, a crucial, differentiating aspect of their measure is that in addition to experts' assessments they use, where available, public opinion surveys- mainly the World Values Survey (Say, WVS). Indicators based on the surveys predominate heavily in the political participation and political culture categories, and a few are used in the civil liberties and functioning of government categories. In addition to the WVS, other sources that can be leveraged include the Euro-barometer surveys, Gallup polls, Asian Barometer, Latin American Barometer, Afro barometer and national surveys. In the case of countries for which survey results are missing, survey results for similar countries and expert assessments are used to fill in gaps collect and fulfill the data for scoring.

DATA SOURCES AND DATA STRUCTURE

Primarily, it was the same data the Economist Intelligence Unit collected from all over world by their assigned sources and gathered and analyzed by them on

determining the type of democracy of any of the 167 countries of the world. The Economist Intelligence Unit's Democracy Index provides a snapshot of the state of democracy worldwide for 165 independent states and two territories— this covers almost the entire population of the world and the vast majority of the world's states (micro states are excluded). To verify and evaluate, we have used our own methodology and techniques: a). the K- Means Method [13] to cluster the country and b). Statistical model, Multinomial Logistic Regression[14] that is quite efficient, a digital algorithm that is based on the proposed model that once someone enter the data then automatically receive the probability of falling in one of the four categories that will identify the type (one of the four regimes) of democracy the country is classified.

The importance of the subject study is that, if anyone use the model to identify and classify the type of country's regime type then they have to put in the scores for the five categories and it will give you score to identify that country as one of the four regimes mentioned earlier. This model can be used by the following organizations and officials such as- NATO, Country Administrators, financial Institutions (Such as, IMF, World Bank, ADB, etc., UN, and Private Industries to invest their valuable resources as correctly as possible with better confidence and peace of mind not to worry about their monetary returns. Following Fig-1 is presented to give an idea of the data collection and data flow process by EIU:
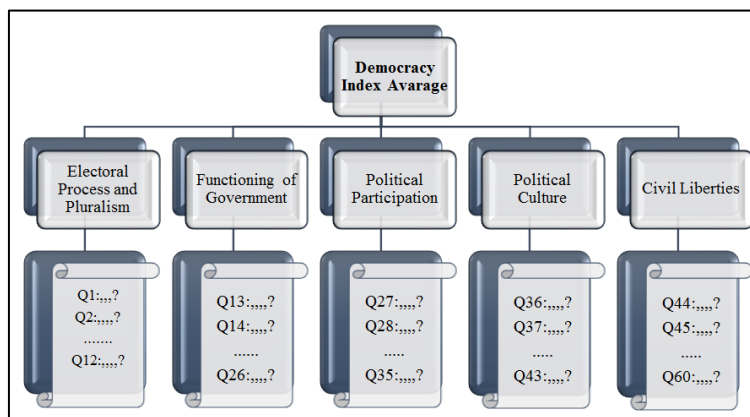

Fig-1: Data Diagram of EIUs Democracy Index Scores

PROPOSED CLUSTERING METHOD
Then we plotted the simple bar chart of the classification/clustering done by the Economist Intelligence method. Fig-2 below shows the simple bar chart that shows the number countries in each of the classification of government. According to economists the following is the frequency table for various formats of democratic countries.
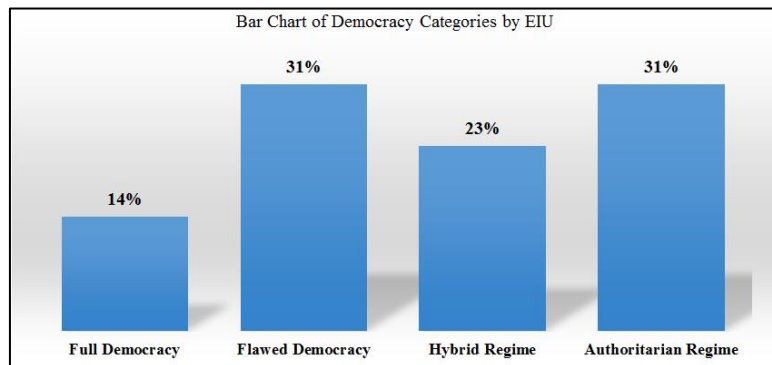

