Understaning Happiness

Xue Feng, Appilineni Kushal and Girish Kumar

June 2021

1 Abstract

In this project, we apply various machine learning algorithms to the data collected from world happiness report between the years 2007 - 2020. First, we find the correlation between the happiness score and each of the 6 indicators (logged GDP, social support, healthy life expectancy, freedom to make life choices, generosity and lower corruption) and plot a time series. We see that happiness is strongly correlated to logged GDP, social support, healthy life expectancy and weakly (or negatively) correlated to freedom, generosity and perception of corruption. Furthermore, there are very little changes in these correlations over the years as seen in the time series plot. We then apply Kmeans, Agglomerative Clustering, Affinity Propagation clustering algorithms to the data to see any underlying clustersing. We find no meaningful clustering in the data. Next, use a suite of modelling approaches to find the best model which could predict happiness score when fed with raw data of a country. Using partial dependency plots, we determine that happiness is strongly correlated to logged GDP, social support, healthy life expectancy, as concluded earlier. We explain the erroneous predictions even in our best fit model using LIME. Lastly, we model countries belonging to a particular region to check for features previously not captured. We see that happiness score is not related (weakly dependent) on health life expectancy in European countries, which is a deviation from previous results.

2 Introduction

World Happiness Report is an annual survey of the state of happiness that ranks 156 countries by how happy their citizens perceive themselves to be. The happiness score is based on the answer of individuals to a survey conducted by the "Gallup World Poll" where people across different countries were asked to rate their own happiness in life on a scale of 0 to 10. Also, a weighted estimate was then calculated amongst the 6 factors - logged GDP, social support, healthy life expectancy, freedom to make life choices, generosity and lower corruption - determining the extent to which each of them contributed to their happiness, in each country. This report gains global recognition as governments, organizations and civil society use happiness indicators to inform their policy-making decisions. Experts from various diverse fields of economics, psychology, survey analysis, national statistics, health, public policy etc., describe how these measurements of well-being can be effectively used to assess the progress of nations. The reports shows the state of happiness in the world today and informs how the new science of happiness explains national variations in happiness.

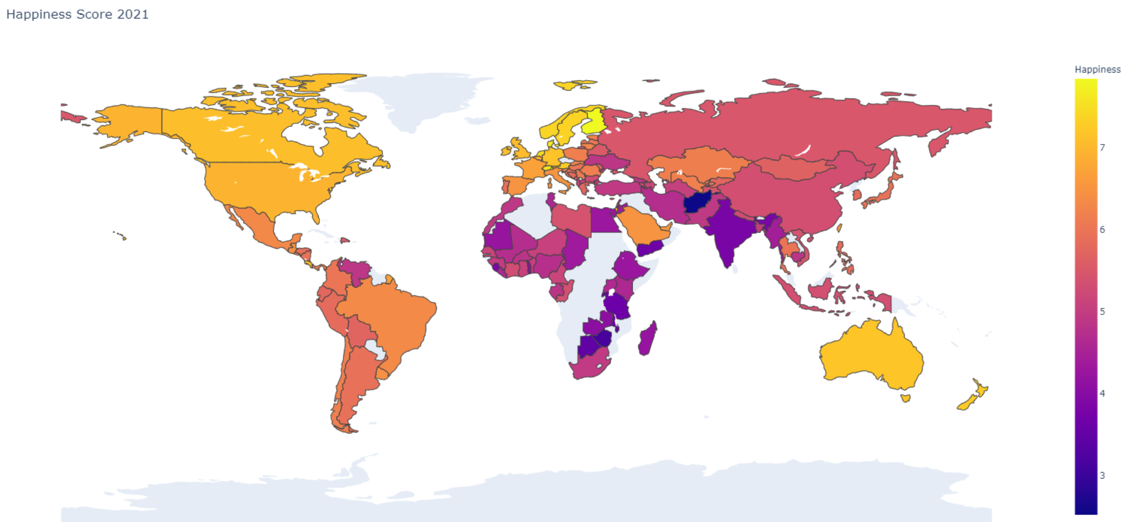
We found this happiness data of 150 countries across the globe for the years 2007 - 2020 (with some missing values) on Kaggle. In this report, we've performed 3 main tasks - EDA and clustering (Section 3), fit various models and predict happiness (Section 4), region based modelling (Section 5).

In section 3, we consider some preliminary analysis of the dataset which helps us to understand the relation between these factors. Firstly, we analyzed correlation between happiness and different factors to find the factors which are strongly related to the happiness and the change of correlation values from year 2007 to year 2021. Also, we compared the data of year 2021 with the data before that, which gives us better understanding of the affect of COVID19 to our happiness. Secondly, we tried to find some patterns behind the happiness score using clustering techniques like kmeans. Specifically, we want to explore if there are some different patterns for countries to have high or low happiness score. Unfortunately, the result is disappointing.

In section 4 we consider the task of predicting happiness score based on the various features of a country for a particular year. This task would prove to be very challenging because of reasons ranging from low data availability (section 4.1) to a requirement of non-trivial cross-validation method (section 4.2). We try to fit many machine learning models (section 4.5) but because of the challenges, most models do not perform well. We do an in-depth analysis of the best model and explain the feature dependence (section 4.6). We also do an anlysis of wrong predictions by the best fit model (section 4.7).

In section 5, we group countries belonging to a particular region and apply the fit and predict methods (described in section 4) to all countries in this region. Then, we repeat the same by grouping all geographically close regions in a particular continent and then try modelling all countries that belong to this continent. We then discuss the success and failure of such region and continent based modelling approaches and how it is different from those described in Section 4.

Figure 1: The worldwide happiness score of year 2021



3 Preliminary analysis

This section we consider some preliminary analysis of the dataset.

3.1 Correlation

In statistics, correlation or dependence is well-known statistical relationship between two random variables. The value of a correlation coefficient ranges between -1 and +1. The correlation coefficient is +1 in the case of a perfect direct (increasing) linear relationship (correlation), -1 in the case of a perfect inverse (decreasing) linear relationship (anti-correlation),[5] and some value in the open interval $(-1, 1)$ in all other cases, indicating the degree of linear dependence between the variables. As it approaches zero there is less of a relationship (closer to uncorrelated). The closer the coefficient is to either 1 or 1, the stronger the correlation between the variables.

We have six feathers like GDP in our data. Intuitively, all of these factors contribute to the happiness score. Here we analyze correlation between happiness and different factors to find the factors which are strongly related to the happiness. Also, the contribution from different factors may vary in different year. Especially, since a worldwide pandemic caused of COVID19 started at year 2020, we consider the year before 2020 as normal data. It is interesting to compare the data of year 2021 with normal data, which gives us better understanding of the affect of COVID19 to our happiness.

3.2 data processing

In this section, we consider the data of different from different year as different data point and we extracted only columns from 'Ladder score', 'Logged GDP per capita', 'Social support', 'Healthy life expectancy','Freedom to make life choices', 'Generosity','Perceptions of corruption'. After dropped data with missing values, there are 2149 data remained.

3.2.1 overall correlation

The correlation is shown in figure 3. We can see that all of the factors except for generosity are strongly related to happiness score. Among them, GDP, Healthy life expectancy, Freedom to make life choices and social support are positively related to happiness score, with decreasing impact. Perceptions of corruption is negtively related, which is within our expectation. Also, based on figure 3, there are some correlation between these factors. For example, GDP are highly related to Healthy life expectancy and social support.

3.2.2 comparison between 2020 and 2021

From Figure 4 we can see that the impact of corruption also decreased from -0.51 to -0.42, which is beyond our expectation.

3.2.3 tendency

From Figure 5, we can see the change of the correlation of these factors with ladder score from year 2007 to 2021.

3.3 clustering

In this section, we try to find some patterns behind the happiness score. Specifically, we want to explore if there are some different patterns for countries to have high or low happiness score. clustering is the task of grouping a set of objects in such a way that objects in the same clusetr are more similar to each other than to those in other clusters which matches our goal. Thus, we applied some clustering techniques to the dataset.

