A Real Data-Driven Analytical Model to Predict Happiness

Aditya Chakraborty (adityachakra@usf.edu)
University of South Florida
Chris P. Tsokos
University of South Florida

Research Article

Keywords: Gallup world poll, Subjective well being (SWB), Nonlinear statistical modelling, Machine learning regularization techniques

Posted Date: February 9th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-224834/v1

A Real Data Driven Analytical Model to Predict Happiness

Aditya Chakraborty (Doctoral Candidate)
adityachakra@usf.edu
Chris P. Tsokos (Distinguished University Professor)
ctsokos@usf.edu
Department of Mathematics & Statistics
University of South Florida
Tampa,FL-33620

February 9, 2021

Abstract

Purpose: Philosophers and many modern-day researchers are convinced by the fact that the pursuit of happiness is the ultimate goal for humankind. Aristotle believed that the utmost goal of human life was *eudaimonia* (interpreted as **"happiness," "human flourishing," or "a good life."**). Recently, many economists and physiologists have been doing applied research in the areas of *subjective well-being (SWB)* or happiness and trying to understand how it improves the quality of life of individual beings. Thus, searching for a data-driven analytical model is crucial to predict SWB and enhance the quality of life

Methods: Our present study utilizes the world happiness database obtained from the Gallup World Poll on the happiness of 156 countries. However, our study focuses on using only the data of fifty-four developed countries, based on the human development index **(HDI)**. We have developed a non-linear analytical model that predicts the average happiness score based on eleven risk factors with a high degree of accuracy. We also compared our analytical model with three other statistical models, and our model outperformed the rest of the three in terms of $RMSE$ and MAE.

Results: Our analytical model includes **five** important findings. The response of the proposed model is the average score of happiness of individuals in developed countries. In addition to predicting the happiness score, our model identifies the individual risk factors and their corresponding interactions that significantly contribute to happiness. We rank these risk factors by their percentage of contributions to the happiness score. We also proceed to rank the developed countries with respect to their predicted happiness score from our developed model. From our study, we found **Finland** being number one, followed by Denmark. The U.S is **fifth** and Romania being **54th**.

Conclusion: The proposed model offers other useful information on the subject area. Our analytical model has been validated and tested to be of high quality, and our prediction of happiness is with a high degree of accuracy. We created a survey questionnaire (appendix 1) based on the data that can be used along with our model by any company for the *strategic planning* or *decision making*.

Keywords: Gallup world poll,subjective well being (SWB), nonlinear statistical modelling, machine learning regularization techniques

1 Introduction

When we think about Happiness in modern life, we might be referring to the feeling we get after the first lick of a delectable ice cream cone or when spending quality time with some of our wonderful friends. This way of thinking about Happiness as satisfaction or amusement suggests that it is a subjective, emotional state, susceptible to the moment-to-moment experience that we are having.

Even though feeling good is a part of Happiness, some old lines of thought have defined Happiness more extensively. Specifically, Aristotle believed that the ultimate goal of human life was a notion of ancient Greeks called *eudaimonia*. The word is often translated as *Happiness*, but more likely means "human flourishing" or "a good life." Being happy is not only associated with personal well-being but also with productivity on a large scale. Studies have been performed to understand the association between Happiness and productivity(1). Happy individuals tend to perform better, and lower Happiness is correlated with less productiveness. Several pieces of evidence, accumulating the complementary strengths and weaknesses, have been consistent with the existence of a causal link between human well-being and human performance(1). A happy mind is also associated with sound mental health. Health and Happiness are essential and possibly related to the pursuits of mankind. Sound health may play a vital role in determining the Happiness or, morbidness/sickness may cause unhappiness. Conversely, a feeling of Happiness may strengthen health conditions(2). Numerous studies on Happiness has been done by social researchers concentrating on psychological and social causal and cognitive factors of Happiness. For instance, Happiness is routinely keep under surveillance in sociological surveys(6), and levels of Happiness have been connected to individual personality and idiosyncrasy(3), living conditions(2), dignity and morale (7), love(4), democracy(5), and also with brain activity of specific individual(8). Some studies have investigated Happiness concerning health in a widespread population. In an epidemiological study of Finnish men, it has been found that life satisfaction (measured through four items assessing whether life is interesting, happy, easy, or lonely) predicts lower mortality(9), but the specific contribution of Happiness was not reported. In the medical literature, the Happiness is often considered a contributing factor of good mental health. For instance, the mental health scale embedded in the Short Form-36 (SF-36) questionnaire includes an item on Happiness, (10), one item from the Bradburn scale of well-being asks whether the respondent is 'depressed or very unhappy' (11; 12), and the validity of the Happiness-Depression scale was tested against a mental health questionnaire(13). Taking the contrasting stand, Saracci proclaimed that the World Health Organization (WHO) definition of health should be more relevant for a definition and interpretation of Happiness, and health needs a narrower definition(14). However, it is has not been proven how the notions of health, well-being, quality of life, and Happiness relate to each other and further study is required in this regard. If we can develop a functional relationship between the subject of study (such as happiness) and other contributing factors (which might affect the subject of interest) by justifying and validating the assumptions, we can predict the happiness given the specific levels of the contributing variables. Also, it is possible to maximize the level of happiness and determine the desired levels of individual risk factors from the model. Some studies also found that the effect of the nationality of levels of Happiness may capture the impact of cultural integration on people's well-being(2). Using an international cross-section of 28 countries, researchers have found a highly significant impact of democracy on the subjective well-being of people(19). Thus, Happiness and democracy, as one would expect, are highly correlated.

In general, personal Happiness and well-being seem to the principal objective of human life. Throughout history, the virtue of Happiness has been considered as the ultimate end of temporal existence. Aristotle's ancient view about Happiness was *"Happiness is so important, it transcends all other temporal considerations"*. Aristotle's prescription for spending a good life was to exercise virtues like being kind, humble, wise, and honest in our actions consistently. In other words, accomplishing different physical and emotional needs, as listed in Figure 1, is the recipe for a happy life. The following Figure 1 illustrates briefly different stages of human needs to achieve Subjective Well

Being.

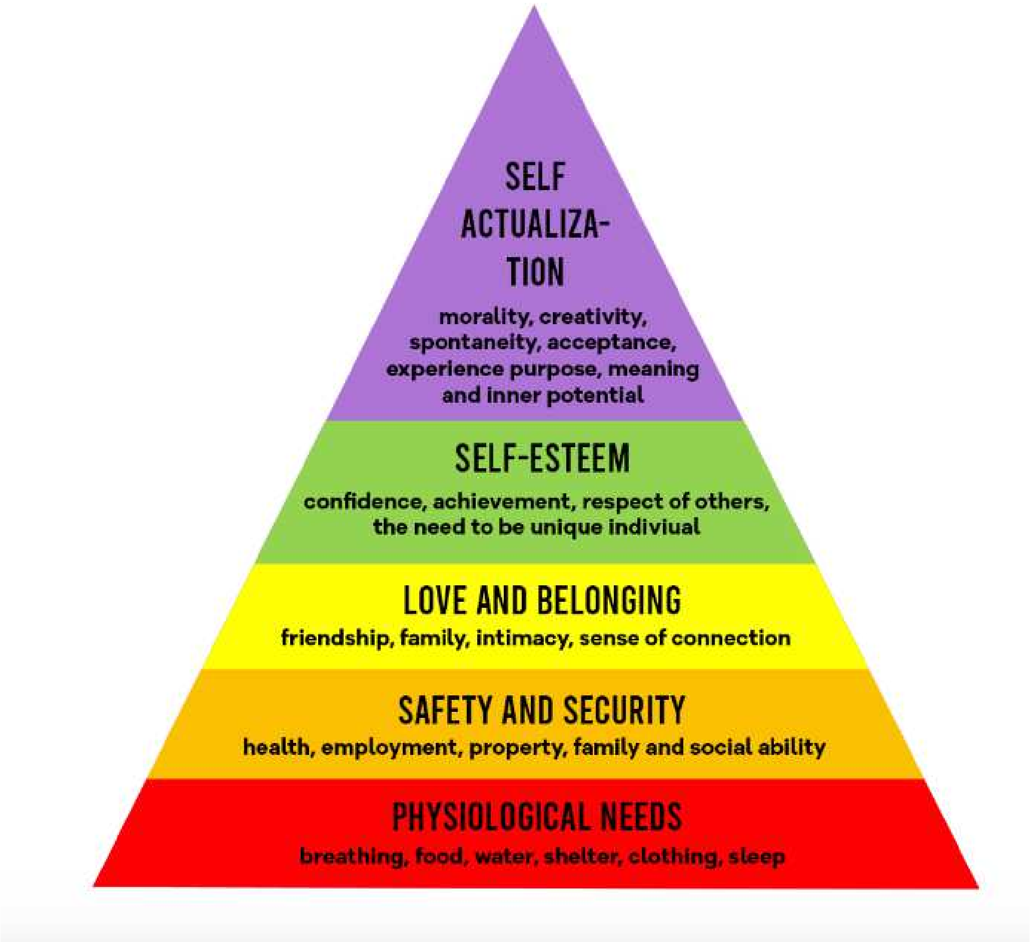

SELF ACTUALIZA- TION
morality, creativity, spontaneity, acceptance, experience purpose, meaning and inner potential
SELF-ESTEEM
confidence, achievement, respect of others, the need to be unique indiviual
LOVE AND BELONGING
friendship, family, intimacy, sense of connection
SAFETY AND SECURITY
health, employment, property, family and social ability
PHYSIOLOGICAL NEEDS
breathing, food, water, shelter, clothing, sleep

Figure 1: Triangle Showing Hierarchy of Needs For Subjective Well Being
source by:`https://irechargeme.com/wellbeing/`

While we build the analytical model, we have the national average of happiness score as the response variable; hence, we proceed to develop an analytical model containing significant risk factors and other significant interactions. The proposed non-linear statistical model is based on several assumptions, such as linearity, multicollinearity, homoscedasticity, and different assumptions concerning statistical methodology. The dataset shows that some of the risk factors are highly correlated, as shown in Figure 4. The parameters of the models become difficult to interpret under the influence of multicollinearity. The parameters also become very unstable when independent variables are highly correlated, which leads to over-fitting the model. Moreover, we use different penalization regression methods: Ridge Regression (L_2)(14), Lasso Regression (L_1)(15), and Elastic net (EN)(16). These machine-learning techniques are vastly used in applied sciences to address several ill-factors of the model (such as over-fitting) . Our proposed statistical model is useful in predicting individuals' Happiness, given the values of the significant risk factors. Also, we ranked the risk factors in accordance with their percentage of contribution to the happiness score. The validation and quality of our proposed analytical model have been statistically evaluated using R square (R^2), R square adjusted $(R^2{}_{adj})$, Mean absolute deviation (MAD), root mean square error (RMSE), and residual analysis. The advantages of using this model has been discussed in the conclusion section. To the best of our knowledge, no such statistical model has been developed under the proposed logical structure to predict Happiness for developing countries. Therefore, searching for an proper data-driven analytical model in the prediction of Happiness is important.

2 Methodology

2.1 The Data

The World Happiness Report is a landmark survey of the state of global happiness that ranks descriptively 156 countries by how happy their citizens perceive themselves to be. The data has been obtained from the World Happiness Report 2019 website(22), where they used the **Gallup Poll** to get the answers to specific questions (*appendix 1*). The data has been collected for a total of 156 countries from 2005 to 2018. However, in our study, we only considered the data of **developed countries**(sorted based on the human development index**[HDI]**) in the world. Individuals were asked specific questions, and as a result of their response as a whole, a score was produced, which is termed as the national average. In our data, the average scores of the developed countries from 2005 to 2018 were tabulated. One of the main goals of our study is to understand what attributable variables significantly affect the happiness of an individual. We have eleven attributable variables and the **Ladder** (which is also called subjective well being [SWB] or happiness score as a measure of response. **For example, let there be an imaginary ladder, with steps numbered from 0 at the bottom to 10 at the top. The top of the ladder represents the best possible life, and the bottom of the ladder represents the worst possible life of an individual. On which step of the ladder is an individual standing currently is reported. This measure is also referred to as the *Cantril life ladder* or just life ladder in our analysis.**

The attributable variables (risk factors) that the data was collected on are given below. The descriptions of the risk factors are the same as provided in the world happiness report 2019.

- **LOG_GDP**(X_1)(Log GDP): Per-capita gross domestic product(in logarithmic scale) in purchasing power parity(PPP).

- **SOC_SUPPORT**(X_2)(Social Support): This variable is defined as is the national average of the binary responses (either 0 or 1) to the GWP question "If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?"

- **LIFE_EXPECT**(X_3)(Life Expectancy): Healthy life expectancies at birth are based on the data extracted from the World Health Organization's (WHO) Global Health Observatory data repository.