Fig-2: Bar Chart of EIUs Classification of Democracy

But, after having a closer and careful look at the dataset, we have decided to do our own classification techniques to cross validate the classification done by the Economists Intelligence Unit (EIU). And we have started with K-Means clustering algorithm that is well appropriate for this kind of classification and clustering. This method is used to verify and evaluate the classification and clustering done by Economists' we have applied K-means clustering and classification technique. The algorithm for this method is as follows:

- The basic step of k-means clustering is to determine number of clusters K
- We assume the centroid or center of these clusters. We can take any random objects as the initial centroids or the first K objects in sequence and can also serve as the initial centroids

- Then the K-means algorithm will perform the three steps given below until convergence
- Iterate until stable:
    a. Determine the centroid coordinate
    b. Determine the distance of each object to the centroids

c. Group the object based on minimum distance

The following flowchart displays in Fig-3 shows the process for this type of clustering and classifications:
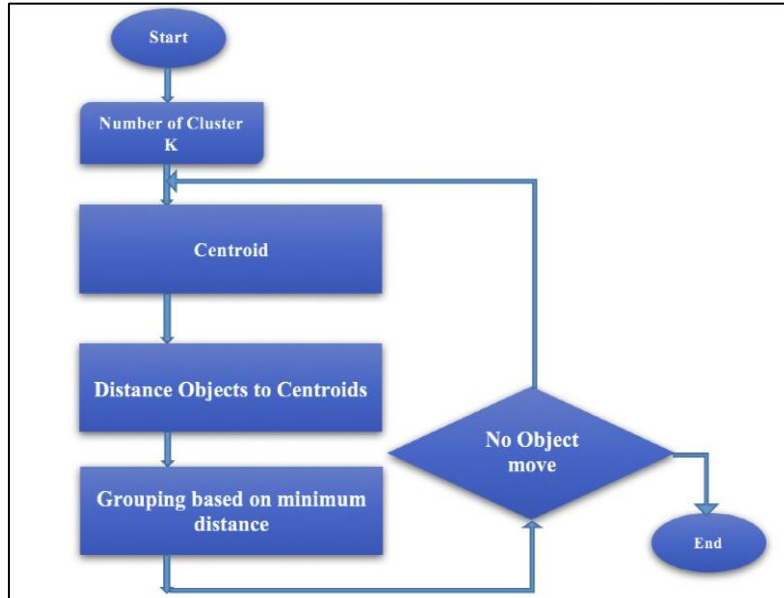

Fig-3: Flowchart design for K-Means clustering technique

In accordance with the steps mentioned above first of all we need to determine the number of appropriate clusters applicable to the data set. In that regards we have used the following graphical technique.

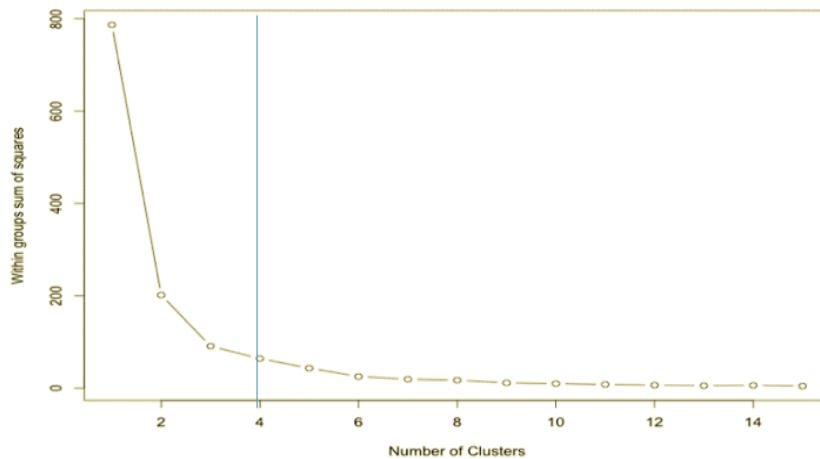**Determining the number of clusters**

Fig-4: Scree plot for determining number of clusters

From the above Fig-4, we see that the bend of the curve became stabled from cluster number 4. So, this graph suggests the appropriate number of clusters would be K= 4. Since we have determined the number of clusters we should have in our dataset, now we are ready to execute our K-means method techniques to identify the correct cluster with all of the individuals distributed into each of the separate clusters.

