Journal of Computer Science

Original Research Paper

Analyze and Predict the 2022 World Happiness Report Based on the Past Year's Dataset

Yifei Zhang

School of Physical Science, University of Liverpool, United Kingdom

Article history
Received: 20-09-2022
Revised: 06-02-2023
Accepted: 15-02-2023

Email: 2658527241@qq.com

Abstract: Through the impact of the COVID-19, people around the world have been affected to various degrees. Thus, it is more interesting to compare the happiness reported between 2022/2021 and before 2019. This article concludes 5 years of happiness scores data including family, Gross Domestic Product (GDP), health, freedom, generosity trust, and dystopia residual. Happiness scores are considered appropriate indicators to measure the progress of social development. This study presents two linear regression models to predict happiness scores across countries in 2022. Data is sourced from the world happiness report dataset from 2015-2021, available in open source. Preliminary exploratory data analysis was carried out to select the most appropriate variables to include in the models. The models' accuracy was tested by comparing the output values to the true 2022 world happiness report data. The experiment results show that the linear regression achieved a Root Mean Square Error (RMSE) = 0.236 and Mean Squared Error (MSE) = 0.056 for 2022.

Keywords: World Happiness Report, Linear Regression, Data Analysis, Machine Learning

Introduction

Measuring happiness is challenging due to the varied definitions of Happiness. The world has suffered from the global COVID-19 Pandemic for three years, producing unusual results in recent happiness measurements especially the 2021-2022 happiness scores, which represent pandemic struggles and are not indicative of previous global happiness trends.

Motivated by the happiness scores of this author's homelands, the author sought to understand how happiness is quantified and how the UN measures happiness, and what countries have a higher score.

Throughout history, the first scientist to try to measure happiness as a mathematical quantity was Bhutan in the 1970s, who calculated his country's Gross Happiness National (GHN) instead of Gross National Product (GNP) and the key core of his calculating that Bhutan believed that the spirit and material improvement are both important for the development of the human society (Bhattacharyya *et al*., 2019). Then several international organizations later build their own happiness index like U.N the world happiness report.

The first world happiness report was published in 2012 by the United Nations, which compared the happiness of people in 156 countries and regions around the world (Helliwell *et al*., 2016). According to the world happiness official website, there are six variables: Economic production Gross Domestic Product, or (GDP); social support; life expectancy; freedom; absence of corruption; and generosity, all of which contribute to each country's Happiness Score also referred to as a ladder score.

The world happiness report is essential for both governments and the public as it provides information that is difficult to quantify in the real world. For the government, happiness scores can provide evidence of their population's well-being. For the civilization, it provides an understanding of their actual life quality in many aspects including health care, work environment, and education (Helliwell *et al*., 2021). Hence it is necessary to study and analyze the data.

In the work of (Jannani *et al*., 2021), the authors have used nine different ways to predict the quality of life through the world happiness Index including lasso regression, multiple linear regression, LSTM, random forest regressor, support vector regression, gradient boosting regressor, XGboost Regressor, MLP regressor, and AdaBoost regressor. The performance to evaluate the models is MAE, MSE, RMSE, and R^2 (the definition of which, is in section 4). The data from 2015-2021 has been collected and divided for training (2015-2020) and testing (2021). Their results show that the best performance is

achieved using the lasso regression with a 0.8954 R^2 score and 0.0656 RMSE.

However, the above paper is mainly focused on the overall happiness scores but not on how the variables such as GDP and health influence the world happiness scores. It is more interesting to find out which variables will influence the world happiness score more rapidly and how they will influence it.

The author aims to analyze the data about the world happiness report from 2021-2022, the distribution of the region for these countries, and the relationship between different variables on the ladder score. By using linear regression, this study will predict ladder scores using each country's GDP and provide the best fit linear equation. These analyses will be carried out in Python.

In the next section, the author will explore some other authors' work applying various machine learning algorithms on predicting the data. In the third section, the author will give some basic background information to help the reader understand the world happiness report and analyze the 2021-2022 world happiness report. In the fourth section, the author will explain the methodology which will be used in the following section. In the fifth section, the author discusses how to build the model and shows the result in Python. In the sixth section, the result is analyzed and discussed. Finally, the conclusion and next research scopes are developed in the last section.