- **FREEDOM**(X_4): Freedom to make life choices is the national average of responses to the GWP question "Are you satisfied or dissatisfied with your freedom to choose what you do with your life?"

- **Generosity**(X_5): Generosity is the residual of regressing national average of response to the GWP question "Have you donated money to a charity in the past month?" on GDP per capita.

- **PER_CORR**(X_6)(Perception of Corruption): The measure is the national average of the survey responses to two questions in the GWP: "Is corruption widespread throughout the government or not" and "Is corruption widespread within businesses or not?" The overall perception is just the average of the two 0-or-1 responses.

- **POS_AFFECT**(X_7)(Positive Affect): Positive affect is defined as the average of three positive affect measures in GWP. These are happiness, laughter, and enjoyment in the Gallup World Poll.

- **NEG_AFFECT**(X_8)(Negative Affect): Negative affect is defined as the average of three negative affect measures in GWP. These are worry, sadness, and anger, respectively.

- **CONF_GOV**(X_9)(Confidence in Government): How much trust and confidence does one have in government when it comes to handling [International problems/Domestic problems] – a great deal, a fair amount, not very much or none at all?

- **DEM_QUALITY**(X_{10}): Democratic quality is the National average of the first two dimensions of World Governance Index **(WGI)**(23) namely,*voice and Accountability* and *Political Stability and Absence of Violence/Terrorism*.

- **DEL_QUALITY**(X_{11}):Delivery quality is the National average of the last two dimensions of World Governance Index **(WGI)** namely, *Government Effectiveness*, *Regulatory Quality*, *Rule of Law* and *Control of Corruption*.

The definitions of the above-mentioned measures under **DEM_QUALITY** and **DEL_QUALITY** (which are also the six dimensions of the World Governance Quality Index **(WGI)** are as follows:

1. Voice and Accountability: Voice and accountability captures perceptions of the extent to which a country's citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association, and a free media.

2. Political Stability and Absence of Violence/Terrorism: Political Stability and Absence of Violence/Terrorism measures perceptions of the likelihood of political instability and/or politically motivated violence, including terrorism.

3. Government Effectiveness: Government effectiveness captures perceptions of the quality of public services, the quality of the civil service as the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies.

4. Regulatory Quality: Regulatory quality captures perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development.

5. Rule of Law : Rule of law captures perceptions of the extent to which government agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence.

6. Control of Corruption : Control of corruption captures perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests.

From Figure 2 below, we see that there are some missing observations in the data set. However, the proportion of missing values is small; we used **predictive mean matching (pmm)** algorithm to perform multiple imputation(20) to our dataset. Predictive mean matching (PMM) is a useful technique to perform multiple imputation(21) for missing data points in a plausible manner, especially for imputing quantitative variables that are not normally distributed.

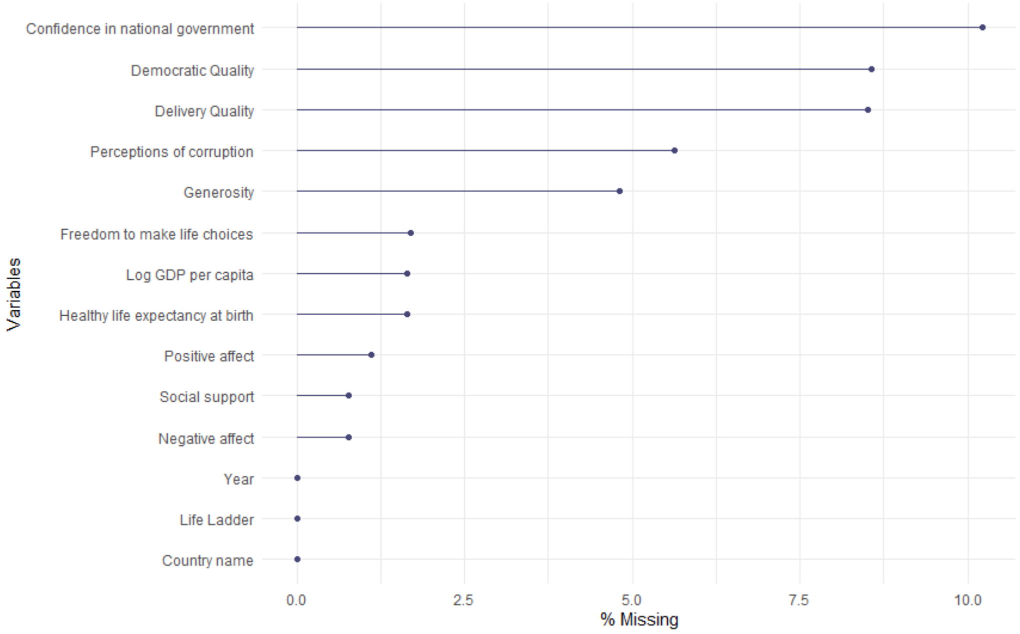


Figure 2: Showing The Distribution Of Missing Values in Happiness Data

While the development of proposed analytical model to predict happiness score as a function of several risk factors, one of the most important assumptions is the normality of response (dependent variable). That is, the response variable Ladder should follow the Gaussian probability distribution. The mid-values of happiness score seem to be reasonably straight, but the ends are skewed to a certain degree, as can be seen from the Q-Q plot shown by Figure 3.

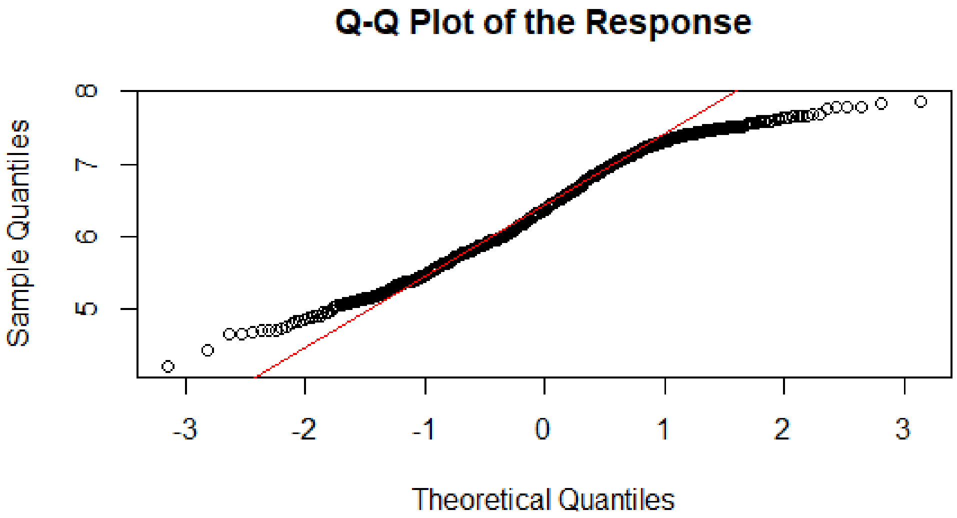


Figure 3: Q-Q Plot Of The Response Ladder

We have also shown through goodness-of-fit test (Shapiro-Wilk normality test, p-value = $5.565e^{-10}$) that the response Ladder does not follow the normal probability distribution. Thus, the Q-Q plot supports that the national average of happiness scores, do not follow the Gaussian probability distribution. The correlation plot of of the risk factors is shown in Figure 4, where negative correlations are presented in red and positive correlations in blue color. The color intensity and the degree of relationship between each pair of risk factors are proportional to the correlation coefficients. From the following correlation matrix in Figure 4, we see that there are strong positive associations between the variables LIFE_EXPECT and DEL_QUALITY, Generosity, and DEL_QUALITY and DEM_QUALITY and DEL_QUALITY. Also, there is a strong negative association between the variables LOG_GDP and PER_CORR and PER_CORR and CONF_GOV. Thus, we would implement some regularization techniques such as Ridge Regression (L_1), Lasso Regression (L_2), and Elastic net regressions to take into account the over-fitting issue and compare their performance in terms of $RMSE$ and MAE.

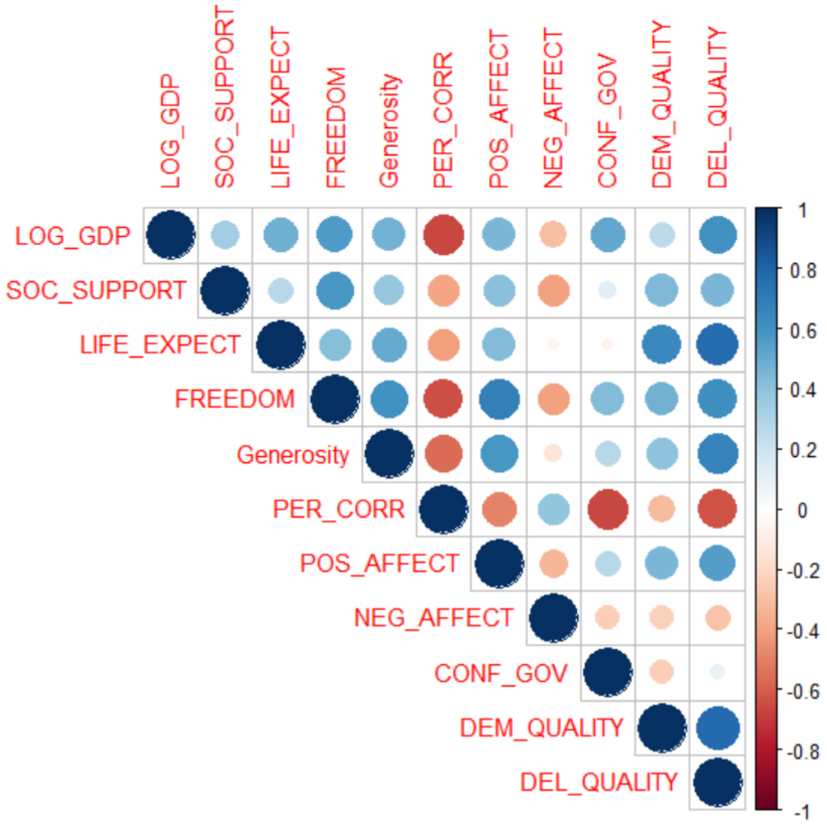


Figure 4: Correlation Matrix of The Attributable Variables

2.2 Development of Statistical Model

We now start developing the non-linear analytical model, which is driven by the national average of happiness score as a function of the eleven risk factors and all possible interactions, as discussed previously. The general structure of our non-linear model with all possible interactions and additive error structure, is given by:

$$Ladder = \beta_0 + \sum_i \alpha_i x_i + \sum_j \gamma_j k_j + \epsilon \quad , \tag{1}$$

where β_0 is the intercept term of the model, α_i is the coefficient of i^{th} individual risk factor x_i, γ_j is the coefficient of j^{th} interaction term k_j and ϵ is the random error term of the model, that follows a normal distribution with zero mean and constant variance.

One of the main suppositions to develop the above model is that the response variable should follow the Gaussian probability distribution. As we have shown above, the dependent variable Ladder does not follow the Gaussian probability distribution initially. Therefore, we utilize a non-linear transformation to filter our happiness data so that it follows the normal probability distribution. We used Johnson transformation for our response, which is given by:

$$z = \gamma + \delta ln\Big(\frac{x-\epsilon}{\lambda+\epsilon-x}\Big) \quad , \qquad \epsilon < x < \epsilon + \lambda$$

and

$$TLadder = -0.43 + 0.87 ln\Big(\frac{x-4.2}{3.72+4.2-x}\Big) \quad . \tag{2}$$

Here, *T Ladder* denotes the new response variable(transformed) after the use of Johnson S_U transformation to our old response. We now estimate the coefficients (weights) of the risk factors for the processed data in equation 2. To develop our analytical model, we initially proceed with the full statistical model, including all eleven risk factors and ten possible interactions between each pair. Thus, initially, we start structuring our model with $\binom{n}{k} = 55(n = 11, k = 2)$ terms that include the primary contribution of the risk factors and every possible interactions. As we began with the full statistical model (fifty-five terms), as mentioned, we have applied the backward elimination(26) method to identify the most significant contributions of both the individual attributable variables and interactions by eliminating the less important risk factors gradually. Furthermore, backward elimination is deemed one of the best traditional methods for a set of feature vectors to encounter the problem of overfitting and carry out feature selection.