3.3.1 data processing

we consider the whole dataset and extract only the cloumns from 'Logged GDP per capita', 'Social support', 'Healthy life expectancy','Freedom to make life choices', 'Generosity','Perceptions of corruption'. That is, data from the same country but different years are considered as different data. Note that the 'Ladder Score' is not included during clustering, but it's a reference when we compare different clusters.

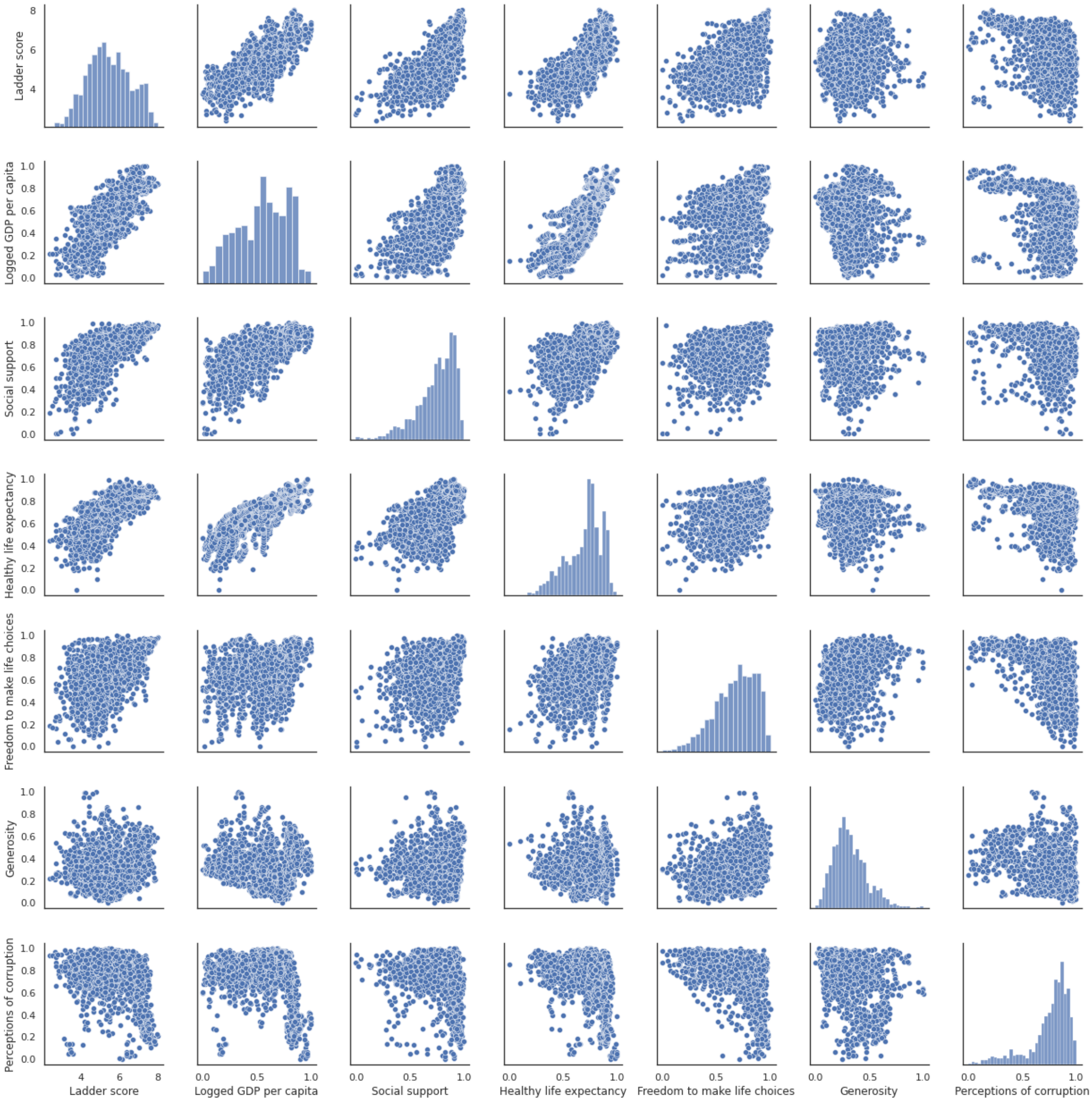
Since clustering is distance based but the data from different columns have different scales, We applied Minmaxscaler() to each column of our data which projects each column to [0,1].

3.3.2 clustering methods

The clustering methods we considered are summarized below.

Kmeans: K-Means is probably the most well-known clustering algorithm. It aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean which are cluster centers.

Figure 2: the correlation between 'Ladder score', 'Logged GDP per capita', 'Social support', 'Healthy life expectancy','Freedom to make life choices', 'Generosity','Perceptions of corruption'



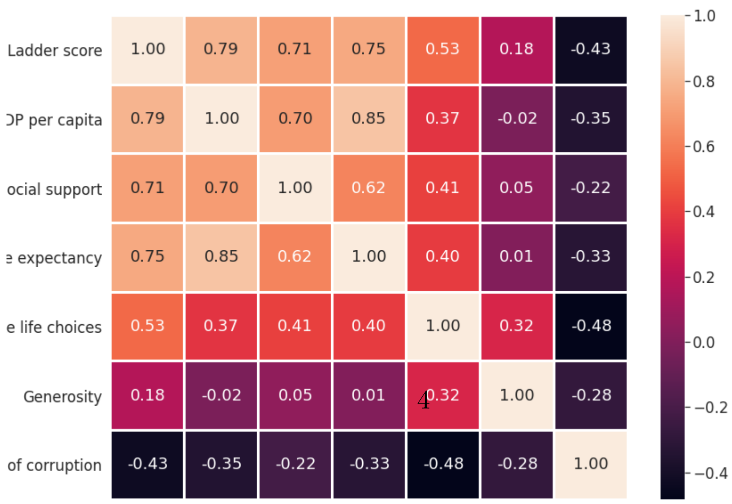
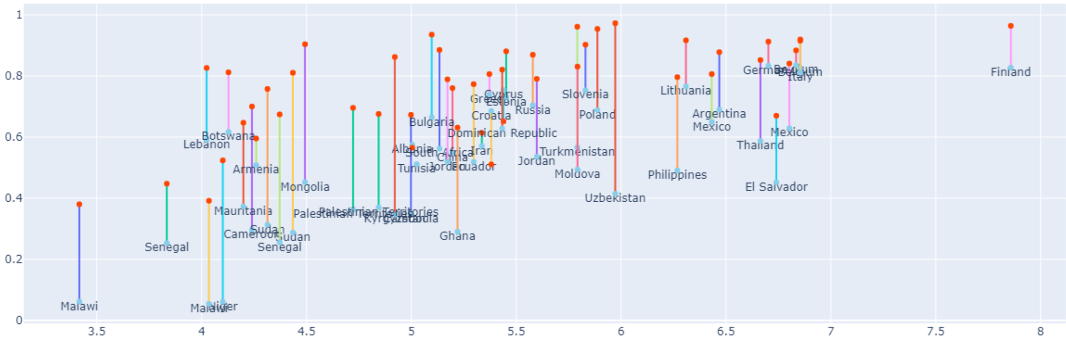
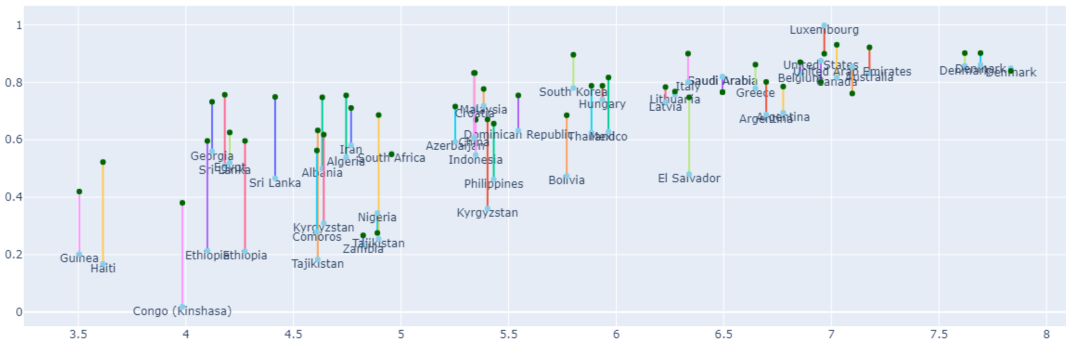