In research presented by Khder *et al*. (2022), the authors identified the essential issues in using data from the importance of the variables included in the dataset. The research uses various machine learning methods such as Neural Network (NN), Random Forest (RF), and XGboost (GB) to classify the GDP as one of the primary indicators of life happiness scores. Also, the insight gained from the study is that high life expectancy may lead to a higher happiness score by classifying as the first rule, while the use of one classification method and its results by evaluating different performance indicators increases the finding of reinforcement.

Researchers (Sandvik *et al*., 1993) found that it is different for a human being to get the value of happiness from family and friends. The research team found that there is a discrepancy between the three evaluations that must take into account.

Bullen (2022) used machine learning methods to divide the UN world happiness dataset into a training set and test set, which used the K-Means cluster to reach.

Millard (2011) uses various approaches based on the scope of machine learning to analyze global happiness. The Principal Component Analysis (PCA) was used to analyze gender equality and life satisfaction. In the feature, selection trees were used as well as for life satisfaction prediction. The findings of this study are key characteristics of life expectancy, incoming distribution, and freedom summarized using permutation tests. The results show happiness in life in the form of a visual map.

Tan *et al*. (2020) aimed to find the most accurate model to predict ladder scores based on the dataset for the top 30 countries and regions in the world. The traditional linear regression model has been used to show that it is risky to be used because of the correlation between various explanatory variables. Ridge regression lasso regression and elastic net based on machine learning are applied to get an accurate prediction and the elastic net model has been found to be the most accurate model in this study.

Khemraj *et al*. (2021) mainly focus on three methods: Artificial Neuron Network (ANN), Support Vector Machine (SVM), and Regression Tree (RT). The most significant prediction method is the ANN # 3 6-20-1 model which has an accuracy of 83.68% and the significant test for SVM2 is found as R 0.15 and RMSE 0.5454. The regression of the Tree2 model has significant testing of error RMSE 0.57815 that the lowest and enclosed 0. The author has indicated that SVM is simpler to apply than ANN but the first choice of option for the prediction of the world happiness score is ANN.

Salma *et al*. (2020) studies various factors to determine their importance in GDP growth and develops a forecasting model to forecast the future using the gaussian process, decision table, random tree, multilayer perception, and random tree and achieved an MAE of 1.801% using linear regression.

It can be concluded that research about world happiness prediction applied various machine learning methods to analyze the results. The most used ML algorithms are LR, SVM, RF, and NN such as LR is used in the context. However, each paper is based on a unique dataset and applies different methods to predict which might need a high-level knowledge about machine learning, this study is focused on the basic linear regression methods in machine learning to help beginners in machine learning studies to analyze the happiness score data.

Before proceeding with the analysis of the dataset, the author provides some brief intuition to the reader, to understand the world happiness report for 2021-2022. Figure 1, it has shown the happiness score 2022 for the countries and regions in the dataset which has been used through a world map.

Figure 2, it has shown the happiness score 2021 for the countries and regions in the data set which has been used through a world map.

It seems like no countries and regions have dramatic changes between 2021-2022. The happiness scores are distributed to similar scores in different states. Figure 3 details the different regions on the happiness scores.

Next, the author explores the dataset to search for drastic changes between these two years, shown in Fig. 4.

Figure 4, what can be concluded is that the basic distribution line is not changed in a dramatic way. The mean happiness score for 2022 is higher than the 2021 happiness score. The score intersection between 5-7 has a noticeable increase.