Validation of the K-means Method Clustering and Classification:

To validate our method, we should look at the convergence and if we can attain the convergence on or before 10 iterations showed in Table-1, then we should conclude that the clustering and classification technique we have utilized here as K-means method is one of the good if not the best solutions if we look at the iteration history of this clustering algorithm. Following table

shows the convergence of K- Means clustering method:

Table-1: Convergence of K-means clustering

Iteration History				
Iteration	Change in Cluster Centers			
	FD	HR	FLD	AR
1	2.240	3.657	3.114	4.492
2	0.436	0.170	0.380	0.367
3	0.319	0.049	0.188	0.065
4	0.270	0.085	0.133	0.069
5	0.229	0.069	0.184	0.000
6	0.263	0.068	0.242	0.000
7	0.255	0.142	0.319	0.000
8	0.000	0.224	0.096	0.069
9	**.000**	**.000**	**.000**	**.000**
FD = Full Democracy, HR = Hybrid Regime, FLD = Flawed Democracy, AR = Authoritarian Regime				

Comparison of Democracy Categories by EIU and proposed method

After executing the algorithm of K-means clustering method we have found that all the classification done by our k-means method exhibiting a difference in the number of members for each of the clusters. The Bar chart given in Fig-5 for type of Democracy by classified by K-Means Method:
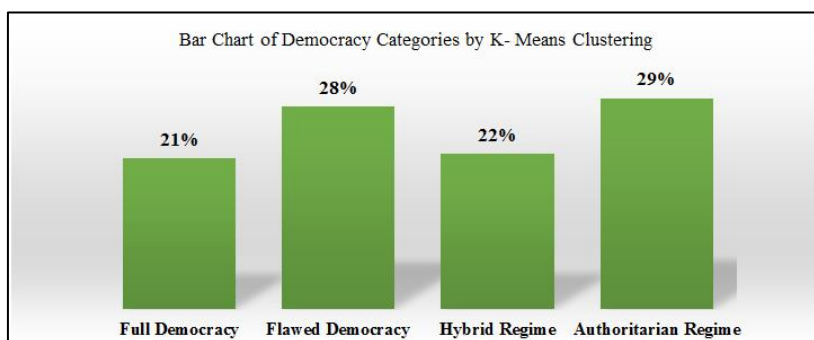

Fig-5: Bar Chart of EIUs Classifiaction of Democracy after Clustering

We have showed the comparative analysis for the number of members of the each cluster in the following Table-2 is compared with the results of Economists':

Table-2: Comparison of cluster members of EIU and K-means

Type of Democracy	Economists Classification	Our Classification (K-means Method)
Full Democracy	24	35
Flawed Democracy	52	47
Hybrid Regime	39	36
Authoritarian Regime	52	49
Total	167	167

Now if we want to compare the proportions for each of the clusters for both classifications we have postulated a side- by- side comparison of the above phenomena given in the Fig-6 as follows:
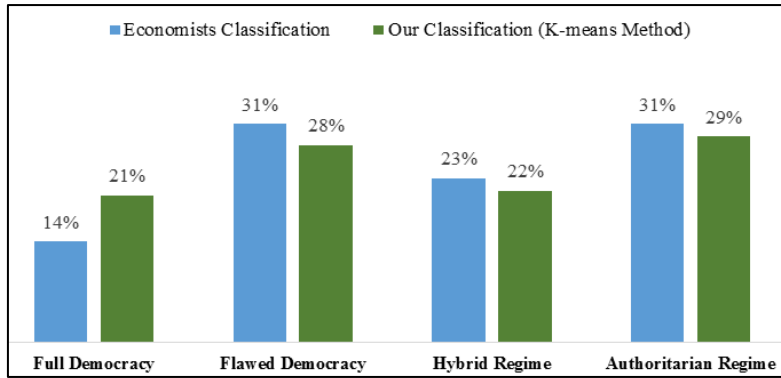
Fig-6: Comparison of Cluster proportion

In addition, if anyone wants to investigate the percentage amount of disagreement for each of the regime categories then he/she should look at the following Table-3:

Table-3: Disagreement Percentage between EIU and K-means Clustering

Type of Democracy	Economist Classification	Our Classification (K-means Method)	Disagreement (Per Class)	Overall
Full Democracy	14%	21%	7%	13%
Flawed Democracy	31%	28%	3%	
Hybrid Regime	23%	22%	1%	
Authoritarian Regime	31%	29%	2%	

As we can see from the table above that the overall misclassification is 13% by the method of classification by EIU. Also, we have run the Non-parametric Wilcoxon- sign rank test and marginal-homogeneity test and sign- test. From Table-4, we see that in all non- parametric tests all the null- hypothesis have been rejected.