Though, the statistical estimation method of our data analysis has indicated that only seven out of the eleven risk factors significantly contribute and twenty-eight interaction terms, we can not omit the risk factors that are not significant, and simultaneously include any risk factor interacting with it in the model. Thus, the best proposed statistical model with all risk factors and significant interactions that estimates the average happiness score accurately are eleven risk factor individually, and the twenty-eight interaction term, which is given by:

$$\widehat{TLadder} = \begin{cases} -0.45 + 0.42X_1 + .12exp(X_2) + .04X_3 + .27X_4 + \\ .16X_5 + .03exp(-X_6) + .12exp(-X_7) - .07X_8 + .03X_9 \\ -.07X_{10} - .1X_{11} - .17X_1X_3 + .14X_1X_4 - .13X_1X_5 + \\ .27X_1X_6 - .03X_1X_7 - .11X_1X_8 + .22X_1X_{11} - .1X_1X_5 \\ +.19X_5X_6 + .14X_6X_8 + .22X_2X_9 + .16X_2X_{10} + .15X_2X_{11} + \\ .07X_3X_7 - .06X_3X_{10} - .43X_4X_6 - 0.19X_4X_8 - 0.29X_4X_9 + \\ 0.10X_4X_{10} - 0.30X_4X_{11} + 0.18X_5X_6 + 0.10X_5X_9 - .2X_5X_{10} + \\ .32X_5X_{11} - .02X_6X_{11} + 0.1X_7X_8 + 0.1X_7X_9 + 0.1X_7X_{11} \end{cases} \tag{3}$$

The $\widehat{TLadder}$ can be computed from equation (3) and is based on the Johnson transformation(18) of the data. We now proceed to utilize the anti-transformation on equation (3) to estimate the actual estimate national average of happiness score $\widehat{Ladder}$ as follows:

$$\widehat{Ladder} = 4.2 + \frac{3.72}{1 + exp\left(\frac{T\widehat{Ladder}+0.43}{0.87}\right)} \quad . \tag{4}$$

The proposed analytical model will assist social scientists and economists acknowledge how the happiness score changes when any of the eleven risk factors is varied by keeping the other risk factors fixed at the same time. Likewise, with the variation of significant interaction. Anyone, interested to know the optimum levels of the risk factors at which the happiness score is maximized, can do the same by using any analytical optimization technique. We now illustrate the percentage of contributions of the risk factors and the interactions to the happiness score as shown below in Table 1.

Rank	Risk Factors	Contr.(%)
1	LOG_GDP	7.15
2	$FREEDOM \cap PER\_CORR$	5.58
3	$LOG\_GDP \cap PER\_CORR$	5.00
4	$FREEDOM$	4.94
5	$EXP(POS\_AFFECT)$	4.63
6	$FREEDOM \cap CONF\_GOV$	4.46
7	$FREEDOM \cap NEG\_AFFECT$	4.13
8	$CONF\_GOV \cap SOC\_SUPPORT$	3.89
9	$NEG\_AFFECT \cap SOC\_SUPPORT$	3.72
10	$GENEROSITY$	3.72
11	$FREEDOM \cap DEL\_QUALITY$	3.59
12	$GENEROSITY \cap DEL\_QUALITY$	3.45
13	$EXP(SOC\_SUPPORT)$	3.30
14	$GENEROSITY \cap DEM\_QUALITY$	3.16
15	$LOG\_GDP \cap DEL\_QUALITY$	2.96
16	$PER\_CORR \cap SOC\_SUPPORT$	2.87
17	$LOG\_GDP \cap LIFE\_EXPECT$	2.55
18	$POS\_AFFECT \cap NEG\_AFFECT$	2.45
19	$LOG\_GDP \cap NEG\_AFFECT$	2.44
20	$CONF\_GOV \cap POS\_AFFECT$	2.38

Table 1: Ranking of Individual Risk Factors and the Interactions With Respect to The Percentage of Contribution to The Response

To evaluate the quality of the proposed analytical model(17), we use both the coefficient of determination, R^2, and adjusted R^2, which are the basic criteria to evaluate the model performance. The sum of squares due to regression(SSR) is the squared sum of the differences between the predicted response and the mean response. It captures the observed variability of the model. The sum of squared errors (SSE), also termed as the residual sum of squares, is the variation that remains unexplained. We always try to minimize this error in a model The total sum of squares (SST) = SSE + SSR. R^2, the coefficient of determination, is defined as the proportion of the total response variation that is explained by the proposed model, and it measures how well the regression process approximates the real data points. Thus, R^2 is given by

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad .$$

However, R^2 itself does not consider the number of variables in the model. Also, there is the problem of the increasing R^2 with addition of variables in the model. To address these issues, we have the adjusted R^2, which considers the number of parameters and is given by

$$R^2_{adj} = 1 - \left[\frac{(1-R^2)(n-1)}{n-k-1}\right] \quad ,$$

where n is the number sample data points, and k is the number of independent risk factors used in the model, excluding the constant. For our final statistical model, the R squared is 88.8%, and R squared adjusted is 87.8%. Both R squared and R squared adjusted are very high and very close to each other. That is, the developed statistical model explains 88.8% of the variation in the response variable, a very high-quality model. Similarly, the risk factor that we included in the model, along with the relevant interactions, estimates almost 89% of the total variation in the happiness score. In Figure 4, we rank the individual attributable variables and interactions (top 20) with respect to their contribution to the national happiness score. That is, we listed those terms based on their percentage of contribution to the response. In a survey or experiment, if the group of experimenters or surveyors know beforehand the most significant variables which account for the response, they might be interested in collecting information on those important variables only, which might save some experimental resources.

2.3 Verifying Model Assumptions

Once the statistical model has been developed, it is necessary to check the model assumptions (if any). In our case, we have proposed a multiple non-linear regression model, which is very useful and conveys to us accurately some important information on the subject matter. However, multiple linear regression has some important assumptions which must be satisfied with the correctness of the proposed model. In this section, we will verify the important model assumptions.

2.3.1 Mean Residual should be Close to Zero

When one performs multiple linear regression (or any other type of regression analysis), one obtains a linear function that best fits the data. The entire data points usually don't fall exactly on this regression plane, but they are scattered around it. The residual(error)$\hat{\epsilon}$ is defined as:

$$\hat{\epsilon} = \text{residual} = \text{observed value-predicted value} = y - \hat{y} \quad ,$$

where y and $\hat{y}$ are the observed and predicted response. $\hat{e}$ is the estimated residual error from the linear fit. The sum of the residuals equals zero, assuming that the regression function is actually the "best fit."In our case, the mean residual is $-1.56 * 10^{-18}$, implying that it is almost zero. Figure 5 below illustrates the behavior of the residual estimator.

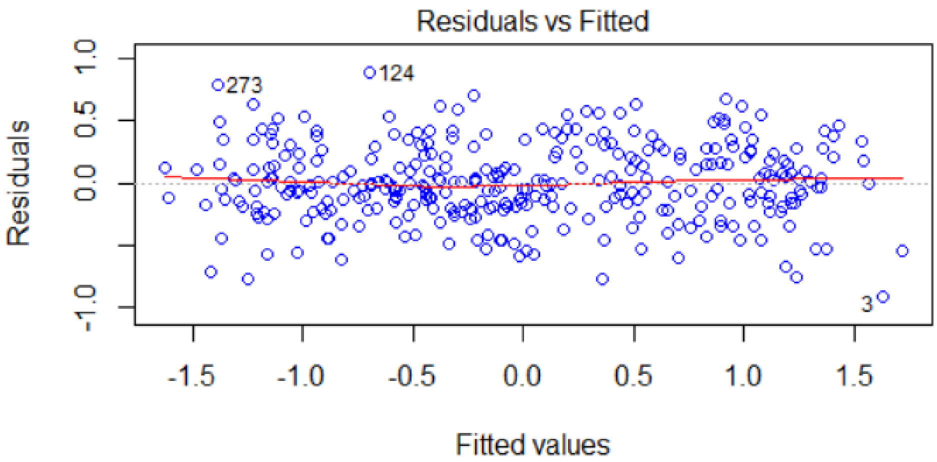


Figure 5: Fitted Vs. Residual Plot

2.3.2 Homoscedasticity of residuals

One of the main assumptions of the linear regression model is the homoscedasticity of the residuals or equal variance. That is, $Var(\hat{e}) = \sigma^2$ which is constant. From the above Figure 5, we see that residuals vary as the fitted values increase. It seems that the pattern is more or less uniform, which is shown by the red line. There is no increasing or decreasing trend. Hence, the assumption of the constant variance of the residuals has been satisfied.

Breusch-Pagan Test: Breusch-Pagan (BPG) test is used to test for heteroskedasticity of the error terms in a regression model. We obtained a p-value of $.35173$ by testing the null hypothesis of constant error variance against the alternative that the error variance changes with the level of the response (fitted values) or with a linear combination of predictors. Hence, we have significant reason to believe the error variance is constant.

2.3.3 Normality of residual

One important assumption of linear regression is normality of residual. From Figure 6 and Figure 7, we see that the studentized residual follows a normal pattern.

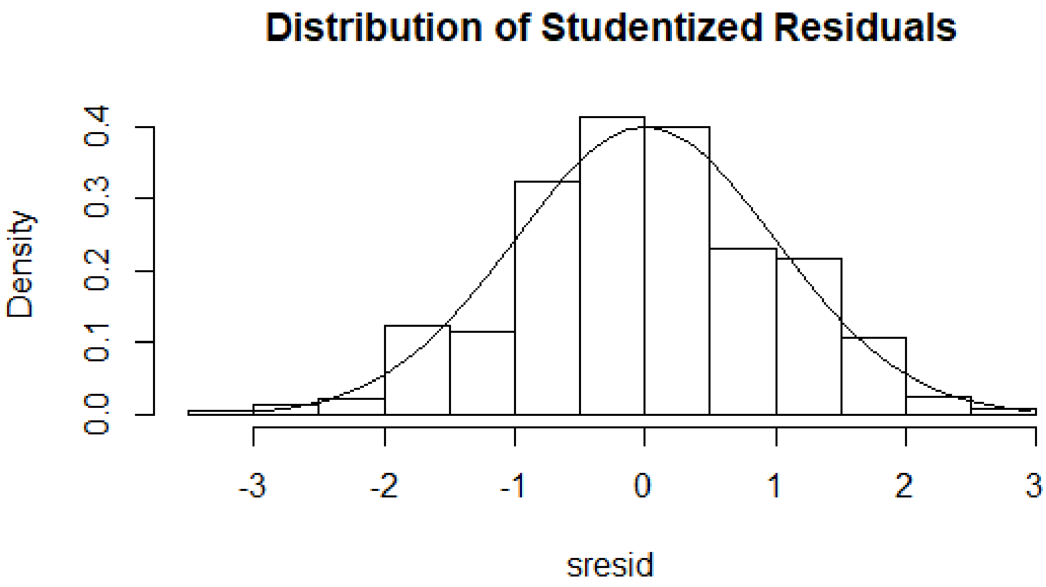


Figure 6: Normality of Studentized Residual(sresid) Plot

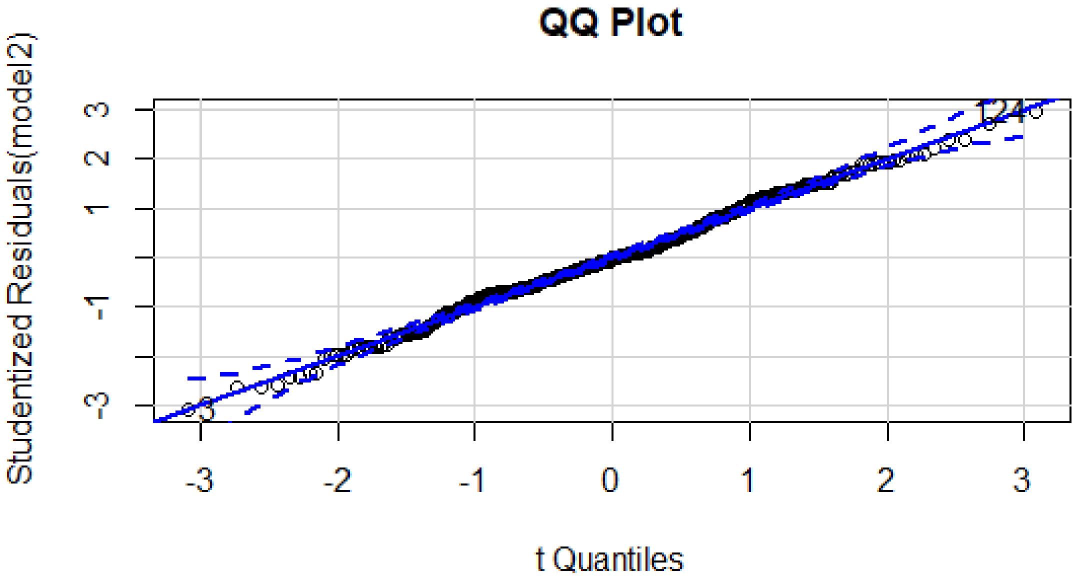


Figure 7: Q-Q Plot of Studentized Residuals

2.3.4 No auto-correlation between the residuals

We proceed to test the auto-correlation between the error terms of our model. The correlation between two error terms is defined as,

$$corr(\hat{e_i},\hat{e_j})=\begin{cases} 0, & \text{if } i \neq j. \\ 1, & \text{if } i = j. \end{cases}$$

where $\hat{e_i}$ and $\hat{e_j}$ are the i^{th} and j^{th} error terms in the model.