DOI: 10.3844/jcssp.2023.483.492

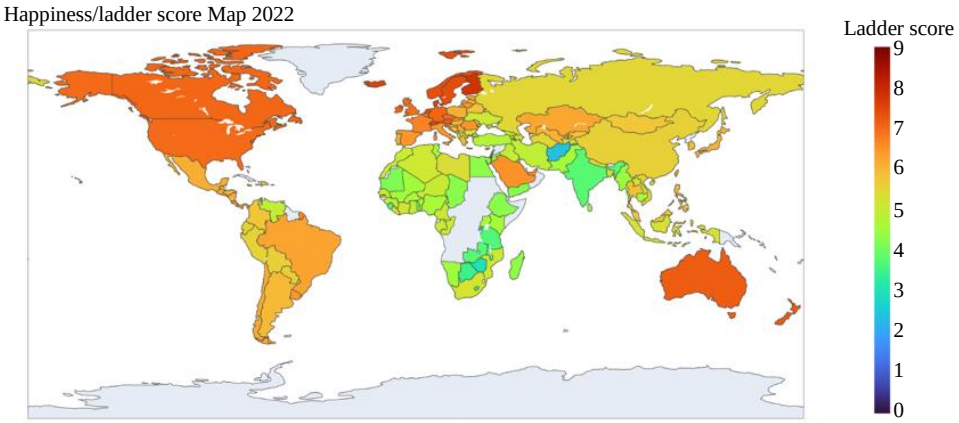

Fig. 1: 2022 happiness score distribution

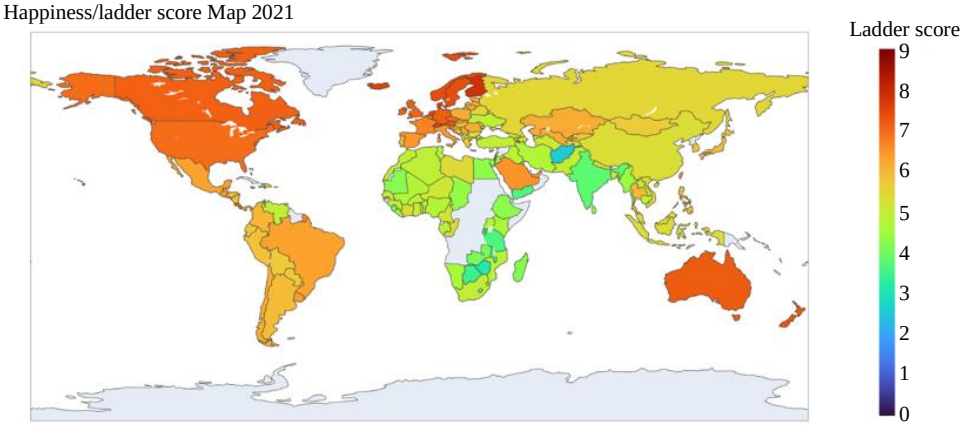

Fig. 2: 2021 happiness score distribution

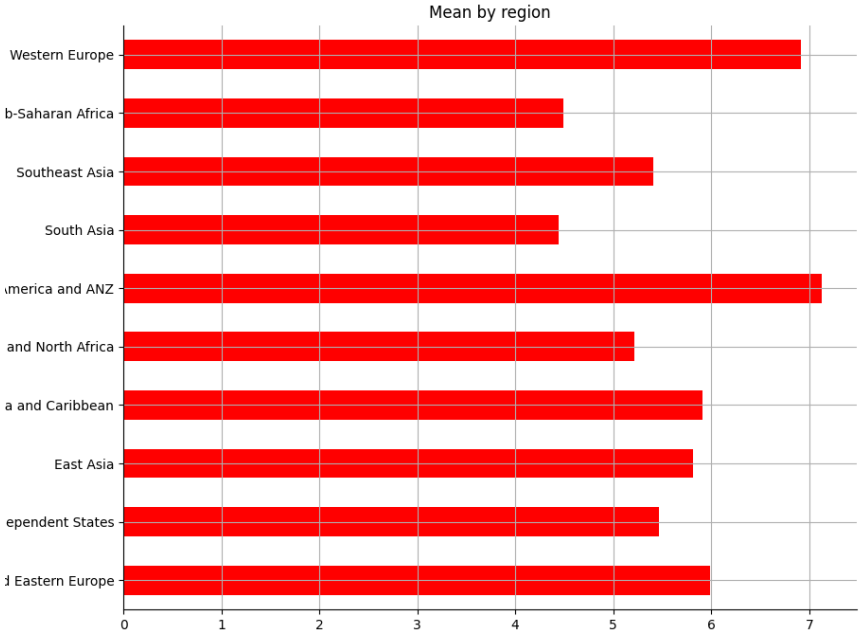

Fig. 3: Happiness score for different regions

DOI: 10.3844/jcssp.2023.483.492

For the 6 variables economic production (GDP), social support, life expectancy, freedom, absence of corruption, and generosity, the next step is to figure out which will influence the ladder scores more rapidly. Thus, the next step is constructing a heat map to get a clear view of the relationship between these 6 variables and the ladder score which will be shown in Fig. 6.