Table-4: Hypothesis test summary between EIU and K-means classification

Hypothesis Test Summary				
	Null Hypothesis	**Test**	**Sig.**	**Decision**
1	The median of differences between EIU's Classifications/ Categories and Proposed Method Classification equals 0.	Related-Samples **Sign Test**	3.588E-5[1]	*Reject* the null hypothesis.
2	The distributions of different values across EIU's Classifications/ Categories and Proposed Method Classification are equally likely.	Related-Samples **Marginal Homogeneity Test**	0.000	*Reject* the null hypothesis.
3	The median of differences between EIU's Classifications/ Categories and Proposed Method Classification equals 0.	Related-Samples **Wilcoxon Signed Rank Test**	0.000	*Reject* the null hypothesis.
Asymptotic significances are displayed. The significance level is .05.				
Exact significance is displayed for this test.				

This indicates that the differences of classification for different categories between EIU and our proposed method are statistically significant. Also, comparative figures are shown in Fig-9 and
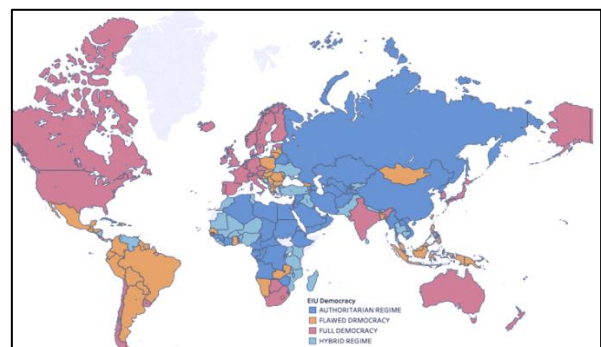

Fig-10 in Appendix A. Moreover, we have ranked the variables according to their importance into the classifications and clustering the Democratic data.

We found that "Electoral Process and Participation" is the most important predictor for Democracy and "Political Participation" is the least important predictor for classifying the type of democracy involved in a country from Fig-7.
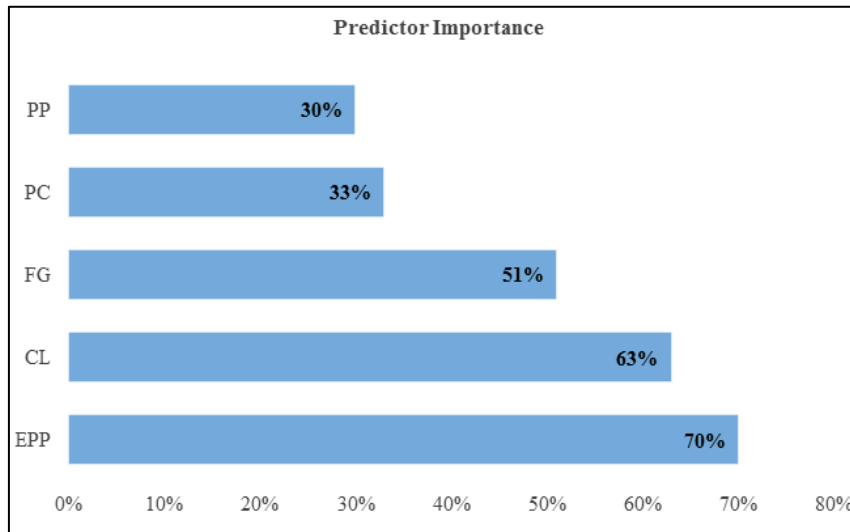

Fig-7: Ranking of Attributable according to importance of contribution to clustering

PROBABILISTIC (STATISTICAL) MODEL FOR DEMOCRACY

So far, we have done the clustering of the countries according to K-means methods successfully, now someone might need a statistical model to predict and identify a country's appropriate democratic strata. To suffice those, we have tried to formulate a statistical model so that anyone can use this model to correctly identify a country's category of democracy with certain level of confidence. To do so, we are proposing the following algorithm which will enable us to predict a new country's appropriate democratic classification. The following algorithm is the first step to identify a proper model from statistical view point. Since the response variable is a categorical variable such as "Full Democracy", "Flawed Democracy", "Hybrid Regime", and "Authoritarian Regime" and our purpose is to classify according to prediction probability for different countries, so, we started with the simple additive logistic model considering no interaction between and among the attributable variables. To do so, we have derived the following step by step algorithm in two separate parts. The reasoning behind two separate parts will become vivid at the end of the total presentation of model formulation algorithm.

