Predictive factors of happiness and policy countermeasure in post COVID-19 pandemic era

Qihang Zheng

School of Statistics, Renmin University of China, Beijing, China

2020201470@ruc.edu.cn

Abstract. The COVID-19 pandemic has severely affected numerous individuals' lives and inevitably lowered their happiness levels. In the process of recovering from the effects of the pandemic, a series of measures must be taken to enhance people's happiness as soon as possible. The research used multiple linear regression and random forest to analyze the data from World Happiness Report 2023 in order to identify some effective predictive factors of happiness, and made recommendations to policymakers. The result of this study shows social support and GDP are the main indexes policymakers should put effort into improving. Governments can enhance the level of social support not only from the classical perspective but also from improving health care. Reformations of the health care system are necessary if current contingency measures are not sufficient for such severe situations as COVID-19. Policies on economics are continuously beneficial to a population's overall well-being and thus are important for policymakers to work on.

Keywords: Happiness, predictive factors, COVID-19 pandemic, policymakers.

1. Introduction

Throughout human history, the pursuit of happiness has been a fundamental concept in society. A metric of happiness that has drawn a lot of attention recently as an assessment of a nation's general prosperity and standard of living is the Happiness Index Score [1]. The United Nations released its first World Happiness Report in April of 2012 after realizing the significance of quantifying happiness worldwide. On the basis of the Cantril ladder survey, the World Happiness Report has been published yearly since then. The study also includes six factors—"GDP per capita", "Social support", "Healthy life expectancy", "Freedom to make life choices", "Generosity", and "Perceptions of corruption"— that are identified to be related to a nation's happiness index based on the survey's data.

The widespread pandemic of COVID-19 in 2020 had a significant impact on people's productivity and quality of life. The continuously shifting relationship between supply and demand, as well as the compulsory prevention and control measures, made many industries tough, and many individuals face the risk of unemployment, which will inevitably lower people's happiness levels [2]. Meenu Gupta et al. studied data from the World Happiness Report during the COVID-19 pandemic and discovered that GDP, happiness, and social support all play a significant effect in a country's COVID-19 impact [3]. In March 2023, the United Nations released the World Happiness Report(WHR) 2023. The data of this report was collected at a time when many countries were gradually recovering from the impact of the

COVID-19 pandemic. Some of them had lifted all of the restrictions, while some others were carrying out routine production and life in daily pandemic prevention and control. This is not the case during the COVID-19 pandemic, and a series of measures must be taken to enhance people's happiness and pull them out of depression as soon as possible. Therefore, this article will study and try to understand the impact of various factors on individual happiness levels based on data from WHR 2023. Furthermore, Khder et al. analyzed the data from the World Happiness Report 2019(before the pandemic) and ranked GDP per capita as the most important variable leading to high happiness levels and Healthy life expectancy as the second [4]. Tarushi et al. Tried to predict happiness scores during COVID-19 and identified Healthy life expectancy to be the most significant affecting factor and GDP the second [5]. These researches offered insight into the situation before and during the pandemic and were helpful for this study to draw a comprehensive conclusion. Hopefully, the findings in this paper could help policymakers and other stakeholders identify areas for intervention and improvement in order to improve the population's overall well-being.

In terms of modeling, this study compared two machine learning methods, multiple linear regression and random forest. Based on these models, it determines which indexes can be improved to increase resident happiness and makes policy recommendations to today's policymakers. Furthermore, the models presented in this paper can be used to predict the happiness score of a specific country based on the parameters defined. This may have implications for policymakers as well.

2. Methodology

2.1. Data
This study used data from the WHR 2023 released by the United Nations, and the dataset was downloaded on the Kaggle website. The report collected data on happiness-related indicators in 137 countries. After filtering out one country with missing key statistical data, a total of 136 pieces of valid data points were used in modeling. Based on the research purpose, this paper selected seven variables in total, namely Ladder score, Logged GDP per capita, Social support, Healthy life expectancy, Freedom to make life choices, Generosity, and Perceptions of corruption. Among them, the Ladder score is the dependent variable, and the rest are independent variables. Table 1 offers a brief description of these seven variables. 80% of the data is used for training, and the remaining 20% is used for testing.