Figure 5 has shown, the most essential variable that will influence the happiness score is the economy (GDP per capita) and the Family and it has the highest correlation with happiness score which is 0.78 and 0.76. The least important variable is generosity which only has a correlation with a happiness score of 0.18. To have a clearer view, the next step is constructing the plot diagram (Fig. 6) for every two variables.

So, the author chooses the economy (GDP per capita) to be the key variable to predict happiness scores.

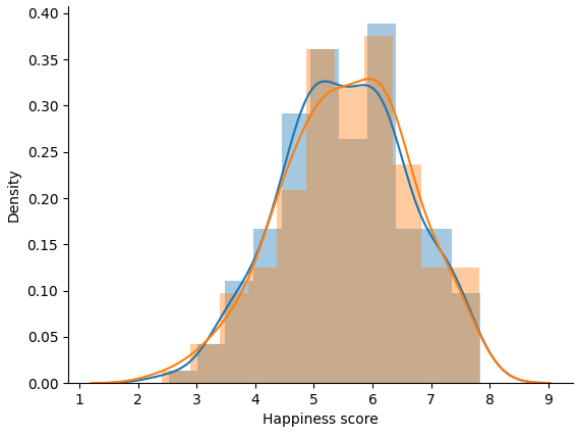

Fig. 4: Density distribution for happiness score between 2021-2022, with blue representing 2021 and orange 2022