Figure 3: sample data visualization



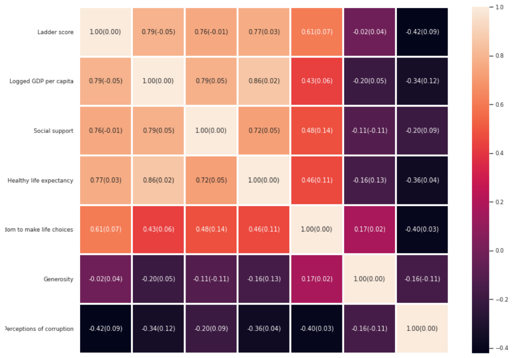
(a) ladder score v.s, GDP(in blue)/social support(in red)



(b) ladder score v.s, GDP(in blue)/healthy life expectancy(in green)

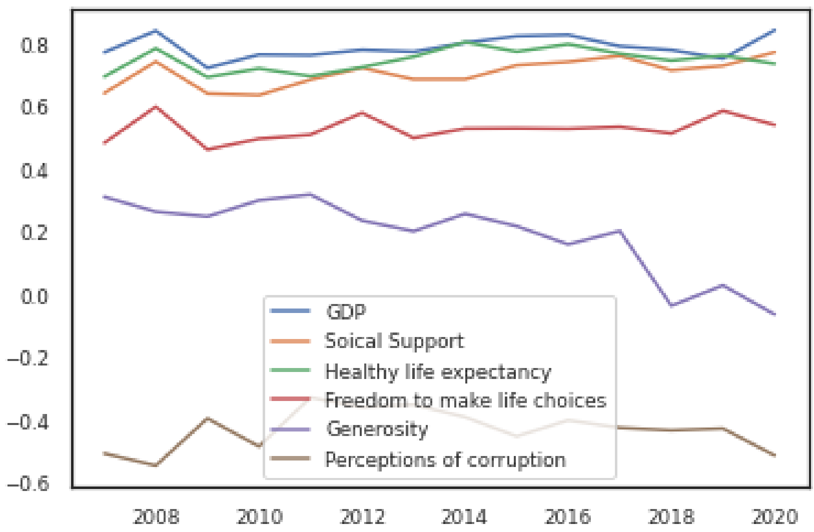
Figure 4: The correlation comparison

(a) year 2020



(b) year 2021

Figure 5: The change of correlation with ladder score



Agglomerative Clustering: This is a "bottom-up" clustering: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.

Affinity Propagation: In statistics and data mining, affinity propagation (AP) is a clustering algorithm based on the concept of "message passing" between data points.[1] Unlike clustering algorithms such as k-means or k-medoids, affinity propagation does not require the number of clusters to be determined or estimated before running the algorithm.

Meanshift: Mean shift is a non-parametric feature-space analysis technique for locating the maxima of a density function, a so-called mode-seeking algorithm.

3.3.3 clustering result

Among all of the clustering techniques, only kmeans needs to initialize the cluster numbers. Thus, we adjusted the cluster numbers from 2 to 5 during our experiments. In Figure 6, only the results for 2 clusters and 3 clusters are given, but the others are similar. From the figure, we can see that the clusters are incremental in terms of happiness score. That is, more advanced clusters, have better GDP, Social support, Healthy life expectancy, Freedom to make life choices and lower corruption. Thus, we conclude that all of these factors are balanced.

From Figure 7, we can see that the result of Agglomerative Clustering are similar to kmeans with 2 clusters. The results of Affinity Propagation and meanshift are messy and inclusive.

4 Happiness Prediction

We consider the following regression problem - Given features of a country for a particular year, can we predict the country's happiness score? The solution to this regression model will also help us understand which features contribute the most (either positively or negatively) in determining the happiness score.

4.1 Challenges in this prediction problem

Some particular aspects of the given dataset, make it very challenging for us to conquer the above mentioned regression problem very efficiently. These challenges are discussed below.

- **Insufficient data** - As we will see in the discussion of our training models, linear models do not perfrom very well for this problem. Non-linear models that usually require a good amount of data don't have a very good perfromance on this dataset as the size of the data is considerably small with 2085 datapoints.

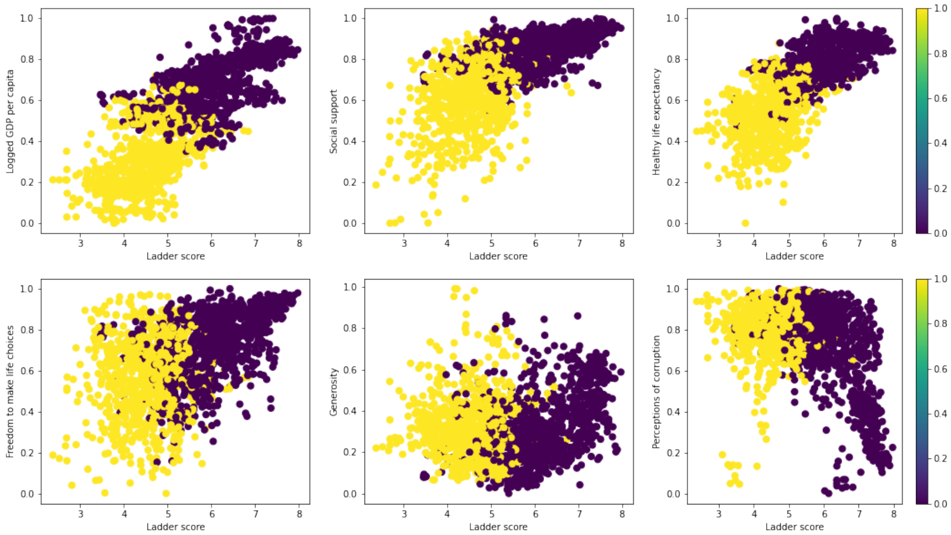
- **Low variation in values for a particular country** - The features that we have for a particular country do not usually change a lot across years. Thus we will have even less datapoints with significant variance in the dataset. This problem has another consequence over the way we have to design our corss-validaiton which is discussed later.

- **Missing Values** - The data contains multiple missing values. We have discussed in a different section how we have dealt with this problem.

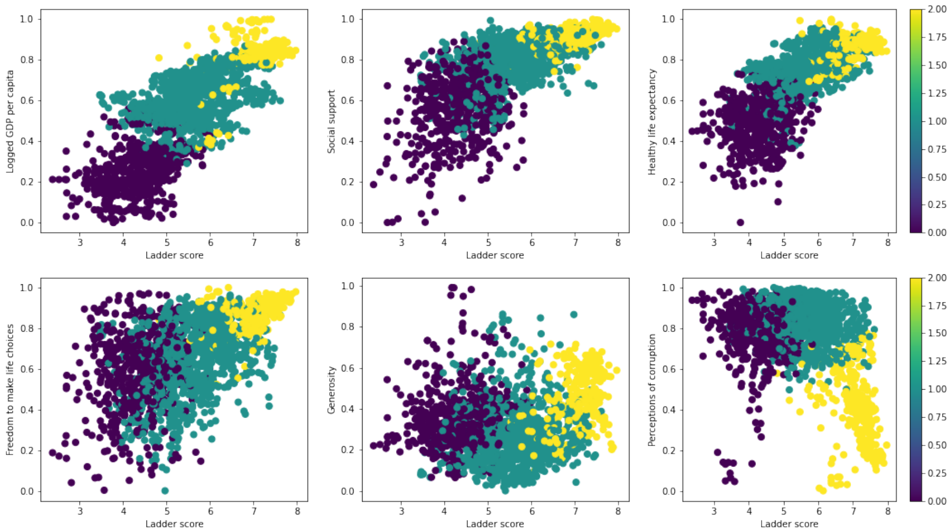
4.2 Cross Validation

As mentioned earlier, the values that we see for a particular country do not have a lot of variation across years. As a consequence, a naive model predicting the mean of happiness score of a particular country will have a very good performance. In our experiments we found that such a model will have an R^2 score of 0.88 over the training dataset. This concludes that even if we have one datapoint for a particular country in the training dataset, and we have a model that is good enough, the model will have good performance on the validation/test data point of the same country.