The following Figure 8 shows the autocorrelation of the residuals vs. lag plot. The X-axis corresponds to the lags of the residuals. The first line to the left shows the correlation of residuals with itself (Lag0); therefore, it will always be equal to 1. If the residuals were **not auto-correlated**, the correlation (Y-axis) from the immediate next line onwards would drop to a near-zero value below the dashed blue line (significance level). Hence, there is no auto-correlation between residuals in our model.

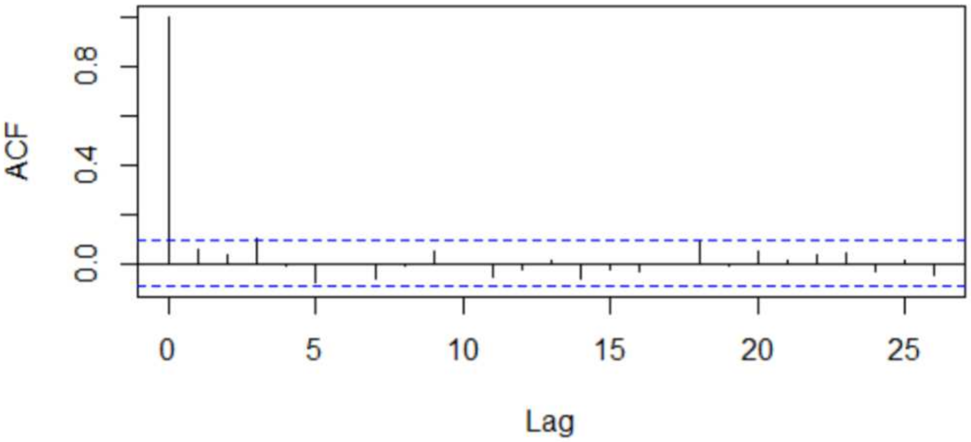


Figure 8: Showing The Auto-Correlation Plot of Residuals

RUN TEST: Also, we can verify the no auto-correlation case of the residuals by Run test (Wald, A. and Wolfowitz, J. (1940)(24). Runs test examines the randomness of a numeric sequence by studying the frequency of runs. We obtained a p-value of 0.9264, which implies that we fail to reject the null hypothesis that residuals are random. Hence, there is no pattern.

Durbin-Watson test: The Durbin Watson Test(25) is a measure of auto-correlation (also called serial correlation) in residuals from the regression analysis. Auto-correlation is the similarity of a time series over successive time intervals. It can underestimate the standard error and can cause one to believe that the predictors are significant when they are not. The Durbin–Watson test statistic is used to detect the presence of autocorrelation at lag 1 in the residuals (also termed as prediction errors) in regression analysis. The test statistic for this test is given by:

$$DW = \frac{\sum_{t=2}^{T} \left(\hat{e_t} - \hat{e_{t-1}} \right)^2}{\sum_{t=2}^{T} \hat{e_t}^2} \quad ,$$

where $\hat{e_t}$ and $\hat{e_{t-1}}$ are the residuals at time points t and $t-1$, respectively.
A rule of thumb is that the test statistic values in the range of 1.5 to 2.5 are relatively normal. Values outside of this range could be a cause for concern. Field(2009) suggests that values under 1 or more than 3 are a definite cause for concern. The value we obtained for the test statistic is 1.89 with a p-value of $.109$, implying that there is insufficient sample evidence to reject the null hypothesis that the true auto-correlation in zero.

5. The regressors and the residuals are nor correlated: We calculated the Pearson's product-moment correlation coefficient between each regressor and the residuals. As expected, every time we obtained an insignificant p-value implying that the true correlation is zero.

We further studied the fact that if there are other statistical models that give better useful results than the proposed nonlinear regression model. Thus, we developed some penalized regression models and compared those with our proposed model. These models are given in the following section.

3 Penalized Regression Models

Penalized regression methods have proven to be a high-yielding area of research in statistics and data sciences. The key idea is to add a 'penalty' to regression to encourage desirable behavior in the model. Often this is done to reduce variability in estimating the parameters. While developing the proposed statistical model for happiness, we used OLS, the ordinary least square technique to obtain an approximate estimate of the coefficients (weights) of the attributable variables. To address the multicollinearity problem (since in our data set, some variables are strongly correlated), the Regularization methods are used. Since these methods are based on adding the regularization parameters(lambda and alpha) to the regression coefficients of the individual risk factors, these the model generalizes the data and prevents over-fitting. To further illustrate our proposed model's quality, we will discuss three machine learning regularization methods and our proposed non-linear analytical model.

3.1 Ridge Regression

For multiple linear regression, the ordinary least squares fitting procedure of the coefficient estimates(weights) $\beta_1, \beta_2, ...\beta_p$ that minimizes the cost function RSS (Residual Sum Of Squares), is given by,

$$RSS = \textstyle\sum_{i=1}^{n} \Big(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij}\Big)^2 \quad .$$

Ridge regression is very similar to least square regression, except that the ridge coefficients are estimated by minimizing a slightly different quantity. In particular, ridge regression coefficient estimates $\hat{\beta^R}$ are the values that minimizes the following function:

$$\sum_{i=1}^{n} \big(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij}\big)^2 + \lambda \sum_{i=1}^{p} {\beta_j}^2 \quad , \tag{5}$$

where $\lambda \geq 0$ is a tuning parameter (sometimes called a penalty parameter that controls the strength of the penalty term in ridge regression) to be determined via cross validation.

3.2 LASSO (Least Absolute Shrinkage and Selection Operator) Regression

The LASSO regression model appends an absolute value of magnitude of a coefficient as penalty term to the loss function that is given by:

$$\sum_{i=1}^{n} \Big(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij}\Big)^2 + \lambda \sum_{i=1}^{p} \mid \beta_j \mid \quad . \tag{6}$$

Comparing (5) to (6), we see that the LASSO and Ridge regression have similar formulations. The only difference is that the β_j^2 term in the ridge regression penalty in (6) has been replaced by $\mid \beta_j \mid$ in the LASSO penalty (6). In statistical literature, the LASSO uses an L_1 penalty where the

Ridge uses L_2 penalty.

3.3 Elastic Net

Elastic Net regression model is the combination of Ridge and LASSO regression methods. The loss function of elastic net model can be defined by:

$$\sum_{i=1}^{n}\Big(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\Big)^2+\lambda\Big[(1-\alpha)\sum_{i=1}^{p}{\beta_j}^2+\alpha\sum_{i=1}^{p}\mid\beta_j\mid\Big]\quad . \tag{7}$$

However, in the above equations (5, 6 and 7) the constructions of the three models will be the same structure as our proposed model in equation (1) with only the coefficient estimation will be different because of the randomness in selecting the training data set.

4 Comparison among different Models

We now proceed to compare the performance of the proposed model with the other three models using the following two matrices.

4.1 Root Mean Squared Error (RMSE)

After each repetition of the cross-validation, the model assessment metric RMSE is computed, which is given by:

$$RMSE=\sqrt{\frac{\sum_{i=1}^{n}\Big(y_i-\hat{y_i}\Big)^2}{n}}\quad ,$$

where y_i and $\hat{y_i}$ are the observed and predicted responses.

4.2 Mean Absolute Deviation (MAE)

The MAE measures the average magnitude of the errors in a set of forecasts, without considering their direction which is given by

$$MAE=\frac{\sum_{i=1}^{n}|y_i-\hat{y_i}|}{n}\quad ,$$

where y_i and $\hat{y_i}$ are the observed and predicted responses.

While comparing the proposed model with the three regularization methods Ridge, LASSO, and Elastic Net, we have found that our proposed analytical model performs better in terms of validations matrices RMSE and MAE, as described above. Table 2 below provides multiple comparisons among the different models in terms of training and testing accuracy.

Table of Comparison				
Matric Models	RMSE		MAE	
	Training	Testing	Training	Testing
Proposed Model	**.31**	**.43**	**.24**	**.31**
RIDGE	.38	.5	.3	.35
LASSO	.36	.52	.27	.37
EN	.36	.52	.29	.37

Table 2: Comparison Among Different Models in terms of RMSE and MAE

From the above Table 2, we see that our proposed nonlinear statistical model gives minimum testing error in terms of RMSE and MAE when compared with the penalized regression models. Thus, our analytical model outperforms the other three models for our happiness data.

5 Validation and Prediction Accuracy of The Proposed Model

We developed our analytical model on 80% training data and validated the model based on 20% testing data. In the testing data (validation data), the test error is the average error that occurs from using the analytical method to predict the response on a new set of observations. That is a measurement that was not used in training the method. The test error gives an idea about the consistency of the analytical model. Moreover, we performed repeated ten-fold repeated cross-validation (10 times) for our validation testing. The primary objective is that we will use 10-fold cross-validation, then we repeated cross-validation ten times, where each of the repetition folds are split differently. In 10-fold cross-validation, the training set is divided into ten equal subsets. One of the subsets is taken as the testing set in turn, and (10-1) = 9 subsets are taken as a training set in the proposed model. The error mean square error E_1 is computed for the held out set. This procedure is repeated ten times; each time, a different group of observations is treated as a validation set. This process results in 10 estimates of the test error, $E_i, \quad i = 1, \ldots 10$. The average error of each set throughout the cross-validation process is termed as a cross-validated error. The following Figure 9, illustrates briefly the idea of 10 fold repeated cross-validation, where $E_i, \quad i = 1, \ldots 10$ is the mean square error (MSE) in each iteration and ACVE is the average cross-validated error.

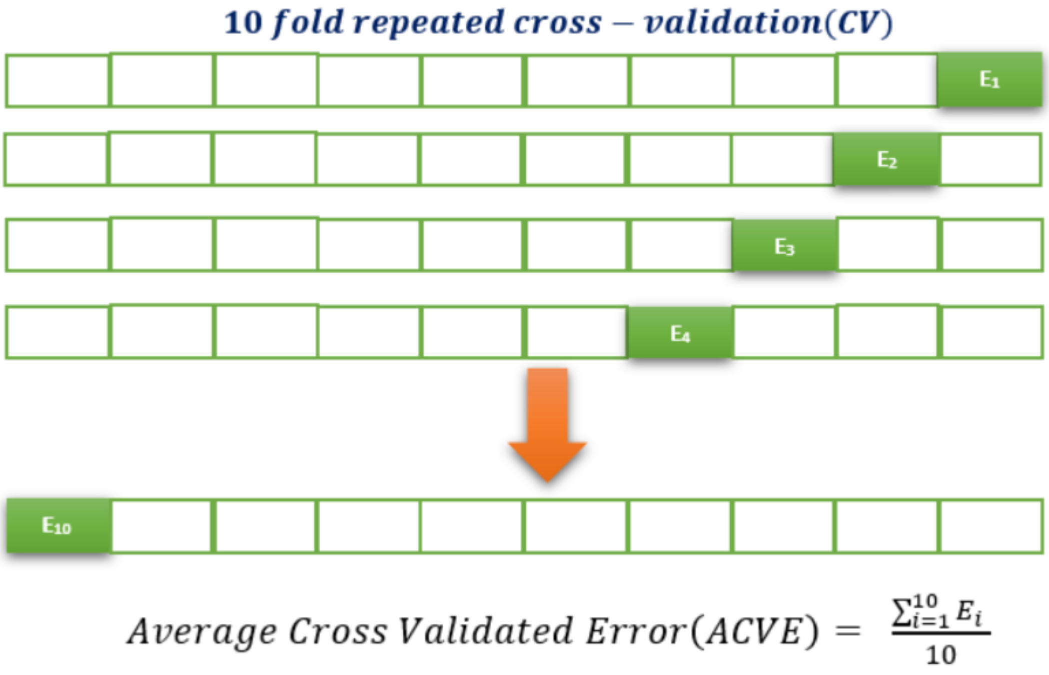

10 fold repeated cross – validation(CV)
E1
E2
E3
E4
E10
Average Cross Validated Error(ACVE) = $\frac{\sum_{i=1}^{10} E_i}{10}$

Figure 9: Brief Illustration Of Repeated Ten Fold Cross Validation

Now we employ the following three methods to illustrate the prediction accuracy of the proposed model.