Table 1. Description of the variables

Ladder score	The happiness score for each country ,ranging from 0 to 10.
Logged GDP per capita	The logged GDP per capita of each country.
Social support	The score of social support for each country.
Healthy life expectancy	The healthy life expectancy of each country.
Freedom to make life choices	The score of freedom to make life choices for each country.
Generosity	The generosity score of each country.
Perceptions of corruption	The score of the perception of corruption in each country.

2.2. Multiple linear regression
It is a supervised machine learning algorithm. It consists of a target (dependent) variable and three or more independent variables. By plotting the data on a linear graph using the equation y=kx+b, where k is the tan slope and b is the intercept (the point where the line intersects the y axis), this method aims to fit the data and find linear relationships between dependent and independent variables. It is simple but practical if the independent and dependent variables have a linear relationship.

2.3. Random forest

Random Forest is a machine learning algorithm that can be used for regression modeling as well as classification. When it comes to regression, RandomForest constructs numerous independent decision trees and then aggregates the forecasted outcomes of each tree to generate an ultimate regression prediction. By averaging the predictions made by each decision tree in the ensemble, Random Forest produces regression predictions. The average value of all decision trees' predictions is the final prediction in a regression.

For regression modeling, Random Forest has the advantage of handling datasets with complex patterns and unstructured features. The result of the modeling is reliable and unaffected by slight adjustments in the dataset. Because modeling uses bootstrapping and a subset of random features, it is capable of preventing overfitting. Random forests also have certain drawbacks, such as low reliability, resistance to overfitting extremely complex data, and sensitivity to unbalanced datasets [6].

3. Results and Discussion

Before the discussion of the modeling results, some descriptive statistical analyses had been made to get fundamental knowledge of the research data. Table 2 shows some of the descriptive statistics of variables in the dataset.

Table 2. Descriptive statistics of variables

Variable	Min	Max	Mean	Std.dev.
Ladder score	1.859	7.804	5.544	1.143
Logged GDP per capita	5.527	11.66	9.455	1.210
Social support	0.341	0.983	0.799	0.130
Healthy life expectancy	51.53	77.28	64.97	5.750
Freedom to make life choices	0.382	0.961	0.788	0.112
Generosity	-0.254	0.531	0.024	0.142
Perceptions of corruption	0.146	0.929	0.725	0.177

The average happiness score was 5.544, and the highest score was 7.804 in Finland which has been ranked the happiest country for six consecutive years. By comparison, lowest score was 1.859 in Afghanistan. Figure 1 demonstrates the correlation of the variables (The variables are presented briefly in the figure, e.g. "Freedom" represents Freedom to make life choices which is mentioned above).
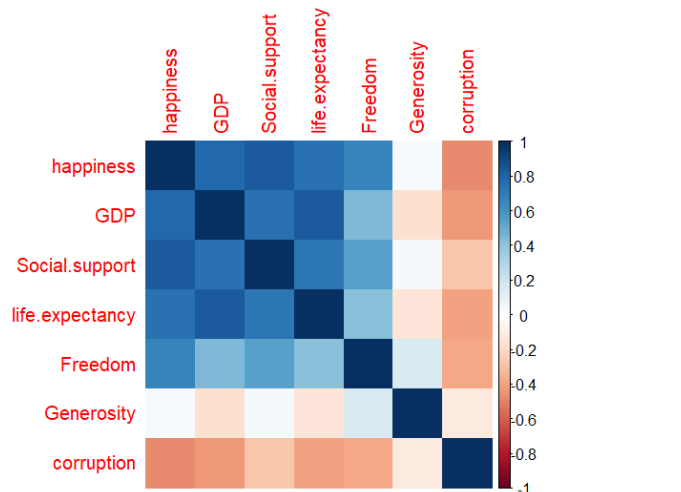

Figure 1. Correlation diagram

The independent variable "Generosity" has a very low correlation with the dependent variable "Happiness score". This could provide evidence that "Generosity" has little impact on a country's happiness level.