Figure 6: kmeans clustering of the dataset

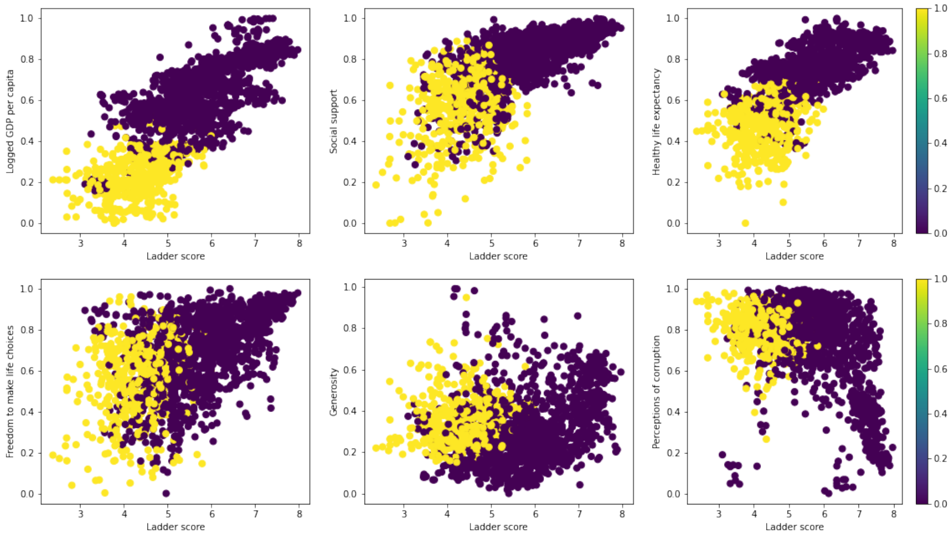


(a) 2 clusters

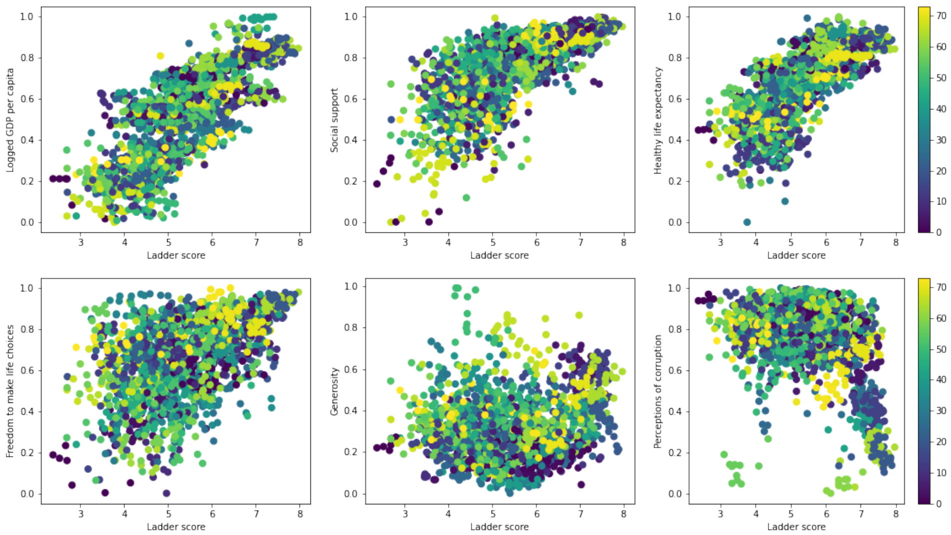


(b) 3 clusters

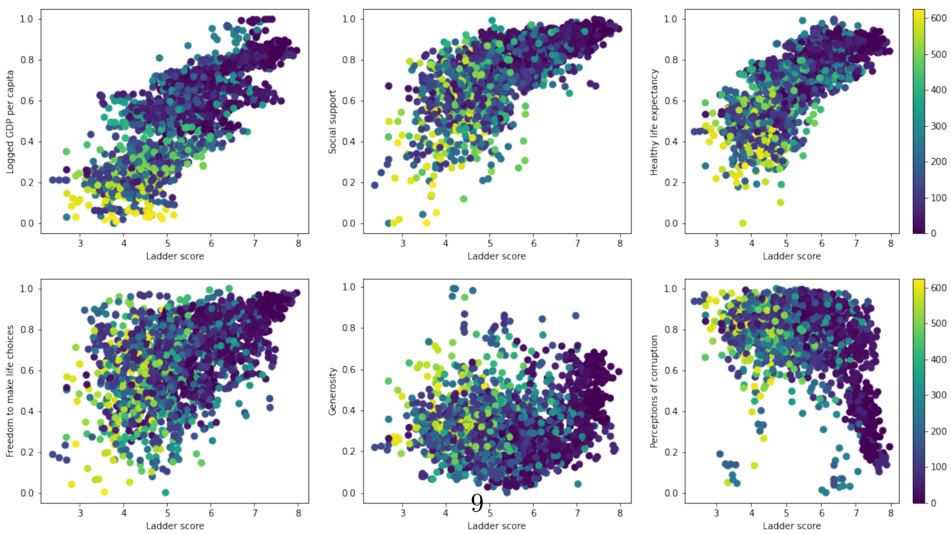
Figure 7: Clustering results except for kmeans



(a) Agglomerative Clustering



(b) Affinity Propagation



(c) Meanshift

We have thus used a cross-validation technique where the datapoints in the training set and the validation set don't have any country in common. We do this by creating 20 data splits (data patition into training and validation dataset) such that in each split we choose any 5 countries and place all of their datapoints into validation data. The remaining goes into the training data. This means we have a $20-$fold cross-validation.

4.3 Evaluation metric - R^2 score

We have used R^2 score as our evaluation metric. For y_i true and $\hat{y}_i$ predicted response variables with $\bar{y}$ denoting the true mean and $i \in [n]$, the R^2 score is given as,

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

The R^2 score is the most standard metric to be used in a regression problem. The score denotes the amount of variance in the true value that is being explained by the predicted values. The best score we can hope for is 1. A score of 0 can be achieve by a model that just predicts the mean of the entire target variable as the predicted value for each data point. As we will see in our results section that we have got negative R^2 score for some models as well. This is because a a model can be arbitrarily worse at explaining the variance.

4.4 Missing Values

As mentioned previously, the data contains multiple missing values. Also mentioned that the values for a particular country do not have a lot of variation across different years. We thus impute the missing values on a country level. We handle the missing value based on whether the feature is categorical or numeric.

- **Categorical features** - We only have one categorical variable and that is the Region indicator. For some years we don't have the regional indicator variable. Since region of a country is supposed to be constant we have just imputed the missing values across years by using an already present value in the dataset for that country. If region of that country is not available for any year, we use the region value as "Unknown".
- **Numeric features** - For numeric features, we fill a missing value for a particular country by using the mean of available values of that feature across years for that particular country.

Any datapoint with missing values present after these steps were dropped from the dataset.