5.1 Min-Max Accuracy

Min-Max-Accuracy is the average of the ratio of minimum value between the actual observation and predicted observation and maximum between actual observation and predicted observation. Mathematically, it can be expressed as follows:

$$Min - Max - Accuracy = mean\left[\frac{min(y_i,\hat{y_i})}{max(y_i,\hat{y_i})}\right] \quad ,$$

where y_i and $\hat{y}_i$ are the observed and predicted response.
It gives an idea about how far the model's prediction is off on an average. For a perfect model, this measure is 1.This can be taken as the accuracy of the proposed model. For our developed model, the Min-Max accuracy is **96.2%**, which is quite impressive.

5.2 Correlation Accuracy

A simple correlation between the original observations and predicted observations can be used as a form of accuracy measure. A greater correlation accuracy implies that the original and predicted observations have analogous directional movement, i.e., when the original observations increase, the predicted observations also increase and vice-versa. We obtained a correlation accuracy of **90.5%** in the test data, which implies that our statistical model is of high quality and should be useful for applied predictive analysis for real data.
Table 3 below provides the two measures of prediction accuracy for our proposed model.

Thus, the above two methods attest to the high quality of our proposed model.

Min-Max-Accuracy	Correlation Accuracy
96.2%	90.5%

Table 3: Prediction Accuracy for the Proposed Model

6 Discussions

After obtaining the significant risk factors along with their significant interactions, we rank them with respect to the percent of contribution to the happiness scores for the developing countries as shown by figure 4. The risk variable that has the largest contribution to the happiness score is the variable **LOG_GDP** which contributes 7.15% of the total variation to the happiness score. The next largest contribution is the combined effect of freedom and perception of corruption with a 5.58% contribution. Numbers 3, 4, and 5 are the combined interaction effect of LOG_GDP, FREEDOM, and exp(POS_AFFECT) with a contribution of 5%, 4.94%, and 4.63%, respectively. Hence, adding these risk factors up, we see that they explain almost 89% of the total variability in the national average happiness score for all developing countries. We can address the usefulness and importance of the proposed model in the subject area in **five** important categories.
These categories are given below.

1. We have identified and tested the individual attributable variables(risk factors) responsible for the change in happiness score across all the developed countries.
2. we have identified the significant interactions that influence the happiness score in our model.
3. we have ranked the individual risk factors and interactions as a percentage of contribution for the the response of the national average of happiness score (Ladder) or subjective well-being (SWB).
4. We can obtain excellent predictions of happiness of individuals given the values of the attributable variables from our analytical model with a high degree of accuracy.
5. Any particular country might use our non-linear statistical model to work on specific risk factors to increase their happiness score. For example, one can work on the variable SOC_SUPPORT (X2) if the value for a particular year is not satisfactory and work on other important aspects to increase the value so that the happiness score can be increased.

We have also ranked all the developed countries based on the **predicted happiness score** of the most recent observations (data) available for the year 2019. The following Table 4, illustrates the ranking of the countries.

Rank	Country	Score
1	Finland	7.67
2	Denmark	7.55
3	Sweden	7.54
4	Iceland	7.38
5	United States	7.35
6	Canada	7.29
7	Ireland	7.17
8	Switzerland	7.16
9	United Kingdom	7.09
10	Germany	7.03
11	Malta	6.98
12	Luxembourg	6.96
13	Oman	6.96
14	Estonia	6.92
15	Singapore	6.89
16	Qatar	6.82
17	France	6.76
18	Uruguay	6.73
19	Slovenia	6.62
20	Malaysia	6.61
21	United Arab Emirates	6.56
22	Saudi Arabia	6.54
23	Netherlands	6.53
24	Argentina	6.51
25	Australia	6.51
26	Austria	6.51
27	Bahrain	6.51
28	Belarus	6.51
29	Belgium	6.51
30	Czech Republic	6.46
31	Norway	6.43
32	Israel	6.38
33	Lithuania	6.35
34	Chile	6.34
35	Spain	6.31
36	Slovakia	6.24
37	Japan	6.24
38	Hungary	6.14
39	Poland	6.12
40	New-Zealand	6.11
41	Cyprus	6.09
42	Italy	6.06
43	Kazakhstan	6.05
44	Russia	5.99
45	South Korea	5.94
46	Kuwait	5.70
47	Turkey	5.60
48	Croatia	5.55
49	Portugal	5.45
50	Montenegro	5.38
51	Latvia	5.35
52	Bulgaria	5.34
53	Greece	5.20
54	Romania	5.04

Table 4: Ranking of Developed Countries based on Predicted Happiness Score

It is interesting to note that Finland and Denmark possess the top happiness scores while the *United States* is fifth. Also, studies(19) has shown a significant influence of democracy on an individuals' subjective well-being (happiness). Finland and Denmark falling into the category of the top democratic countries of the world also validate the fact.

7 Conclusion

We have developed a real data-driven analytical model that very accurately identifies the following very useful findings concerning the happiness of the society of developed countries in the world:

- Identifies the significant attributable variables (risk factors) that drives the degree of happiness.
- Identifies the significant interactions of the risk factors that contribute to the degree of happiness.
- We rank the individual and interactions of the risk factors with respect to their percentage of contribution to the degree of happiness.
- The developed analytical model predicts the degree of happiness very accurately for a given response to a set of questions.

- The developed model can be used strategically to increase the degree of happiness by working with the identified risk factors.
- Furthermore, one can perform surface response analysis to identify the target values of the risk factors so as to be, say, 95 percent sure that we will maximize the degree of happiness based on the identified values.

The developed analytical model has been evaluated by several statistical methods that include the R^2 and $R^2_{adjusted}$ that attest to its high quality. The risk factor **LOG_GDP** is the highest contributor to the happiness score contributing 7.15%, while **DEL_QUALITY** contributes the least with 1.31% to the response. The findings of our study suggest that economists and other social scientists might need to pay more attention to emotional well-being as a causal force. Also, since individual happiness in an organization has a positive correlation with *productivity*, our proposed statistical model can be used for firms' promotion policies, and they may be useful for managers and human resources professionals. Human resource managers can use our model to predict the individual happiness score by using the questionnaire (*attached in appendix 1).* It will help the company to identify those individuals who need to be rewarded and those who need to improve their happiness score. Identifying those individuals are essential for the company as happiness is correlated with an increase in productivity. Our proposed statistical model is also highly useful for *decision making* and *strategic planning* on controlling the factors responsible for causing people to be unhappy and depressed. Finally, since happiness is the most crucial aspect of human life that we seek, controlling the most critical risk factors that significantly contribute to the happiness are essential to control the crime rate of a country, as there is a negative correlation between the individual country's happiness score(Ladder) and crime rate.

8 Conflict of Interest

The authors declare that they have no conflict of interest.

9 Compliance of ethical standard

Not applicable

10 Informed consent

Not applicable

11 Acknowledgements

The authors are grateful to Gallup, Inc. for collecting the happiness data and make it available online(https://worldhappiness.report/ed/2019/).

12 Funding

Not applicable

13 Authors' contributions

AC performed the statistical analysis and structured the manuscript. CT provided the research idea.

14 Availability of data and material

The data can be obtained online from the source `(https://worldhappiness.report/ed/2019/).`

References

[1] Andrew J. Oswald, Eugenio Proto, and Daniel Sgroi,Happiness and Productivity,*Journal of Labor Economics, Vol. 33, No. 4 (October 2015), pp. 789-822*

[2] Thomas V. Perneger, Patricia M. Hudelson & Patrick A. Bovier, Health and happiness in young Swiss adults, *Quality of Life Research 13: 171–178, 2004.*

[3] DeNeve KM, Cooper H. The happy personality: A metaanalysis of 137 personality traits and subjective well-being. *Psychol Bull 1998; 124: 197–229.*

[4] Pettijohn TF II, Pettijohn TF. Perceived happiness of college students measured by Maslow's hierarchy of needs. *Psychol Rep 1996; 79: 759–762.*

[5] Frey BS, Stutzer A. *Happiness prospers in democracy. J Happiness Studies 2000; 1: 79–102.*

[6] Myers DG, Diener E. The pursuit of happiness. *Sci Am 1996; 274: 54–56.*

[7] Cammock T, Joseph S, Lewis CA. Personality correlates of scores on the Depression–Happiness Scale. *Psychol Rep 1994; 75: 1649–1650.*

[8] George MS, Ketter TA, Parekh PI, Horwitz B, Herscovitch P, Post RM. *Brain activity during transient sadness and happiness in healthy women. Am J Psychiatry 1995; 152: 341–351.*

[9] Koivumaa-Honkanen H, Honkanen R, Viinama ki H, Heikkila K, Kaprio J, Koskenvuo M. Self-reported life satisfaction and 20-year mortality in healthy Finnish adults. *Am J Epidemiol 2000; 15: 983–991.*

[10] Berwick DM, Murphy JM, Goldman PA, Ware JE, Barsky AJ, Weinstein MC. Performance of a five-item mental health screening test. *Med Care 1991; 29: 169–176.*

[11] McDowell I, Praught E. On the measurement of happiness. An examination of the Bradburn scale in the Canada Health Survey. *Am J Epidemiol 1982; 116: 949–958.*

[12] Bradburn N. The Structure of Psychological Well-Being. *Chicago: Adline, 1995.*

[13] Walsh J, Joseph S, Lewis CA. Internal reliability and convergent validity of the Depression–Happiness Scale with the General Health Questionnaire in an employed adult sample. *Psychol Rep 1995; 76: 137–138.*

[14] A. E. Hoerl, R. W. Kennard, Ridge regression: Biased estimation for nonorthogonal problems, *Technometrics 12 (1970) 55–67.*

[15] R. Tibshirani, Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society: Series B (Methodological) 58 (1996) 267–288.*

[16] H. Zou, T. Hastie, Regularization and variable selection via the elastic net,*Journal of the royal statistical society: series B (statistical methodology) 67 (2005) 301–320.*

[17] Abu Sheha, M. and Tsokos, C. (2019) Statistical Modeling of Emission Factors of Fossil Fuels Contributing to Atmospheric Carbon Dioxide in Africa. *Atmospheric and Climate Sciences, 9, 438-455. doi: 10.4236/acs.2019.93030.*

[18] Alan M. Polansky, Youn-Min Chou & Robert L. Mason (1999) An Algorithm for Fitting Johnson Transformations to Non-Normal Data, *Journal of Quality Technology, 31:3, 345-350, DOI: 10.1080/00224065.1999.11979933*

[19] DAVID DORN, JUSTINA A.V. FISCHER, GEBHARD KIRCHGASSNER and ALFONSO SOUSA-POZA DEMOCRACY AND CULTURE ON HAPPINESS *Social Indicators Research (2007) 82: 505–526,DOI: 10.1007/s11205-006-9048-4*

[20] Morris, T.P., White, I.R. & Royston, P. Tuning multiple imputation by predictive mean matching and local residual draws.*BMC Med Res Methodol 14, 75 (2014).* https://doi.org/10.1186/1471-2288-14-75

[21] Rubin, D. (1986). Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations.*Journal of Business & Economic Statistics, 4(1), 87-94. doi:10.2307/1391390*

[22] Helliwell, J., Layard, R., & Sachs, J. (2019). *World Happiness Report 2019, New York: Sustainable Development Solutions Network.* (`https://worldhappiness.report/ed/2019/#read`)

[23] Thomas, M. What Do the Worldwide Governance Indicators Measure?. *Eur J Dev Res 22, 31–54 (2010). https://doi.org/10.1057/ejdr.2009.32*

[24] Wald, A., & Wolfowitz, J. (1940). On a Test Whether Two Samples are from the Same Population. *The Annals of Mathematical Statistics, 11(2), 147-162. Retrieved December 9, 2020, from http://www.jstor.org/stable/2235872*

[25] White, Kenneth J. "The Durbin-Watson Test for Autocorrelation in Nonlinear Models." *The Review of Economics and Statistics, vol. 74, no. 2, 1992, pp. 370–373. JSTOR, www.jstor.org/stable/2109675. Accessed 9 Dec. 2020.*

[26] HALINSKI, R.S. and FELDT, L.S. (1970), THE SELECTION OF VARIABLES IN MULTIPLE REGRESSION ANALYSIS. *Journal of Educational Measurement, 7: 151-157. https://doi.org/10.1111/j.1745-3984.1970.tb00709.x*

A Appendix

A.1 In the appendix, our version of the survey questionnaire for World Happiness Report 2019 by Gallup Poll is posted which is the modified version and we request the same type of information

SURVEY QUESTIONNAIRE

Based on who is requesting the information for an individual that is associated with any Government, Company, Organization, Educational Institutions, etc.