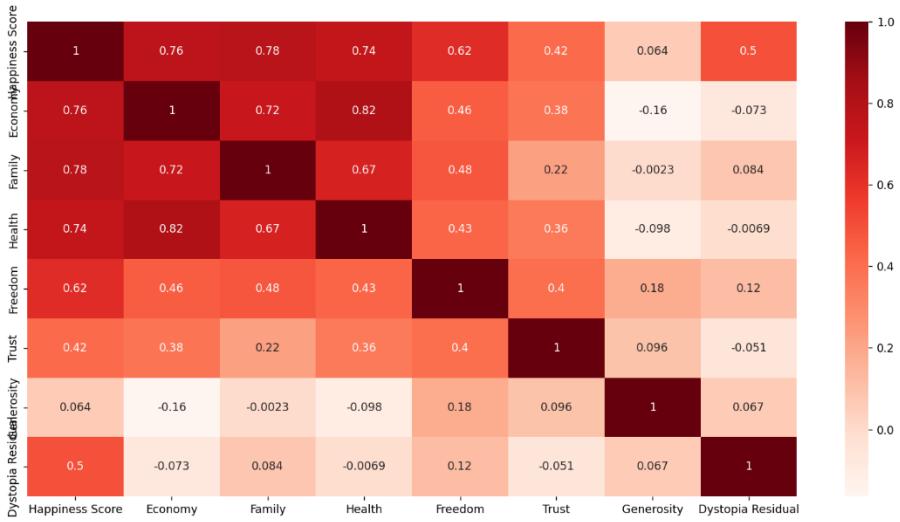

Fig. 5: The heat map of each parameter's correlations

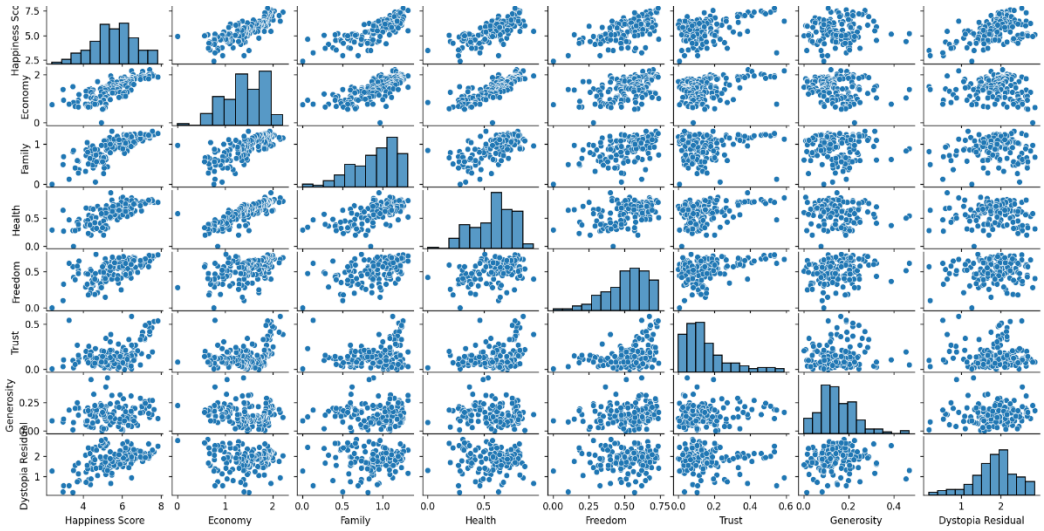

Fig. 6: Scatter diagram of the distribution between each parameter

Materials and Methods

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the save as command. In this newly created file, highlight all of the contents and import your prepared text file you are now ready to style your paper.

Data Availability

The database for this research is the Kaggle world happiness report (Bouzid Ait-Amir and Pougnet, 2020) which is an essential tool to analyze global happiness. This dataset ranks 155 countries and regions in their ladder scores. According to the different survey respondents, their happiness scores were rated on a scale from 0-10 marks to show their best possible life. Besides the dataset includes several insights from variables that will influence the scores.

Describing the Dataset

The dataset features rely on daily life experiences from respondents based on their considerations of the most persuasive lives and the worst lives. These features are described in detail below.

This article will use the world happiness report 2021 as an example to show the top 10 countries' specific data in Table 1 (which will be shown in the last).

The machine learning algorithm that has been used to predict the data is linear regression. Various python packages were used to accomplish different purposes during the research.

NumPy Harris *et al*. (2020) (numerical python) is an extensive library of the python language, which supports a large number of the dimensional array and matrix operations and also provides a large number of mathematical functions libraries for array operations.

Pandas McKinney (2010) is an open source, BSD-licensed library that provides high performance, easy-to-use data structures, and data analysis tools.

Sklearn Pedregosa *et al*. (2011) (sci-kit-learn) is a powerful machine learning library that covers everything from data preprocessing to model training.

Seaborn Waskom (2021) is a data visualization library based on Matplotlib. It builds on Matplotlib with a higher level of API encapsulation to make drawing easier and refined without a lot of tweaking.

In order to evaluate the performance of these models several metrics have been used, which are:

- Mean Absolute Error (MAE) is a measure of errors that corresponds to standard L1 and measures the average of the absolute difference between the actual and predicted values (Bouzid Ait-Amir and Pougnet, 2020)
- Mean Squared Error (MSE) is a criterion that measures the mean square error of the mismatch between predicted and real values (Bouzid Ait-Amir and Pougnet, 2020)
- Root Mean Square Error (RMSE) is the square root of mean squared error (Bouzid Ait-Amirf and Pougnet, 2020)

The author used these metrics to evaluate the proposed models. They are defined as follows:

$$MAE = \frac{1}{m}\sum_{i=1}^{m}| y_i - y_i^{'} |^2$$

$$MSE = \frac{1}{m}\sum_{i=1}^{m}\left(y_i - y_i^{'}\right)^2$$

$$RMSE = \sqrt{MSE}$$

where, m represents the total number of elements of the test data, $y_i^{'}$ is the predicted value and y_i is the corresponding true value of the i^{th} sample.