4.5 Models used and results

We tried the following models - Ridge Regression, Random Forest Regressor, Gradient Boosting Regressor, K-Neighbours Regressor, Support Vector Regressor, and Multi-Layer Perceptron with multiple parameters. We observed that tree based methods have performed better than others. A complete list of all the models trained and score observed are available in the appendix. Some details about the best model are mentioned below:

Model Type	Gradient Boosting Regressor
Max Depth	2
No. of estimators	100
Training R^2 score	0.84
Validation R^2 score	0.64
Test R^2 score	0.89

4.6 Understanding the Random Forest model

Decision trees with multiple features are hard to analyze because of the complexity involved across various dimensions. One standard way to analyze them is with the help of Partial Dependence plots. Let X_S be the set of input features of interest (i.e. the features parameter) and let X_C be its complement. The partial dependence of the response f at a point x_S is defined as: Partial Dependence is defined by the sklearn library [1] as,

$$pd_{X_S}(x_S) = \mathbb{E}_{X_C}\left[f(x_S, X_C)\right] = \int f(x_S, x_C)p(x_C)dx_C$$

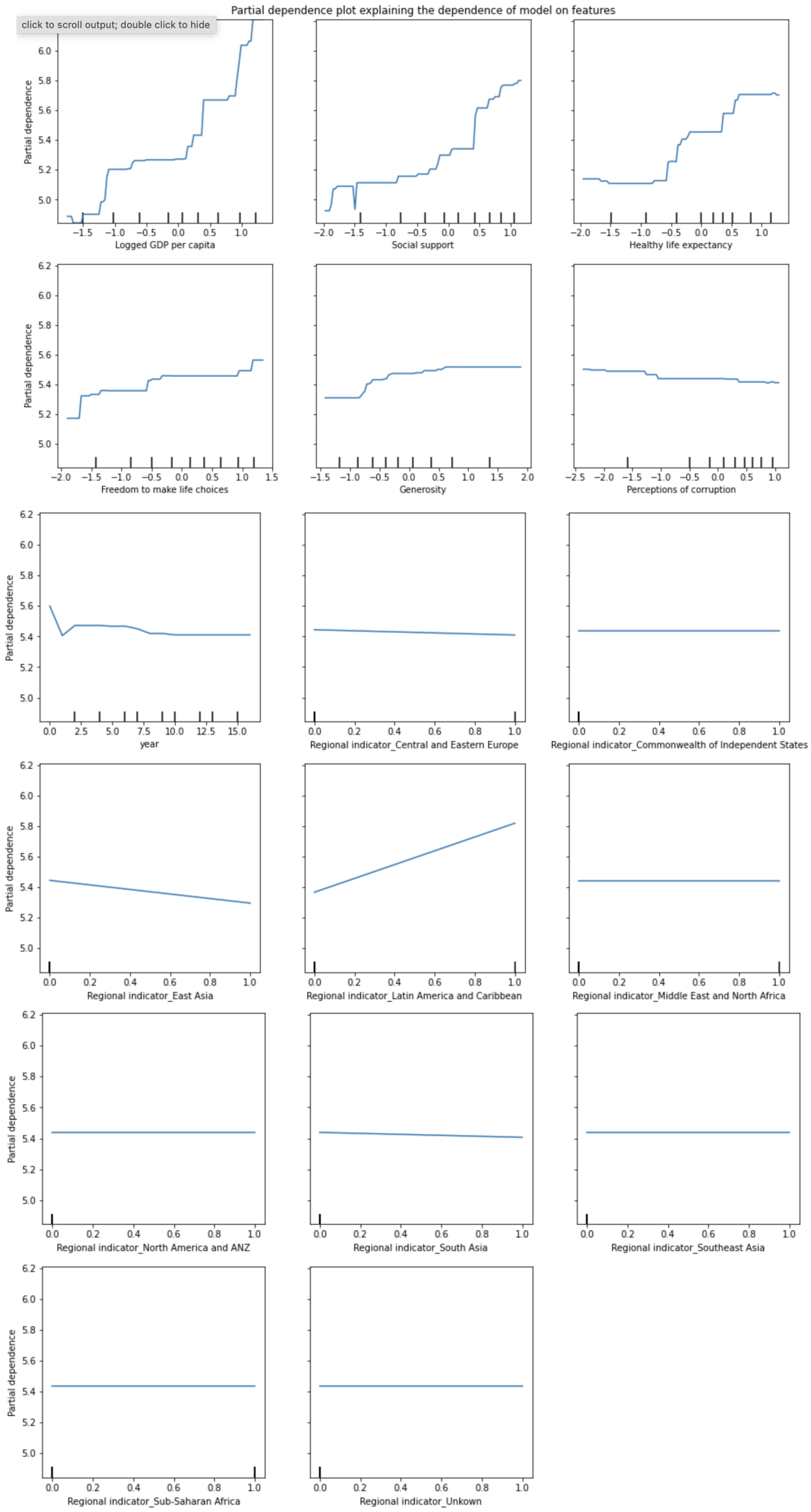
which is approximated as,

$$pd_{X_S}(x_S) \approx \frac{1}{n_{\text{samples}}}\sum_{i=1}^{n} f(x_S, x_C^{(i)}),$$

As noticeable in the partial dependence plots below, we have the similar conclusion as we had during the analysis. Below are some of the observations based on these plots-

- GDP, social support and healthy life expectancy have a very high positive correlation with the happiness score. Moreover, there are specific thresholds in GDP, Social Support and Healthy life expectancy after which there is significant increase in the happiness score for majority of countries.
- Freedom to make life choices and generosity also have positive correlation with the happiness score.
- Perception of corruption has slightly negative effect on the happiness score.
- Happiness score seems to be decreasing on average as years pass by.
- Countries in the Latin America and Caribbean region have significantly higher happiness score than countries with similar features in different regions.
- Countries in the East Asia region have significantly lower happiness score than countries with similar features in different regions.

Figure 8: Partial Dependence of Random Forest on various features



4.7 Explain local instances - LIME

LIME is a visualization technique proposed by [2] to address the issue of machine learning models being black boxes. In Marco Tulio et al's paper, we are given techniques which can be used to visualize any model and provide measurements of how trustworthy a model is. The goal of LIME is to provide some trust in the decisions being made by the model, and in this specific experiment to learn what the model finds important in predicting a particular happiness score. LIME is a local method, it is not meant to be evaluated over the entire input space but instead used on single datapoint.

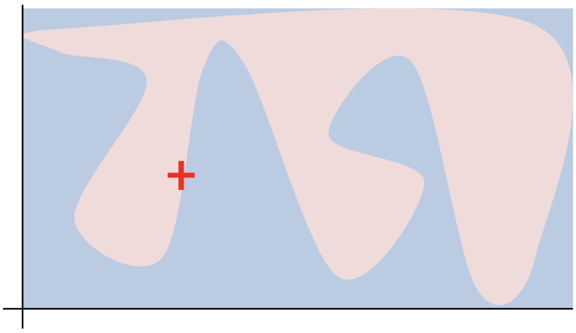
For some instance x, LIME samples around x and weights according to the distance between the two points. It produces explanation ξ according to

$$\xi(x) = \underset{g \in G}{\operatorname{argmin}}\, \mathcal{L}(f, g, \pi_x) + \Omega(g)$$

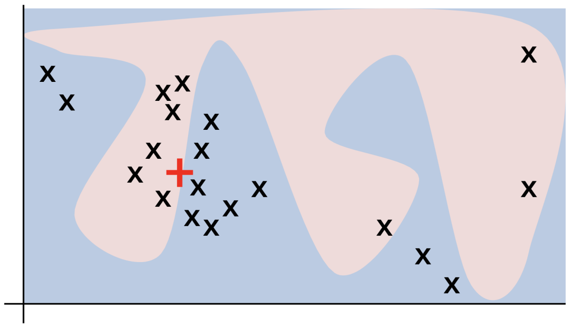
Where G is the class of potentially interpretable models, Ω is the complexity of an explanation g, $\pi_x(z)$ is a proximity measurement between x and z, and $\mathcal{L}$ is a measurement of how faithful g is as an approximation for f. Although Ω is not directly a measure of interpretability, complexity will suffice. The choice of Ω depends on the problem; in a decision tree setting it may be the depth of the tree. When minimizing $\mathcal{L}$ it is important that Ω is low enough for the model g to be interpretable by humans, otherwise the output of LIME will not be meaningful. We will use a linear model wich is simple enough to understand the complex model prediction.

We use LIME to explain a particular prediction instance. In our case we saw that even though the best model is able to perform sufficiently well on the train dataset, there are few instances where the performance is significantly poor. In particular, we notice that the model does consistently bad for the country Sri Lanka. We thus take an instance of Sri Lanka and try to explain it using our LIME implementation.