GDP(X1): Per-capita gross domestic product of the country the individual resides (given information)

Social Support(X2): If you were in trouble, do you have relatives or close friends, you can count on to help you whenever you need them? A)YES □ B)NO □

Life Expectancy(X3): From the attached graph, identify your life expectancy. □Years.

Freedom(X4): Are you satisfied with your freedom to choose what you do with your life?
A) YES □ B) NO □

Generosity(X5): Have you donated money to a charity in the past month? A) YES, B) No. If the answer is YES, then how much?

Corruption Perception(X6): Is corruption widespread throughout your government, your company, or your organization? A) YES□ B) NO □

Positive Affect(X7): Happiness, laughter, and enjoyment.

7.1. On a scale of 1 to 10, how **happy** were you for the last five days?
7.2. On a scale of 1 to 10, how much did you **laugh** for the last five days?
7.3. On a scale of 1 to 10, how much did you **enjoy** for the last five days?

Negative Affect(X8): Worry, Sadness, and anger, respectively.

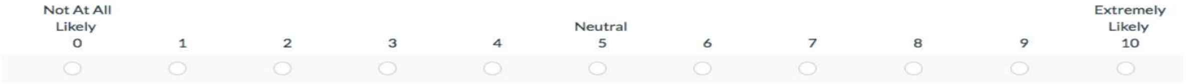
8.1. On a scale of 1 to 10, how **worried** were you for the last five days?
8.2. On a scale of 1 to 10, how **sad** were you for the previous five days?
8.3. On a scale of 1 to 10, how **angry** were you for the last five days?

Confidence in Government(X9): In your government, company, or organization, etc. how much trust and confidence do you have when it comes to handling [International problems/Domestic problems]?

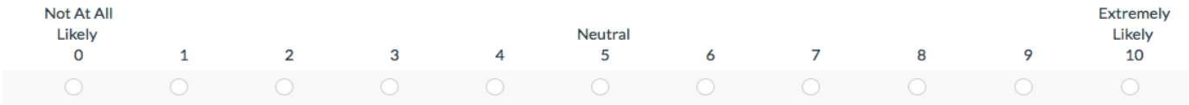
A) A great deal □ B) A fair amount□
C) not very much □ D) none at all □

Democratic Quality (X10):

10.1. On a scale of 1 to 10, how likely do you think that the country's citizens can participate in selecting their government, enjoy Freedom of expression, Freedom of association, and unprejudiced media coverage?



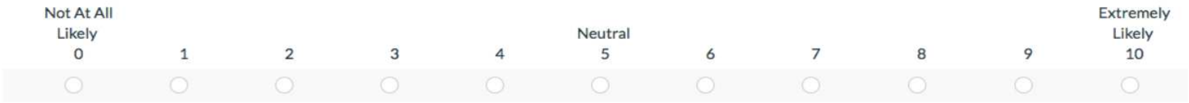
10.2. On a scale of 1 to 10, how likely you think people suffer consequences of political instability and politically motivated violence, including terrorism?



Not At All Likely 0	1	2	3	4	Neutral 5	6	7	8	9	Extremely Likely 10
○	○	○	○	○	○	○	○	○	○	○

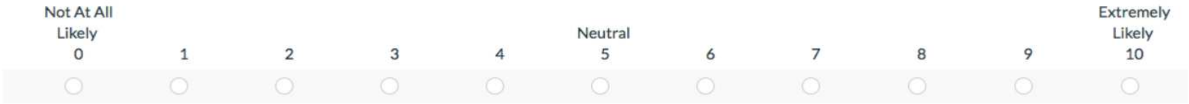
Delivery Quality(X11):

11.1. On a scale of 1 to 10, how likely do you think that your company/organization/government has maintained the quality of public services, the quality of the civil service, the quality of policy formulation and implementation, and the credibility of such policies?



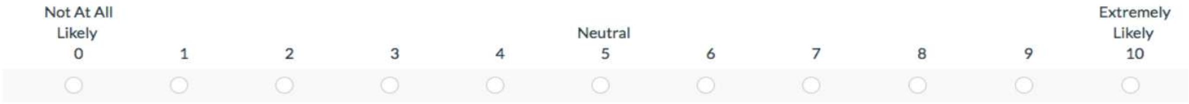
Not At All Likely 0	1	2	3	4	Neutral 5	6	7	8	9	Extremely Likely 10
○	○	○	○	○	○	○	○	○	○	○

11.2. On a scale of 1 to 10, how likely do you think that your company/organization/government can formulate and implement sound policies and regulations that permit and promote private sector development?



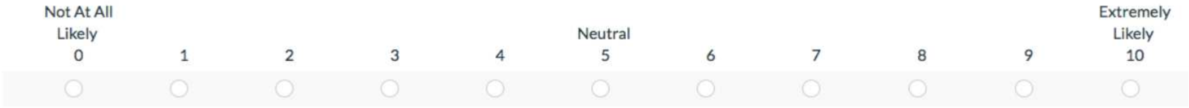
Not At All Likely 0	1	2	3	4	Neutral 5	6	7	8	9	Extremely Likely 10
○	○	○	○	○	○	○	○	○	○	○

11.3. On a scale of 1 to 10, how likely do you think that your company/organization/government agents and law enforcement agencies have confidence in the government and abide by society's rules?



Not At All Likely 0	1	2	3	4	Neutral 5	6	7	8	9	Extremely Likely 10
○	○	○	○	○	○	○	○	○	○	○

11.4. On a scale of 1 to 10, to what extent you think that public power is exercised for private gain, including both petty and grand forms of corruption by the elites for their individual interests?



Not At All Likely 0	1	2	3	4	Neutral 5	6	7	8	9	Extremely Likely 10
○	○	○	○	○	○	○	○	○	○	○

Figures

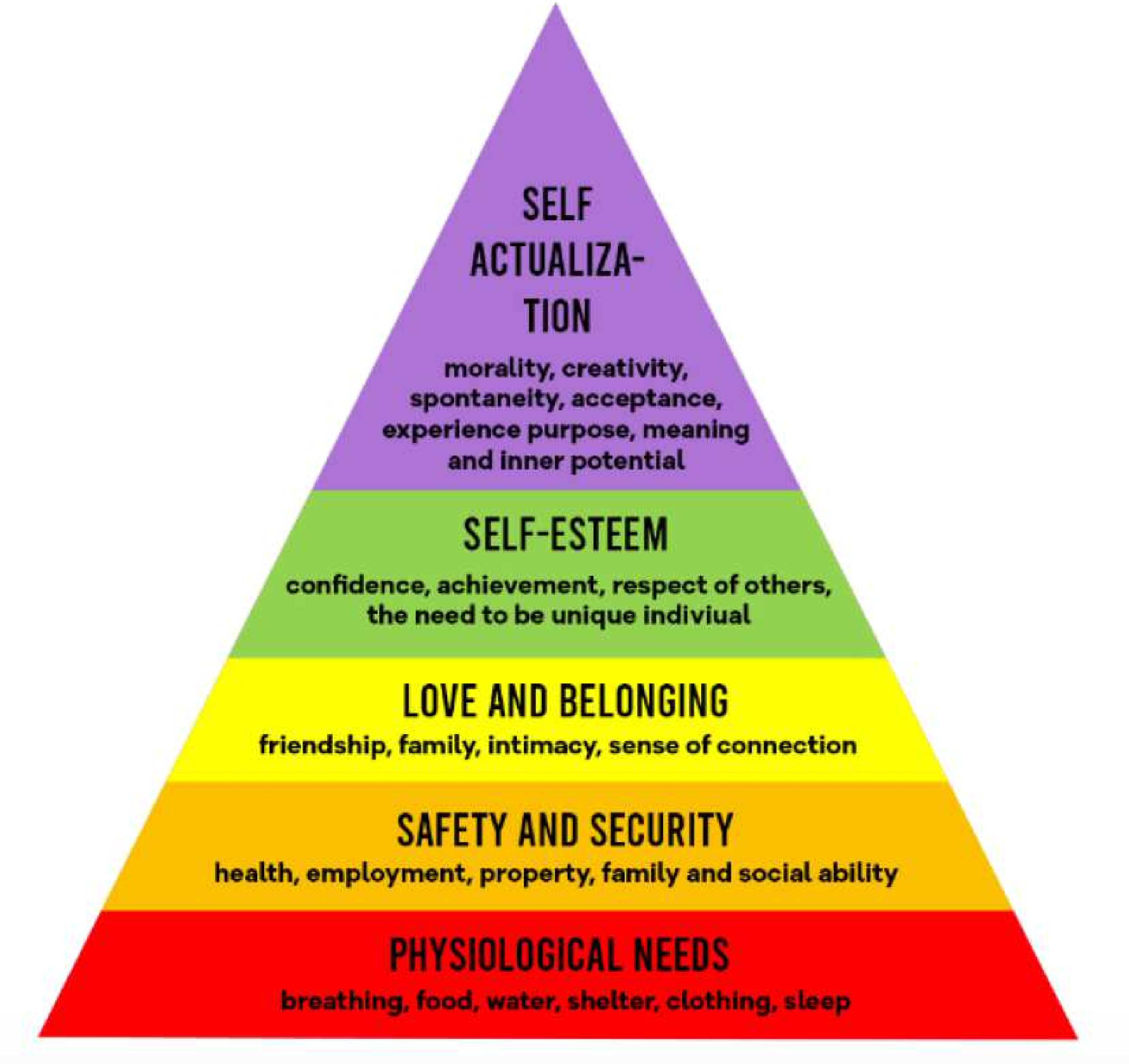

SELF ACTUALIZA- TION
morality, creativity, spontaneity, acceptance, experience purpose, meaning and inner potential
SELF-ESTEEM
confidence, achievement, respect of others, the need to be unique indiviual
LOVE AND BELONGING
friendship, family, intimacy, sense of connection
SAFETY AND SECURITY
health, employment, property, family and social ability
PHYSIOLOGICAL NEEDS
breathing, food, water, shelter, clothing, sleep

Figure 1

Triangle Showing Hierarchy of Needs For Subjective Well Being source by:https://irechargeme.com/wellbeing/