Non- linear Statistical (Multinomial Logistic) Model

In this model, we have used and defined the entire attributable variable as we should and then we have tried to formulate a model with greater accuracy and improvement of the model's prediction quality and confidence.

Defining Variables and analytical model

EPP= *Electoral Process and Pluralism*,   FG= *Functioning of Government*
PP= *Political Participation*   PC= *Political Culture*
CL= *Civil Liberties*

Also, let's define following variables for ease of statistical modeling.

$$X_1 = EPP, \quad X_2 = FG, \quad X_3 = PP, X_4 = PC, \quad X_5 = CL$$

$$
\begin{cases}
(1).\ log\left(\frac{P_1}{P_4}\right) = \alpha_1 + \sum_{k=1}^{5} \beta_{1k}X_k \\[2mm]
(2).\ log\left(\frac{P_2}{P_4}\right) = \alpha_2 + \sum_{k=1}^{5} \beta_{2k}X_k \\[2mm]
(3).\ log\left(\frac{P_3}{P_4}\right) = \alpha_3 + \sum_{k=1}^{5} \beta_{3k}X_k
\end{cases}
$$

To find the alternative log Odds ratios of the equation in equation (1) we can use the following equations:

$$\left\{ \begin{array}{l} log\left(\dfrac{P_1}{P_2}\right) = (\alpha_1 - \alpha_2) + \displaystyle\sum_{k=1}^{5}(\beta_{1k} - \beta_{2k})X_k \\[2ex] log\left(\dfrac{P_1}{P_3}\right) = (\alpha_1 - \alpha_3) + \displaystyle\sum_{k=1}^{5}(\beta_{1k} - \beta_{3k})X_k \\[2ex] log\left(\dfrac{P_2}{P_3}\right) = (\alpha_2 - \alpha_3) + \displaystyle\sum_{k=1}^{5}(\beta_{2k} - \beta_{3k})X_k \end{array} \right\}$$

Here, in the above equations,

P_1 = Probability of country being in Full Democracy Category

P_2 = Probability of country being in Hybrid Regime Category

P_3 = Probability of country being in Flawed Democracy Category

P_4 = Probability of country being in Authoritarian regime Category

Estimating Co-efficients

Table-5: Estimated co-efficients of Mult. Logistic Regression

Category of Democracy		Beta	Estimated Values
Full Democracies	Intercept	α_1	-2745.260
	EPP	B_{11}	61.783
	FG	B_{12}	83.986
	PP	B_{13}	90.044
	PC	B_{14}	89.995
	CL	B_{15}	94.726
Flawed Democracies	Intercept	α_2	-938.165
	EPP	B_{21}	33.324
	FG	B_{22}	33.185
	PP	B_{23}	39.802
	PC	B_{24}	41.377
	CL	B_{25}	42.636
Hybrid Regimes	Intercept	α_3	-401.469
	EPP	B_{31}	24.037
	FG	B_{32}	13.215
	PP	B_{33}	25.297
	PC	B_{34}	19.277
	CL	B_{35}	18.090
Authoritarian Regime	Reference Category		

GOF (Goodness- of- fit) test for proposed Multinomial Logistic Model

After estimating the co-efficients of multinomial logistic regression model as listed in Table-5 of the proposed model we can formulate the following analytical model to identify the proper category of any country of the world.

$$\left\{ \begin{array}{l} (1).\ \ log\left(\dfrac{P_1}{P_4}\right) = -2745.26 + 61.78 \times EPP + 83.98 \times FG + \\ \qquad\quad 90.04 \times PP + 89.99 \times PC + 94.73 \times CL \\[2ex] (2).\ \ log\left(\dfrac{P_2}{P_4}\right) = -938.16 + 33.32 \times EPP + 33.18 \times FG + \\ \qquad\quad 39.80 \times PP + 41.38 \times PC + 42.64 \times CL \\[2ex] (3).\ \ log\left(\dfrac{P_3}{P_4}\right) = -401.47 + 24.04 \times EPP + 13.22 \times FG + \\ \qquad\quad 25.30 \times PP + 19.28 \times PC + 18.09 \times CL \end{array} \right\}$$

As we can see from

Table-6 at the following that the model fitted in equation (3) is a good fit of the proposed model with a better AIC and BIC criterion.