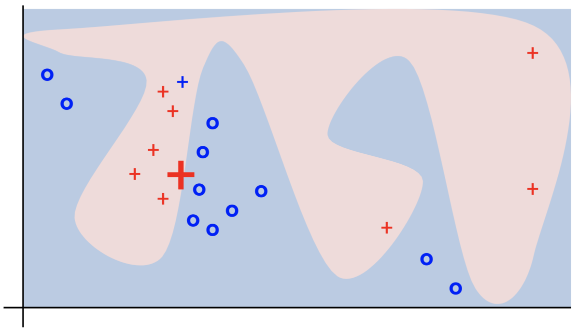
Figure 9: Explain LIME workflow



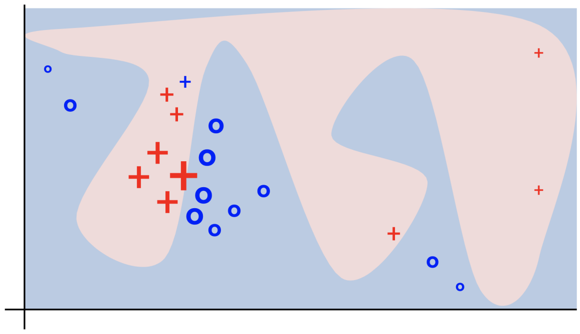
(a) Explaining the prediction for a given point (shown in red) using LIME



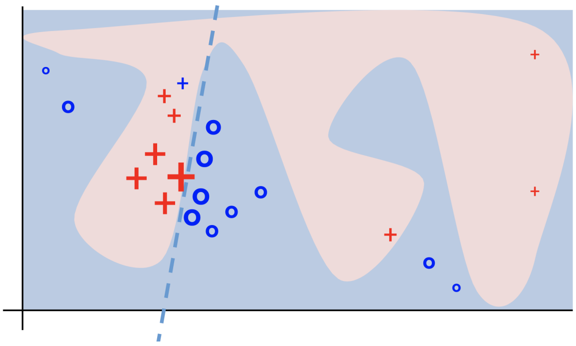
(b) Sample points with higher probability to points near the given point



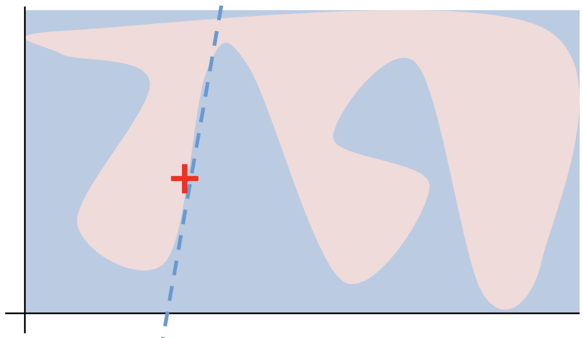
(c) Use black-box model to predict target for the sampled points



(d) Weight points based on their distance from the given point

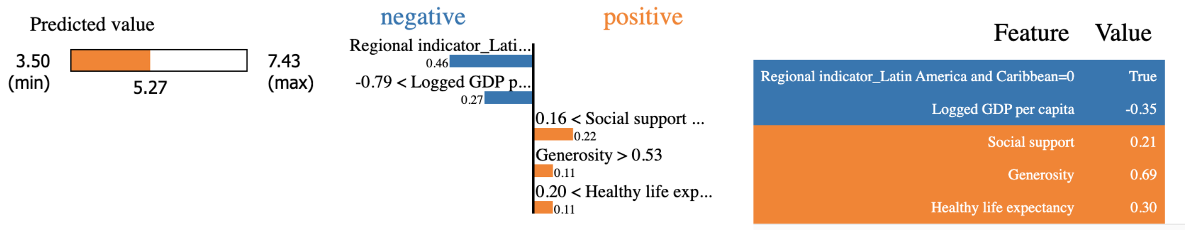


(e) Learn a simple model (usually linear) to explain this prediction

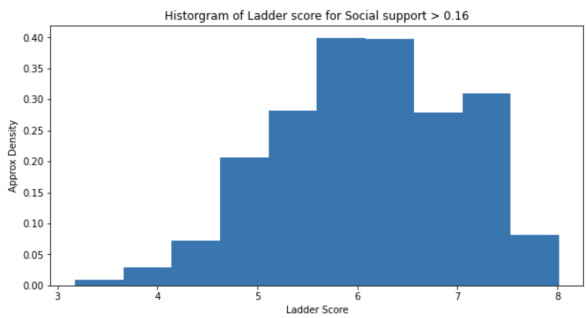


(f) Use this simple model as a local approximation to understand the black-box model

Figure 10: Explain Wrong predictions for Sri Lanka

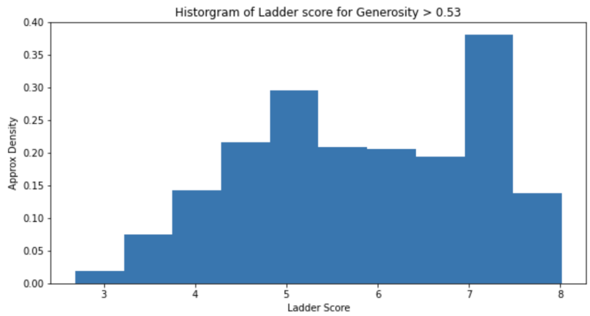


(a) LIME explanation for Sri Lanka



(b) Historgram of Ladder score for Social support > 0.16

(c) Historgram of Ladder score for Healthy life expectancy > 0.20



(d) Historgram of Ladder score for Generosity > 0.53

Note that our analysis concludes that high value of Healthy Life expectancy, Generosity and Social support are the reasons for a predicted high score of happiness for Sri Lanka. But as can be seen from the Figure 10, based on the features of Sri Lanka the 4.415 happiness score of the country is an outlier.

5 Region Based modelling

In this section, we apply the same suite of models to countries belonging to a specific region. We test whether such a region based modelling reveals relationships between variables that might not be captured in our previous modelling approaches. Having already dealt with the missing data earlier, we group all countries belonging to a specific region together. We've 10 different regions namely - South Asia, Central and Eastern Europe, Middle East and North Africa, Latin America and Caribbean, Commonwealth of Independent States, North America and ANZ, Western Europe, Sub-Saharan Africa, Southeast Asia and East Asia. The total data in each region ranges between 100 - 400 points, which further exacerbates our attempt to model features with good statistical accuracy. Nevetheless, we apply the suite of models developed earlier.

5.1 Results from Region Based modelling

After compiling countries in each region, we tried the following models for all countries belonging to a specific region - Ridge Regression, Random Forest Regressor, Gradient Boosting Regressor, K-Neighbours Regressor, Support Vector Regressor, and Multi-Layer Perceptron with multiple parameters. Unfortunately, the modelling wasn't very accurate for all the regions. For example, for the region of South Asia, Central and Eastern Europe and Sub-Saharan Africa, the best fit model details are as given below:

South Asia

Model Type	SVR
estimator kernel	rbf
cross validation R^2 score	-0.93

Central and Eastern Europe

Model Type	KNeighborsRegressor
estimator neighbors	10
cross validation R^2 score	0.33

Sub Saharan Africa

Model Type	Gradient Boosting Regressor
Max Depth	2
No. of estimators	50
cross validation R^2 score	-0.05

Modelling approaches to other regions were similarly unsuccessful. One of the possible reason, why such models were unsuccessful in such a region based modelling was the limited amount of data points we had and the inherent variability in the data. The training data, test data for South Asia (for example) was 74, 17 points respectively. Sub Saharan Africa and Central and Eastern Europe had 379, 47 and 198, 44 points respectively.