For the linear regression, the author set our target equation to be $y_i = \beta x_i + \alpha$ where y_i is the GDP value and x_i is the happiness score of the i^{th} sample.

Table 1: Top 10 countries' specific data (world happiness report 2021)

Rank	Country	Happiness score	Dystopia residual	Economy (GDP per capita)	Family	Health (life expectancy)	Freedom	Generosity	Trust (government corruption)
1	Finland	7.821	2.518	1.892	1.258	0.775	0.736	0.109	0.534
2	Denmark	7.636	2.226	1.953	1.243	0.777	0.719	0.188	0.532
3	Iceland	7.557	2.320	1.936	1.320	0.803	0.718	0.270	0.191
4	Switzerland	7.512	2.153	2.026	1.226	0.822	0.677	0.147	0.461
5	Netherlands	7.415	2.137	1.945	1.206	0.787	0.651	0.271	0.419
6	Luxembourg*	7.404	2.042	2.209	1.155	0.790	0.700	0.120	0.388
7	Sweden	7.384	2.003	1.920	1.204	0.803	0.724	0.218	0.512
8	Norway	7.365	1.925	1.997	1.239	0.786	0.728	0.217	0.474
9	Israel	7.364	2.634	1.826	1.221	0.818	0.568	0.155	0.143
10	New Zealand	7.200	1.954	1.852	1.235	0.752	0.680	0.245	0.483

Results

Through the data from Kaggle, the author collected 6 years of data which are the World Happiness Report from 2016-2021 (WHR, 2021).

The total number of data is 1084 and will be divided into 3 groups: The training set, the testing set, comparing test. The author selected 75% data as the training set and the remaining 15% for testing and 10% for comparison.

The resulting model is shown in Fig. 7.

Through our training, the author obtains the results:

$$\alpha = 0.28465782843416604$$
$$\beta = -0.6227407645545371$$

The target equation will be:

$$y = 0.28465782843416604$$
$$x - 0.6227407645545371$$

From 75% of the data, we fit a line that best describes the data and in order to test the level of fitness 15% of the data is used to test the given target equation.

Subsequently, the next 15% of the data, used for testing, is shown (the green points) in Fig. 8. It can be verified that the test data is in fact evenly distributed on both sides of the target line.