Table-6: Criterion for Multinomial Logistic Regression Model

Model Fitting Information			Likelihood Ration Test		
Model	Model fitting Criteria		-2 log-likelihood	Chi-square	d. f.	Sig.
	AIC	BIC				
Final	36.00	92.124	00.00	449.247	15	0.000

Also, R^2 has been **0.932** and this clearly indicates that the model fitted in equation (3) is the nearly perfect model with a greater accuracy than any other proposed model at the later part of this study.

Table-7: Pseudo R-square for Multinomial Logistic model

Pseudo R-square	
Cox and Snell	**0.932**

Calculating Probability of the Proposed Multinomial Logistic Model

To calculate the Probabilities using the estimated coefficients we will use the following equations:

$$P_1 = P(Y = 1) = \hat{\pi}_1 = \frac{e^{\alpha_1 + \sum_{k=1}^{5} \beta_{1k} X_k}}{1 + e^{\alpha_1 + \sum_{k=1}^{5} \beta_{1k} X_k} + e^{\alpha_2 + \sum_{k=1}^{5} \beta_{2k} X_k} + e^{\alpha_3 + \sum_{k=1}^{5} \beta_{3k} X_k}}$$

$$P_2 = P(Y = 2) = \hat{\pi}_2 = \frac{e^{\alpha_2 + \sum_{k=1}^{5} \beta_{2k} X_k}}{1 + e^{\alpha_1 + \sum_{k=1}^{5} \beta_{1k} X_k} + e^{\alpha_2 + \sum_{k=1}^{5} \beta_{2k} X_k} + e^{\alpha_3 + \sum_{k=1}^{5} \beta_{3k} X_k}}$$

$$P_3 = P(Y = 3) = \hat{\pi}_3 = \frac{e^{\alpha_3 + \sum_{k=1}^{5} \beta_{3k} X_k}}{1 + e^{\alpha_1 + \sum_{k=1}^{5} \beta_{1k} X_k} + e^{\alpha_2 + \sum_{k=1}^{5} \beta_{2k} X_k} + e^{\alpha_3 + \sum_{k=1}^{5} \beta_{3k} X_k}}$$

$$P_4 = P(Y = 4) = \hat{\pi}_4 = 1 - \{\hat{\pi}_1 + \hat{\pi}_2 + \hat{\pi}_3\}$$

From Table-7, we can have the idea of the accuracy of the model which is 93.2% as per our analysis of the dataset at hand.

GENERAL LOGISTIC MODEL

In this part of statistical modeling, we have started with a logistic model in the purpose in mind that all the countries will be classified either "Full democratic" or "Not. Democratic" countries. Following is the detailed algorithm of that model formulation.

Defining Variables

EPP= Electoral Process and Pluralism, **FG**= Functioning of Government
PP= Political Participation   **PC**= Political Culture
CL= Civil Liberties

Also, let's define following variables for ease of statistical modeling.

$$X_1 = EPP, \quad X_2 = FG, \quad X_3 = PP, X_4 = PC, \quad X_5 = CL$$

$$log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$$

Estimating Co- efficients of Logistics Regression

Table-8: Estimated Co-efficients of Logistic Regression

Attributes	Co-efficient ($\hat{\beta}_i$)

	$i = 1, \dots, 5$
Constant ($\hat{\beta}_0$)	-1684.666
EPP	25.193
FG	47.065
PP	46.833
PC	47.229
CL	48.771

Analytical Model of Logistic Regression

$$log\left(\frac{p}{1-p}\right) = -1684.67 + 25.193 \times EPP + 47.065 \times FG + 46.833 \times PP + 47.229 \times PC + 48.771 \times CL$$

Calculating Probability from the Logistic Model

If anyone wants to know the correct classification of any country to be fully democratic country or not can use the probability calculation formula given in the Table-9 at the following. So, after inputting all the necessary information in the following formula if the outcome come out closer to 1 then that country will be defined as democratic country and if the value comes out close to zero, then that country should be defined as not fully democratic country.