In an attempt to still have some modelling successes within regions, we clubbed regions that belong to the same continent together, and apply the suite of models to all countries that belong to a continent. As seen in (), we clubbed the regions (especially in Europe, Africa and Asia) which belong in the same continent. We now apply the same suite of models to each of these 3 continents. We find significant improvement in our model accuracy Europe, and little improvement in Africa. We still didnt have much success in Asia because even after combining all regions in Asia, the total datapoints resulted to about 304, which is still small. In contrast, the total data points in Europe and Africa were 716 and 654. The best fit model for each of these continents are given below:

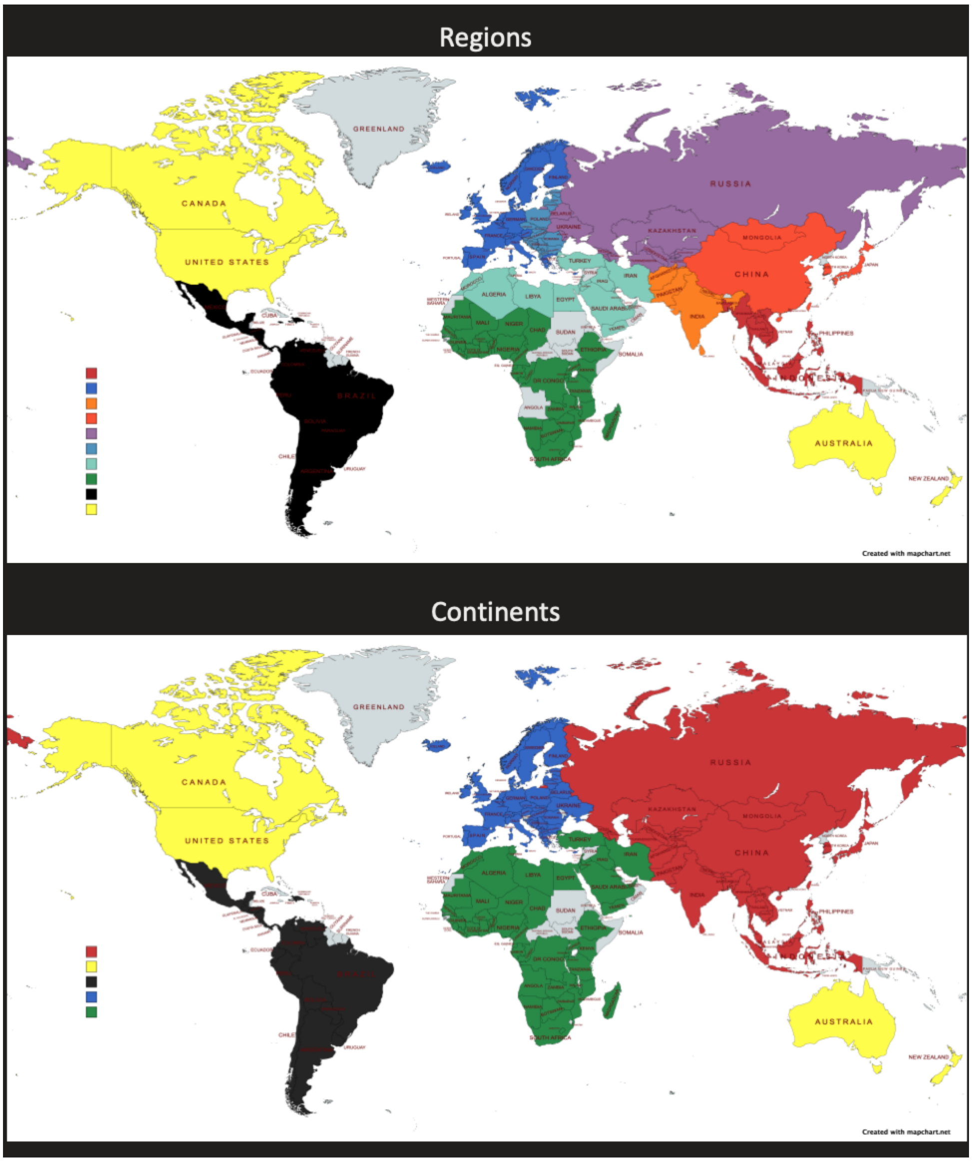


Figure 11: The upper half shows a map where countries belonging to each region are marked in the same color. Different colors represent different regions in which the survey was conducted. In order to show which regions were grouped together for continent based modelling, we've used different shades of same color. For example - regions in Africa defer in their shade of green, regions in Europe differ in their shade of blue and regions in Asia differ in their shade of Red. The lower half shows the map where regions belonging to each continent are marked in the same color. Asia is marked in red, Africa is marked in green and Europe is marked in blue.

Europe

Model Type	Gradient Boosting Regressor
Max Depth	2
No. of estimators	50
Training R^2 score	0.90
Validation R^2 score	0.66
Test R^2 score	0.84

Africa

Model Type	Gradient Boosting Regressor
Max Depth	2
No. of estimators	100
Training R^2 score	0.73
Validation R^2 score	0.25
Test R^2 score	0.59

Asia

Model Type	Ridge
Estimator alpha	10
Training R^2 score	0.60
Validation R^2 score	-0.22
Test R^2 score	0.74

As we see, there were improvements in our modelling approaches when we grouped regions belonging to the same continent together. Essentially, we see that the Gradient Boosting Regressor gives a model which explains the countries belonging to Europe well. Now that we have a good model for Europe, we plot partial dependency of happiness on all estimators for Europe and compare it with the partial dependency plots of all data (shown earlier).

As noticeable in the partial dependence plots (), there are some similarities and some new trends which were not captured in the model for all data. Happiness score still remains strongly correlated with Logged GDP per capita and social support. However, health life expectancy isn't important in Europe (as shown by the flat line in the partial dependency plot of Europe), as compared to all data, where there's a strong positive correlation. This is a new feature, which wasn't captured by the modelling approaches earlier.

Note: In the original data, US, Canada, Australia and New Zealand were marked in a single region (named North America and ANZ) which had a total of 62 data points and thus, we ignored those data points while doing the continent level analysis (see Fig 10).