Furthermore, all 6 independent variables were included in the full model using multiple linear regression. The result can be seen in Table 3.

Table 3. Results of the full MLR model

Variable	Estimate	P value	
Logged GDP per capita	0.203	0.005	**
Social support	3.923	0.000	***
Healthy life expectancy	0.020	0.154	
Freedom to make life choices	2.340	0.000	***
Generosity	0.143	0.658	
Perceptions of corruption	-0.798	0.005	**
Intercept	-0.281	0.004	**

The P values of both Healthy life expectancy and Generosity are greater than 0.05, which means there is no sufficient evidence to demonstrate that they have a significant impact on happiness score. Thus, these two variables were eliminated in the new model. The result of the reduced model can be seen in Table 4.

Table 4. Results of reduced MLR model

Variable	Estimate	P value	
Logged GDP per capita	0.254	0.000	***
Social support	4.199	0.000	***
Freedom to make life choices	2.332	0.000	***
Perceptions of corruption	-0.870	0.002	**
Intercept	-1.416	0.011	*

In the reduced model, all P values of the independent variables are less than 0.05, and there is reason to believe that they have a significant impact on a country's happiness score.

The random forest regression was also carried out using all six independent variables. Table 5 and Figure 2 show the measurements of the importance of the variables in terms of two different metrics.

Table 5. Measurements of importance

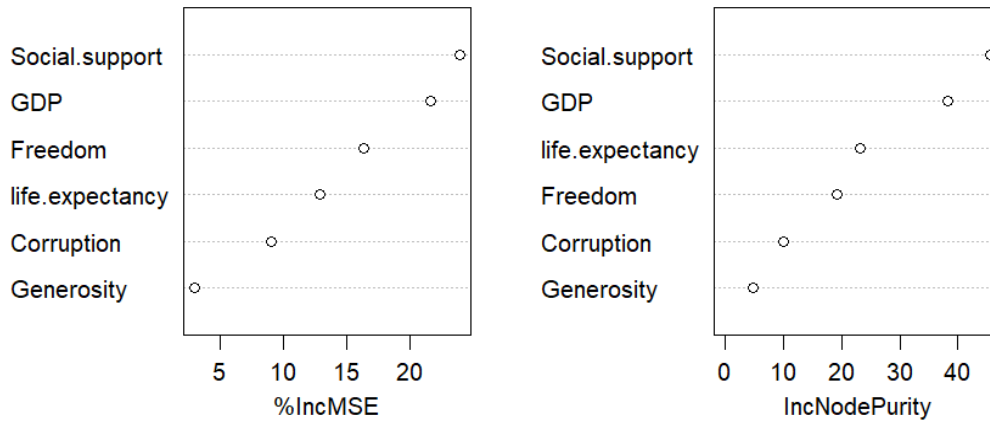| Variable	%IncMSE	IncNodePurity
Logged GDP per capita	21.634	38.126
Social support	23.929	45.483
Healthy life expectancy	12.888	23.238
Freedom to make life choices	16.401	19.209
Generosity	2.990	4.802
Perceptions of corruption	9.034	10.062

Figure 2. Measurements of importance

Both of these two metrics indicate that Social support and Logged GDP per capita are the first and second most important predictive variables respectively, while Generosity and Perceptions of corruption are of little importance.

Compared with the results before and during the pandemic of COVID-19[4,5], the importance of Healthy life expectancy decreased and the importance of Social support increased.

This discrepancy could be explained by a number of factors. After the COVID-19 pandemic, the importance of health to people's well-being should have increased, while the importance of life expectancy actually decreased. This could be due to the fact that Healthy life expectancy is a comprehensive indicator. People's pursuit of health is reflected more in Social support, such as the government's policy of providing healthcare to residents during the pandemic.