Table-9: Probabilistic Structure of Logistic Regression

Estimated Probability calculations	Code	Equation
$P(Y = 1)$	1= Full Democratic	$P(Y = 1) = \dfrac{e^{\beta_0 + \sum_{i=1}^{5} \beta_i X_i}}{1 + e^{\beta_0 + \sum_{i=1}^{5} \beta_i X_i}}$
$P(Y = 0)$	0= Not Fully Democratic	$1 - P(Y = 1)$

Also, this model will give us some sort of statistical classification but it might not be the best model, since we are losing and overlooking some important information of the classification by means of combining three categories of democracy (Flawed, Hybrid, and Authoritarian) in one category defined by "Not Fully Democratic"

Residual Analysis of the Logistic Model

Now, we are at the final stage of evaluating the proposed statistical model (6) and in the purpose of doing that we have tried the ROC curve showed in to quantify the proposed model and if the area under the curve is close to 1 then the proposed model is one of the best model and as we can see from Table-10 that the model is one of the better model if not the best with area of 0.951 and we can observe that the proposed statistical model is statistically significant.
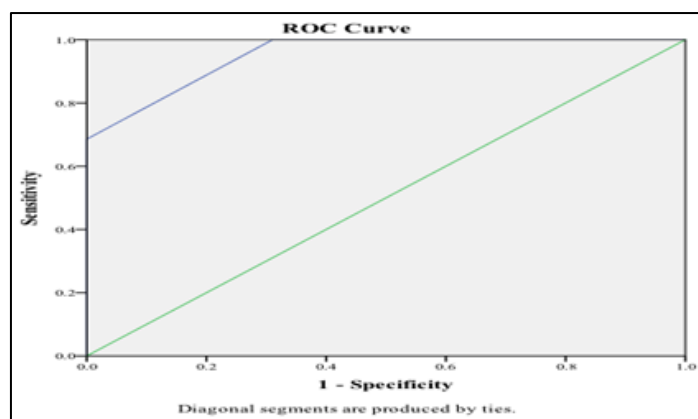

Fig-8: ROC of Logistic Regression

Table-10: Logistic Regression Model Accuracy

Area Under the Curve				
Test Result Variable(s): Category of Democracy				
Area	Std. Error	Asymptotic Sig.	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound

0.951	0.017	0.000	0.918	0.984

Estimating Co-efficients for Logistic model with interactions

Table-11: Estimated Co-efficients of Logistic Regression Model

Attributes	Co-efficient ($\hat{\beta}_i$), $i = 1 \dots, 8$
Constant ($\hat{\beta}_0$)	-347.731
EPP	-61.605
FG	-180.892
PP	50.506
PC	45.642
CL	39.455
CL*EPP	-5.172
CL*FG	8.344
EPP*FG	16.082

Analytical Structure of the Logistic model with 2- way interactions

So, after finding the co-efficients we have constructed the statistical model with interactions. The model in equation (6) is the one we have found.

$$log\left(\frac{p}{1-p}\right) = -347.731 - 61.605 \times EPP - 180.892 \times FG + 50.506 \times PP$$
$$+ 45.642 \times PC + 39.455 \times CL - 5.172 \times (CL \times \text{EPP}) + 8.344$$
$$\times (CL \times \text{FG}) + 16.082 \times (EPP \times FG)$$

Calculating Probability from Logistic model with 2-way interaction

Now, if anybody still wants to use this model and calculate the probability for determining the probability of falling in one of the classification of democracy then the following table might the useful one with consideration of significant interaction terms.

Table-12: Probabilistic Structure of Logistic Regression model with Interactions

Estimated Probability calculations	Code	Equation
$P(Y = 1)$	1= Full Democratic	$P(Y = 1) = \dfrac{e^{\beta_0 + \Sigma_{i=1}^{8} \beta_i X_i}}{1 + e^{\beta_0 + \Sigma_{i=1}^{8} \beta_i X_i}}$
$P(Y = 0)$	0= Not Full Democratic	$1 - P(Y = 1)$

CONTRIBUTION AND CONCLUSION

We have revisited the methodology and experimental design to collect the information using EIU data. But, our K − means clustering is statistically significant and commonly used method and have found better classification than the EIU. We have utilized a logistic regression model subject to data and this only classifies in to two categories. But to be consistent with the EIU's standard definition we have used multinomial logistic regression model and this model has explained the variation as good as 95.1% accuracy. The following

Table-13 shows a comparison of all the methods applied to the all the models.