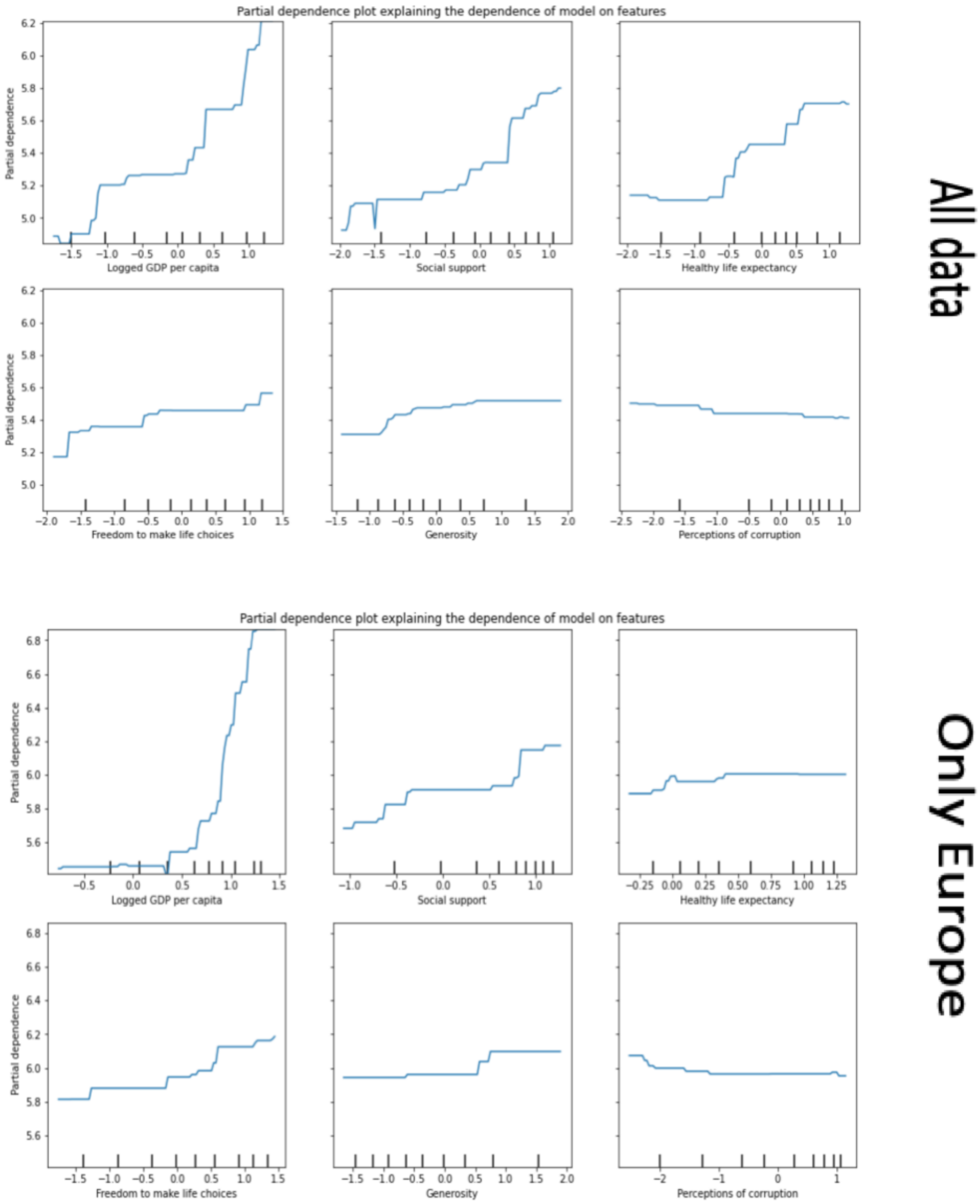


Figure 12: The upper half shows the partial dependency of happiness score on each of the 6 estimators for all data (similar to Figure 7), where as the lower half shows the partial dependency of happiness score on each of the partial estimators only for countries belonging to Europe. We see no dependence of happiness score on Health Life expectancy for Europe, which is a new feature.

6 Conclusion

Predicting happiness score from the features of a country proved to be a very challenging task because of many reasons mentioned in the report in section 4.1. As expected, due to non-linear dependence on features and low data availability, tree-based methods perform the best for the prediction task. The best fit method reveals a similar feature dependence for happiness score as discussed during data analysis. Some countries such as Sri Lanka are certainly an outlier based on their features and make an already challenging prediction task even more difficult.

In addition, with scarce data, region based modelling wasn't very successful. However, with grouping geographically close regions together (by grouping countries into continents), we were able to find a model that fits the data with varied success. Our best model was for Europe (with an R^2 score of 0.84 on test data) and worst was Asia owing to less data from that continent. In addition, for the countries in Europe, we saw little dependence of Happiness score on health life expectancy, which was a new feature. Thus, we believe that region based modelling in principle would reveal novel features that might not be captured if we look at data from all countries. But to successfully reveal these features, we need large amounts of data to train these models. With limited data, it's a challenging task to model or predict happiness score.

References

[1] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[2] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. ""Why Should I Trust You?": Explaining the Predictions of Any Classifier". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*. 2016, pp. 1135–1144.

Appendix

#Index	Rank	Mean Test Score	Parameters	Mean Fit Time
0	12	0.602195	{'estimator': Ridge(), 'estimator__alpha': 0}	0.008005
1	10	0.602311	{'estimator': Ridge(), 'estimator__alpha': 0.1}	0.006249
2	8	0.603309	{'estimator': Ridge(), 'estimator__alpha': 1}	0.004314
3	6	0.609959	{'estimator': Ridge(), 'estimator__alpha': 10}	0.004423
4	22	0.463208	{'estimator': RandomForestRegressor(), 'estimator__max_depth': 2, 'estimator__n_estimators': 50}	0.161464
5	23	0.461095	{'estimator': RandomForestRegressor(), 'estimator__max_depth': 2, 'estimator__n_estimators': 100}	0.33537
6	24	0.46002	{'estimator': RandomForestRegressor(), 'estimator__max_depth': 2, 'estimator__n_estimators': 200}	0.696887
7	4	0.613725	{'estimator': RandomForestRegressor(), 'estimator__max_depth': 5, 'estimator__n_estimators': 50}	0.379015
8	7	0.606943	{'estimator': RandomForestRegressor(), 'estimator__max_depth': 5, 'estimator__n_estimators': 100}	0.74692
9	5	0.611372	{'estimator': RandomForestRegressor(), 'estimator__max_depth': 5, 'estimator__n_estimators': 200}	1.545581
10	9	0.602993	{'estimator': RandomForestRegressor(), 'estimator__max_depth': None, 'estimator__n_estimators': 50}	1.951695
11	14	0.599279	{'estimator': RandomForestRegressor(), 'estimator__max_depth': None, 'estimator__n_estimators': 100}	4.02855
12	13	0.601489	{'estimator': RandomForestRegressor(), 'estimator__max_depth': None, 'estimator__n_estimators': 200}	10.927079
13	3	0.620445	{'estimator': GradientBoostingRegressor(max_depth=2), 'estimator__max_depth': 2, 'estimator__n_estimators': 50}	0.240358
14	1	0.635403	{'estimator': GradientBoostingRegressor(max_depth=2), 'estimator__max_depth': 2, 'estimator__n_estimators': 100}	0.483516
15	2	0.634718	{'estimator': GradientBoostingRegressor(max_depth=2), 'estimator__max_depth': 2, 'estimator__n_estimators': 200}	1.043488
16	11	0.60229	{'estimator': GradientBoostingRegressor(max_depth=2), 'estimator__max_depth': 5, 'estimator__n_estimators': 50}	0.756305
17	15	0.594654	{'estimator': GradientBoostingRegressor(max_depth=2), 'estimator__max_depth': 5, 'estimator__n_estimators': 100}	1.574213
18	17	0.58168	{'estimator': GradientBoostingRegressor(max_depth=2), 'estimator__max_depth': 5, 'estimator__n_estimators': 200}	3.262967
19	25	0.389642	{'estimator': GradientBoostingRegressor(max_depth=2), 'estimator__max_depth': None, 'estimator__n_estimators': 50}	6.110217
20	28	0.382788	{'estimator': GradientBoostingRegressor(max_depth=2), 'estimator__max_depth': None, 'estimator__n_estimators': 100}	15.32109
21	27	0.383536	{'estimator': GradientBoostingRegressor(max_depth=2), 'estimator__max_depth': None, 'estimator__n_estimators': 200}	34.972327
22	26	0.386868	{'estimator': KNeighborsRegressor(), 'estimator__n_neighbors': 2}	0.004556
23	19	0.505012	{'estimator': KNeighborsRegressor(), 'estimator__n_neighbors': 5}	0.004588
24	18	0.530203	{'estimator': KNeighborsRegressor(), 'estimator__n_neighbors': 10}	0.004533
25	16	0.592218	{'estimator': SVR(), 'estimator__gamma': 'auto', 'estimator__kernel': 'rbf'}	0.771744
26	21	0.468643	{'estimator': SVR(), 'estimator__gamma': 'auto', 'estimator__kernel': 'poly'}	15.32265
27	34	NaN	{'estimator': SVR(), 'estimator__gamma': '0.01', 'estimator__kernel': 'rbf'}	0.004551
28	35	NaN	{'estimator': SVR(), 'estimator__gamma': '0.01', 'estimator__kernel': 'poly'}	0.004201
29	33	-3.210518	{'estimator': MLPRegressor(max_iter=100), 'estimator__activation': 'logistic', 'estimator__hidden_layer_sizes': (10, 5)}	3.753728
30	32	-2.260032	{'estimator': MLPRegressor(max_iter=100), 'estimator__activation': 'logistic', 'estimator__hidden_layer_sizes': 5}	3.281158
31	20	0.497604	{'estimator': MLPRegressor(max_iter=100), 'estimator__activation': 'relu', 'estimator__hidden_layer_sizes': (10, 5)}	4.058621
32	31	-0.755117	{'estimator': MLPRegressor(max_iter=100), 'estimator__activation': 'relu', 'estimator__hidden_layer_sizes': 5}	3.080073
33	29	-0.289043	{'estimator': MLPRegressor(max_iter=100), 'estimator__activation': 'tanh', 'estimator__hidden_layer_sizes': (10, 5)}	4.174816
34	30	-0.343321	{'estimator': MLPRegressor(max_iter=100), 'estimator__activation': 'tanh', 'estimator__hidden_layer_sizes': 5}	3.21095

Figure 13: Model Fitting Results