A recent study found that social isolation and quarantine had adverse effects on people's mental health and social support can act as a protection against those effects [7]. Ramkissoon's paper suggest that interventions in reinforcing and fostering social bonding are needed during the post pandemic era [8]. These research findings are consistent with the outstanding importance of Social support.

Based on the results of these models and the discussion above, a series of recommendations can be made for policymakers. Firstly, governments can enhance the level of social support not only from the classical perspective, such as promoting family harmony or fostering friendships but also by improving health care, which may significantly increase the level of social support perceived by the population. Reformations of the health care system are necessary if current contingency measures are not sufficient for such severe situations as COVID-19. Furthermore, the domestic economic situation continues to have a significant impact on people's well-being. How to improve economic conditions should still be one of the main concerns of governments and policymakers for countries in the process of recovering from the effects of the pandemic.

Furthermore, the three models discussed above were compared in terms of R-squared and RMSE (for both the training set and testing set). The results can be seen in Table 6.

Table 6. Comparison of the three models

Evaluation Metrics	MLR Full Model	MLR Reduced Model	Random Forest
R-Squared	0.828	0.825	0.770
RMSE Training	0.481	0.482	0.605
RMSE Testing	0.435	0.454	0.404

R-squared and RMSE for the training set both indicate that the linear model performs better than the random forest in regression modeling. RMSE for testing set of the random forest regression model

is only slightly lower than linear models, so there is not much difference between the predictive function.

4. Conclusion

This paper identified the statistical significance in the correlation between happiness score and the variable included in the World Happiness Report 2023. The full and reduced model using multiple linear regression shows that Generosity and Perceptions of corruption have no significant impact on a nation's well-being. This result is proved by the measurements of importance drawn from the random forest regression model. After referring to previous research, a series of recommendations can be made for policymakers based on the model result and variation of feature importance. In addition to the traditional approach, governments can raise social support levels through enhancing healthcare. If the current backup plans are insufficient to handle emergencies like COVID-19, reforms to the health care system are inevitable. Economic policies are always advantageous to the general well-being of a people, making them one of the priorities for policymakers. Among the three models in this paper, the two linear models are fitted better with the training data, and the prediction accuracy of all three models is similar.

The future work of this research is to apply more machine learning or even deep learning methods to get more insight into the driver factors of happiness after the pandemic. The data provided by the World Happiness Report 2024 or later should be used to further test the predictive function of the models in this paper.

References

[1]   Pandit, Vishwanath, and Vishwanath Pandit. "Prosperity and Happiness." Ethics, Economics and Social Institutions, 2016, 73-89.

[2]   Omer, Saad B., Preeti Malani, and Carlos Del Rio. "The COVID-19 pandemic in the US: a clinical update." 2020, 1767-1768.

[3]   Gupta, Meenu, et al. "Machine Learning-Based Comparative Analysis of COVID-19 Infected Cases with GDP and World Happiness Report." Micro-Electronics and Telecommunication Engineering: Proceedings of 6th ICMETE 2022. Singapore: Springer Nature Singapore, 2023. 345-355.

[4]   Khder, Moaiad Ahmad, Mohammad Adnan Sayfi, and Samah Wael Fujo. "Analysis of World Happiness Report Dataset Using Machine Learning Approaches." International Journal of Advances in Soft Computing & Its Applications 14.1, 2022.

[5]   Agarwal, Tarushi, et al. "Predicting Happiness Score During Covid-19 Using Machine Learning." 2022 4th International Conference on Artificial Intelligence and Speech Technology (AIST). IEEE, 2022.

[6]   Sihombing, Pardomuan Robinson, et al. "Comparison of Regression Analysis with Machine Learning Supervised Predictive Model Techniques." Jurnal Ekonomi dan Statistik Indonesia 3.2, 2023, 113-118.

[7]   Ferreira, Lara N., et al. "Quality of life under the COVID-19 quarantine." Quality of Life Research 30, 2021, 1389-1405.

[8]   Ramkissoon, Haywantee. "COVID-19 adaptive interventions: Implications for wellbeing and quality-of-life." Frontiers in Psychology 13, 2022, 810951.