Table-13: Comparison of R2 among different methods

Modeling Methods	Pseudo R^2
Logistic Regression	0.561
Multinomial Logistic Regression	0.935

In this study, to do the clustering and classification we have used K-means clustering technique that turns out to be a very good classification technique particularly for this type of dataset. Also, based on the data generated by k-means method, we have done a probabilistic classification of DIS data with multinomial logistic regression, classical logistic regression with and without the considering the interaction terms in the model. For, each of the clustering and classification techniques we have used in this study, it turns out that K-means and multinomial logistic regression are the best statistical clustering and classification techniques for analyzing this type of data.

REFERENCES

1. Bahmueller CF. Elements of democracy: The fundamental principles, concepts, social foundations, and processes of democracy. Center for Civic Education; 2007.
2. Bühlmann M. The quality of democracy; crises and success stories. In IPSA-ECPR joint conference 2011 Feb 16 (pp. 16-19).
3. Bühlmann M, Merkel W and Wessels B. "The quality of democracy: democracy barometer for established democracies," 2008.
4. Habermas J. Constitutional democracy: a paradoxical union of contradictory principles?. Political theory. 2001 Dec;29(6):766-81.
5. Kalenborn C, Lessmann C. The impact of democracy and press freedom on corruption: Conditionality matters. Journal of Policy Modeling. 2013 Nov 1;35(6):857-86.
6. Dunning WA and Barker E. "The Political Thought of Plato and Aristotle." Polit. Sci. Q., 1907.
7. Tsokos CP. "The catalyst for democracy: Hellenism and Hellenic Paideia," Univ. South Florida, 2016.
8. Kekic L. "The Economist Intelligence Unit's index of democracy," Econ, 2007.
9. Gugiu MR, Centellas M. The democracy cluster classification index. Political Analysis. 2013;21(3):334-49.
10. Gründler K, Krieger T. Democracy and growth: Evidence from a machine learning indicator. European Journal of Political Economy. 2016 Dec 1;45:85-107.
11. Rahman MS. Statistical analysis of democracy index. Humanomics. 2014 Nov 10;30(4):373-84.
12. Bashar AR, Tsokos CP. Statistical Parametric Analysis on Democracy Data. Open Access Library Journal. 2019 Oct 1;6(10):1-8.
13. A Moore. "K-means and Hierarchical clustering," Stat. Data Min. Tutorials, 2001.
14. Böhning D. "Multinomial logistic regression algorithm," Ann. Inst. Stat. Math, 1992.
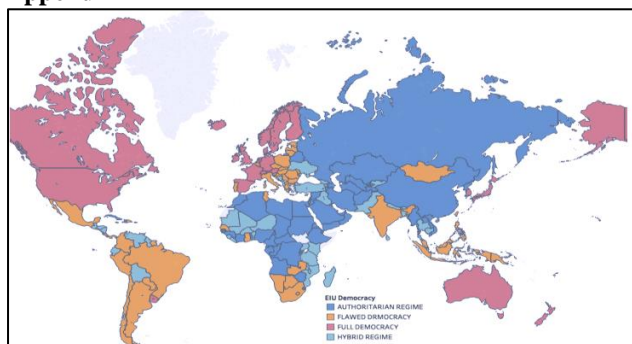
Appendix

Fig-9: Country Clustered based on EIUs original Data
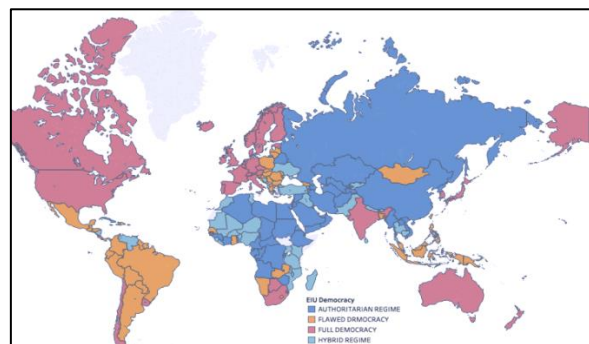

Fig-10: Country Clusters based on K-means & Multinomial Logistic Regression Model