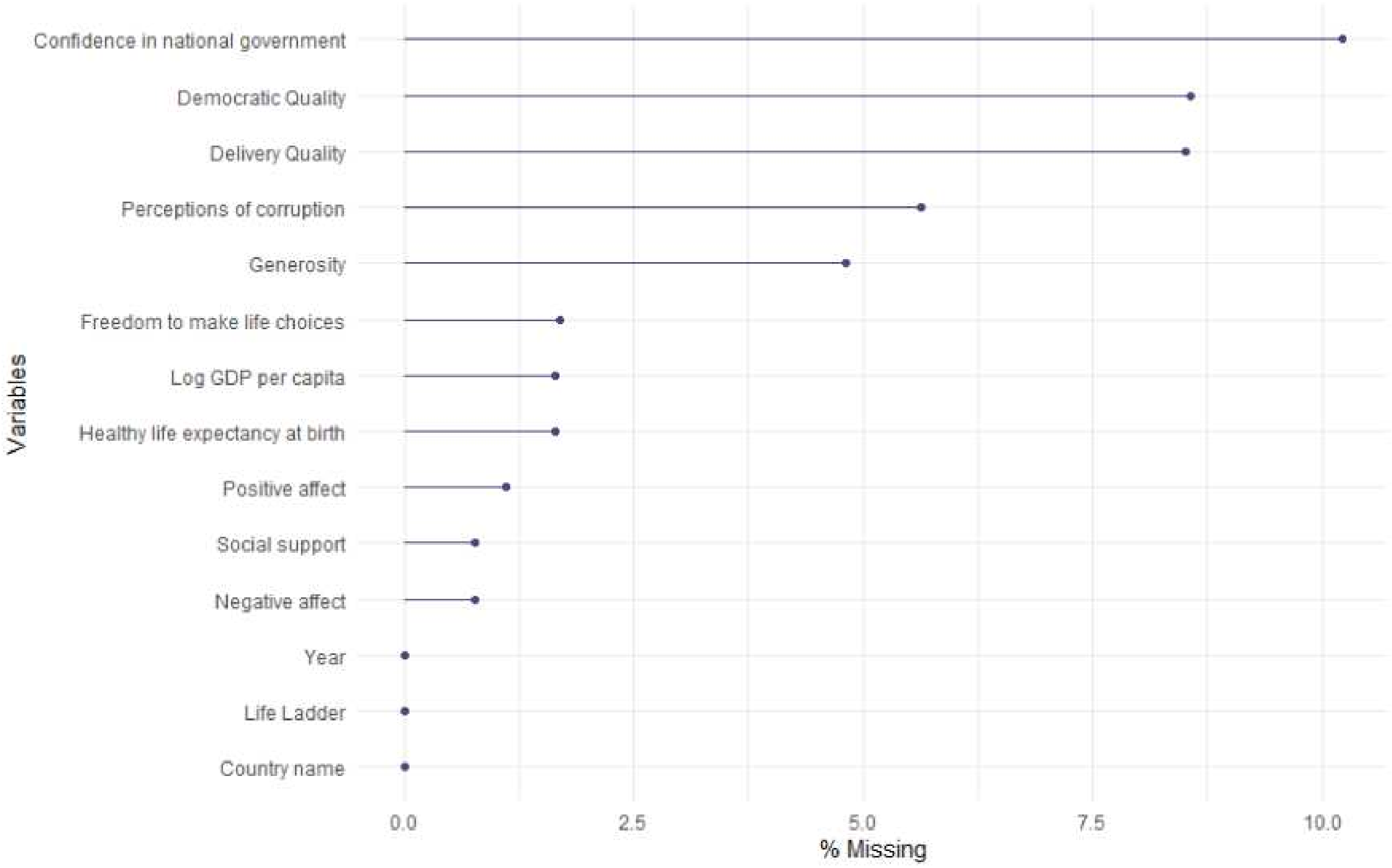


Figure 2

Showing The Distribution Of Missing Values in Happiness Data

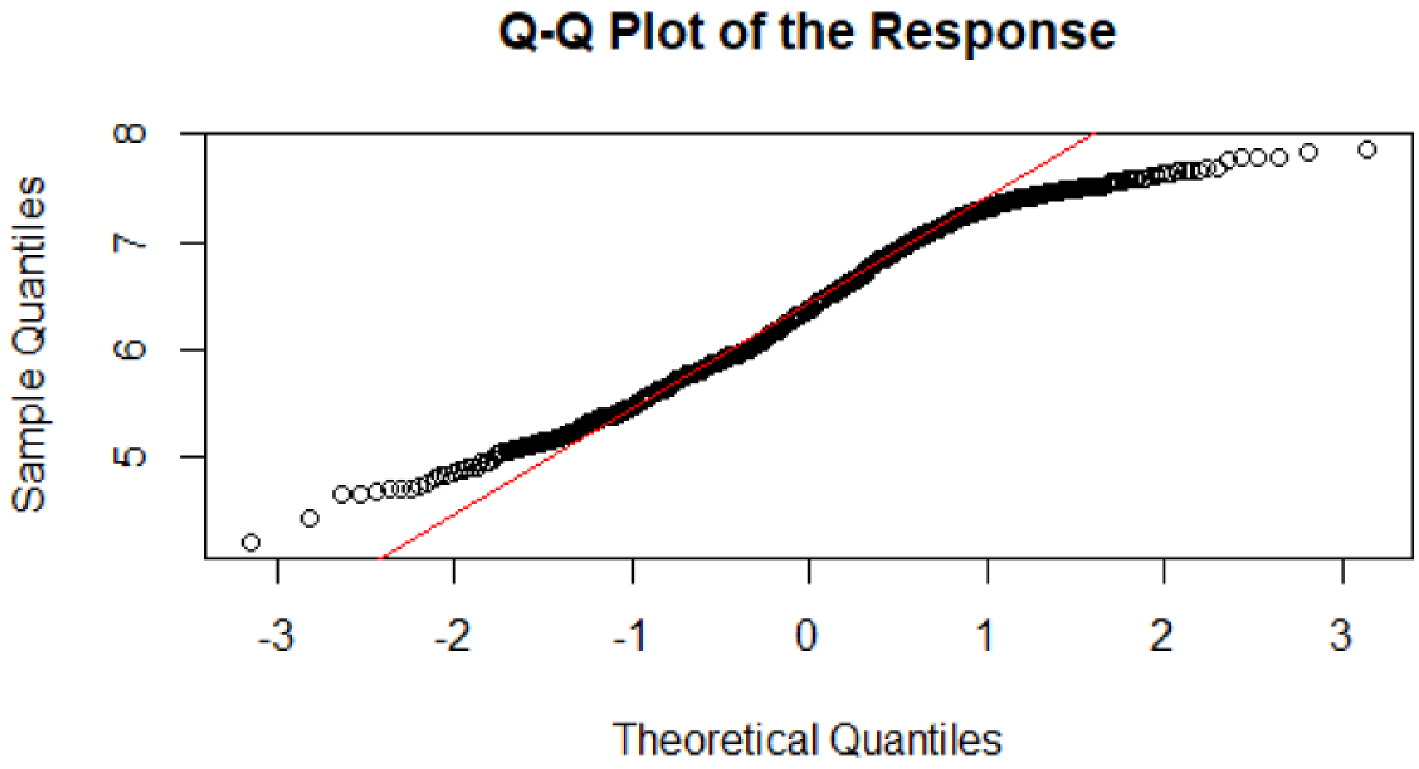


Figure 3

Q-Q Plot Of The Response Ladder

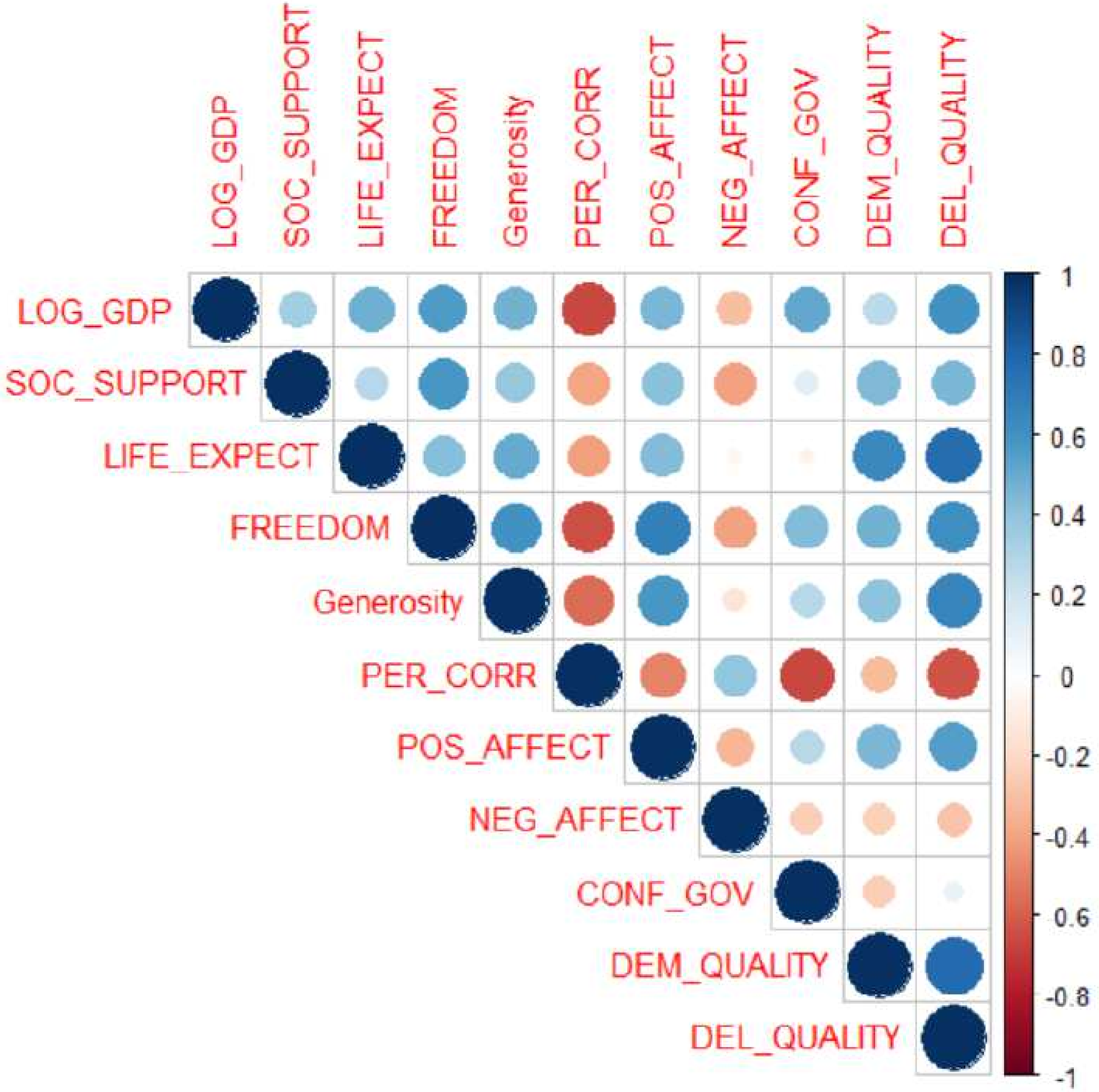


Figure 4

Correlation Matrix of The Attributable Variables

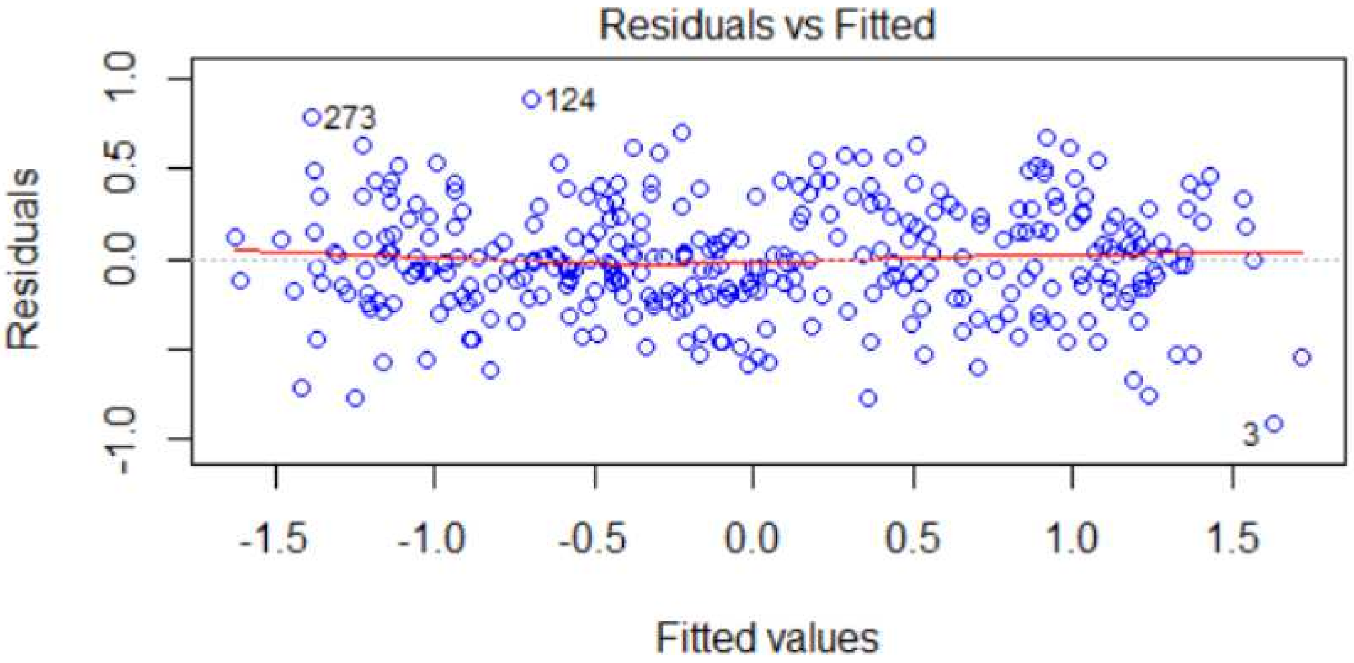


Figure 5

Fitted Vs. Residual Plot

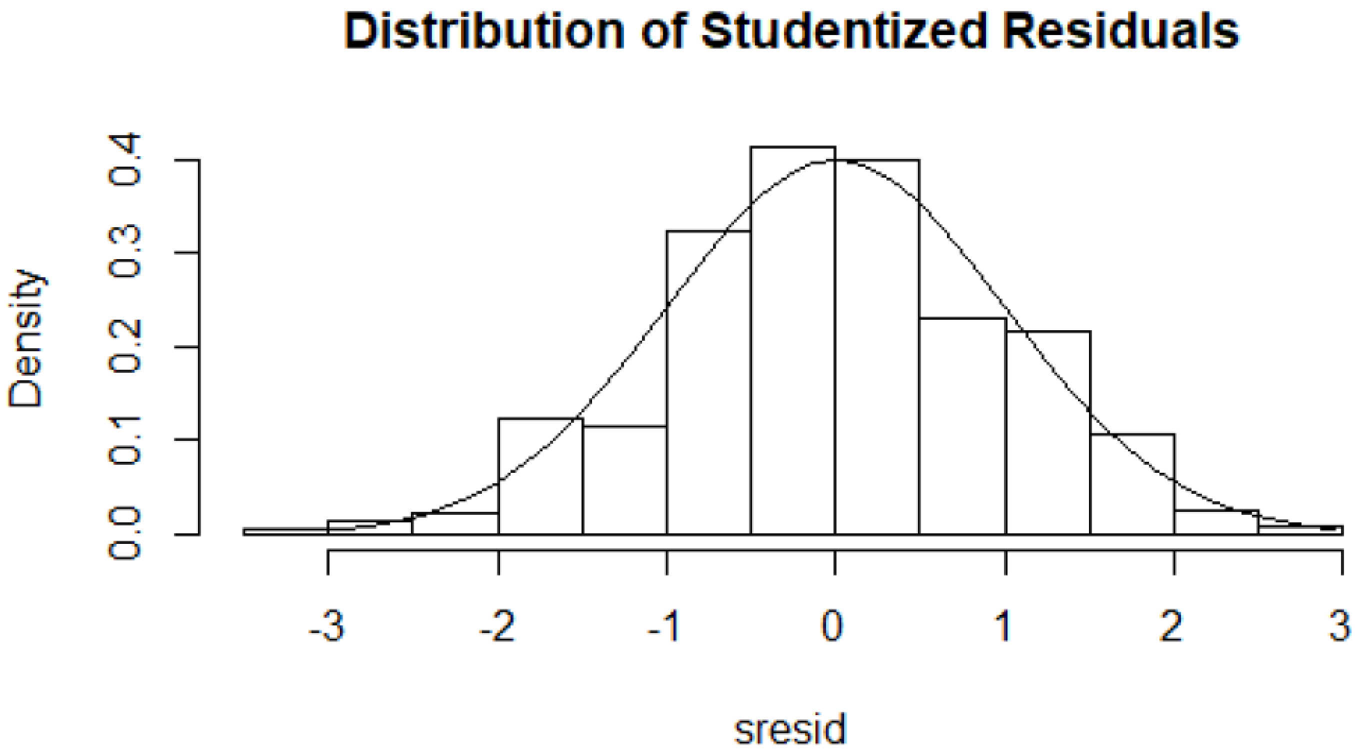


Figure 6

Normality of Studentized Residual(sresid) Plot

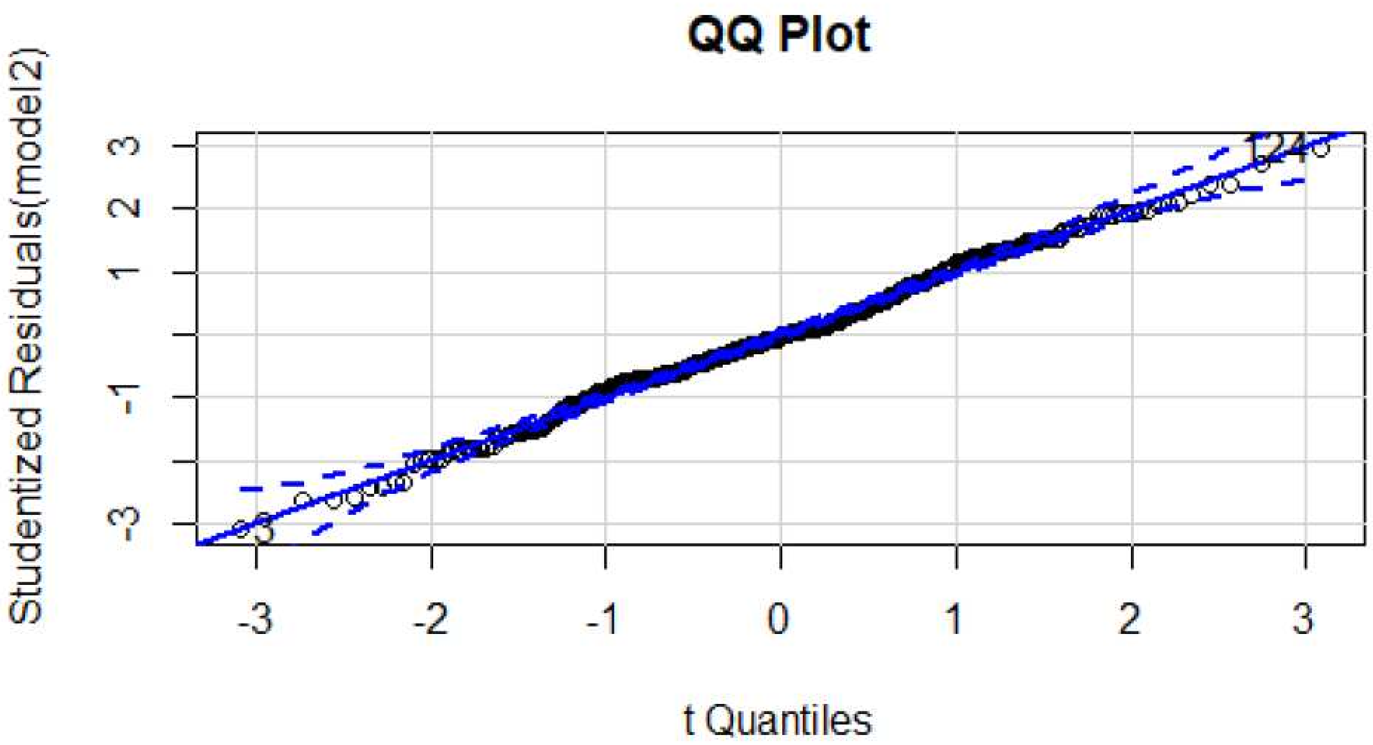


Figure 7

Q-Q Plot of Studentized Residuals

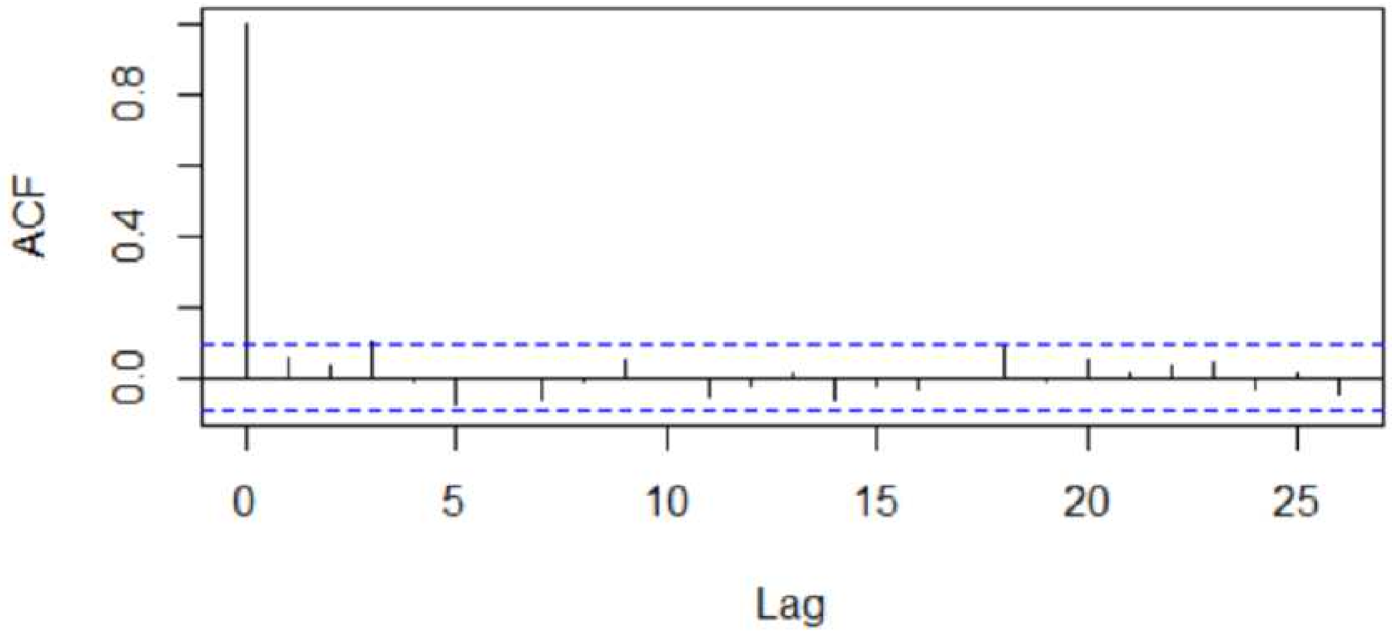


Figure 8

Showing The Auto-Correlation Plot of Residuals

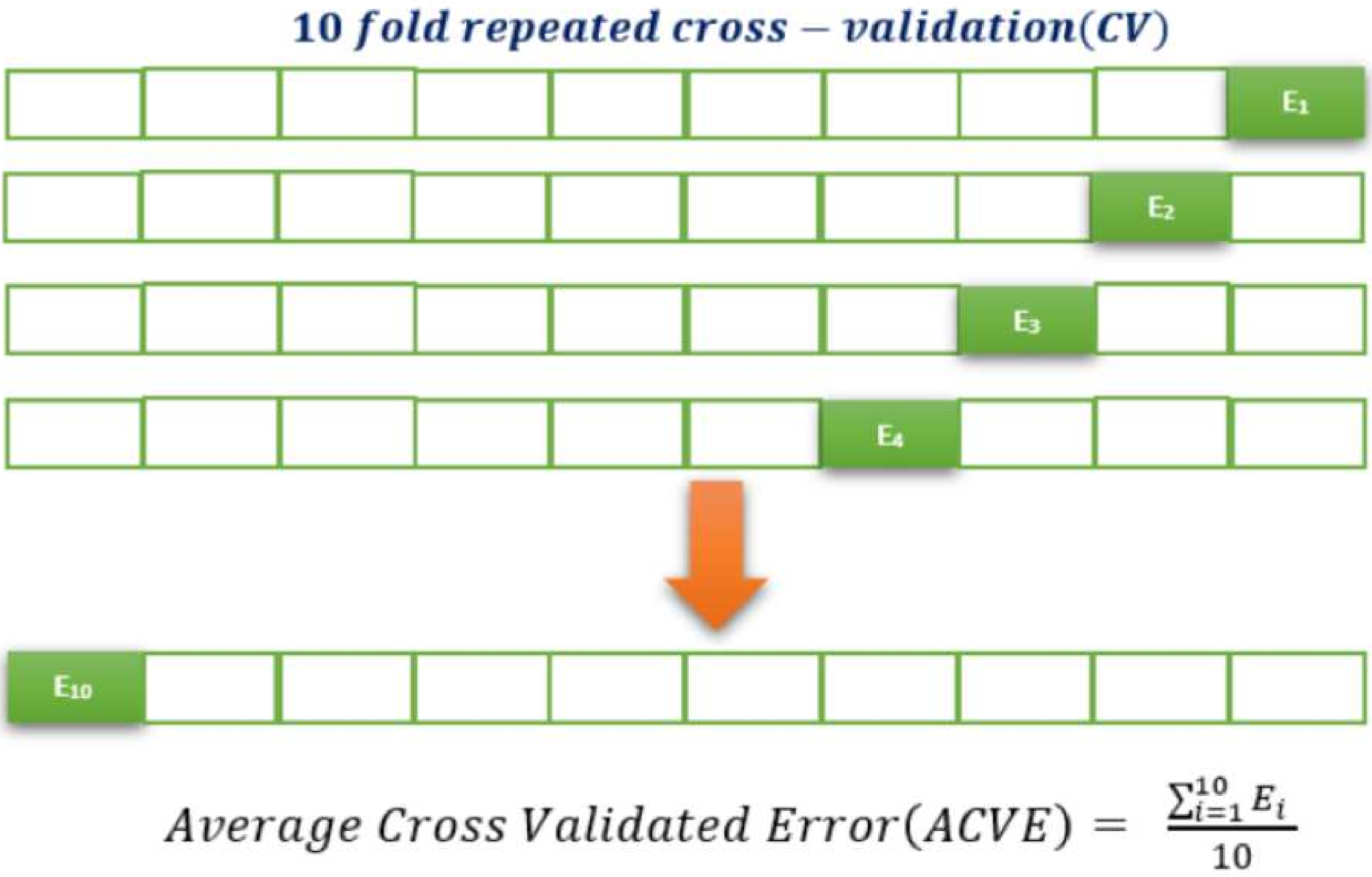

$10\ fold\ repeated\ cross - validation(CV)$
E_1
E_2
E_3
E_4
E_{10}
$Average\ Cross\ Validated\ Error(ACVE) = \frac{\sum_{i=1}^{10} E_i}{10}$

Figure 9

Brief Illustration Of Repeated Ten Fold Cross Validation