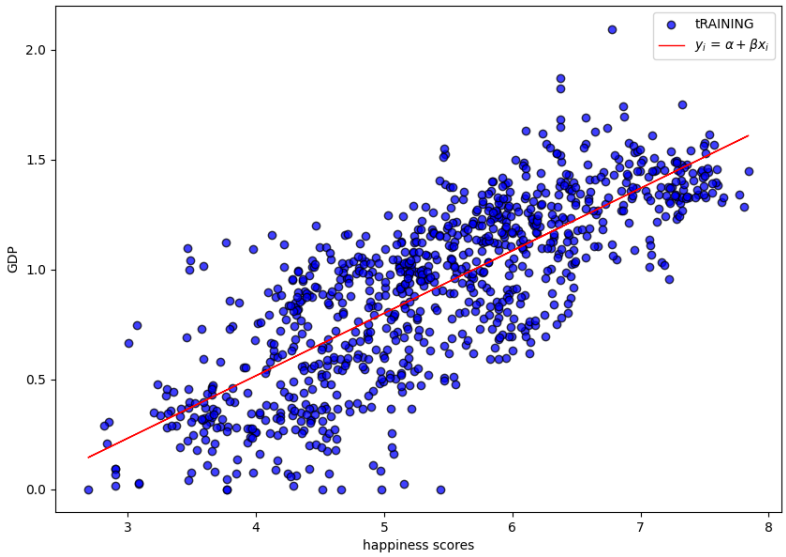

Fig. 7: Distribution of the training model

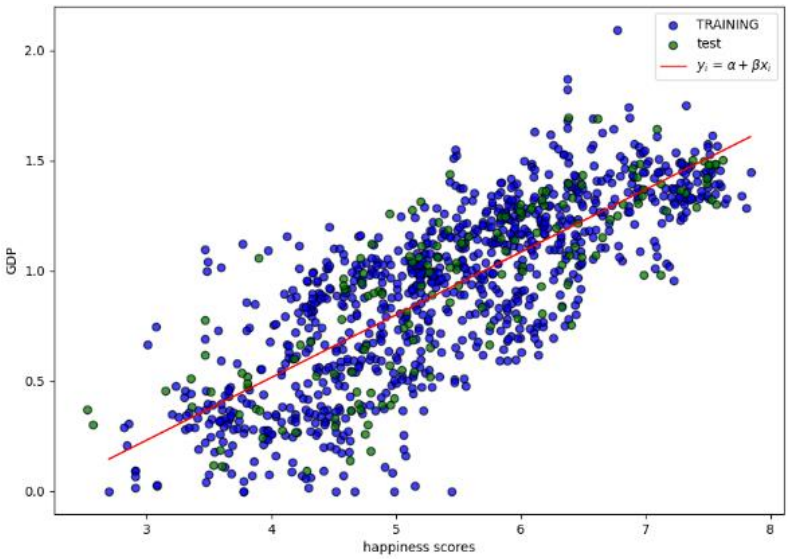

Fig. 8: The training and test data distribution

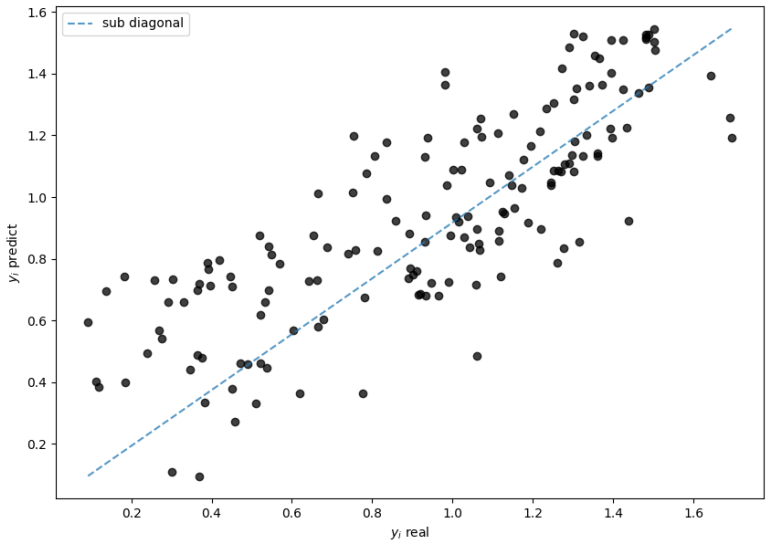

Fig. 9: Testing the model with a comparison between predicted and actual values for 2022 happiness scores

Finally, the author set the remaining 10% of the dataset to be used to predict the happiness score through the target line. The results of the validation are shown in Fig. 9, which shows that the data is appropriately fitted by the linear regression model which has been generated.

After applying the algorithms to the data, three performance equations have been chosen which were mentioned earlier to evaluate our model. The results obtained are the following:

$$MAE : 0.1954020331$$
$$MSE : 0.0562180722$$
$$RMSE : 0.236332901$$

Discussion

From the above result, *MAE*, *RMSE*, and *MSE* are all expected to be as small as possible. Khemraj *et al*. (2021), the author gets the other authors' results about the *RMSE*, *MSE*, and *MAE* which have been cited above which is *RMSE* = 0.5454.

By comparing these two models, our model's *RMSE* is smaller than the ANN model (Khemraj *et al*., 2021).

The results the author obtain *MAE* = 0.195 *MSE* = 0.056 and *RMSE* = 0.236 is in a satisfactory range.

However, it remains uncertain whether GDP is the best variable to predict the happiness score. From the heat map in Fig. 5, it can be observed that the variable of family also displays a high correlation to happiness score. Thus, the author constructs another linear regression model utilizing family as the predictive feature and assesses how this model performs in comparison to the one made using GDP.

Figure 10 shows the distribution between happiness scores and family. The gradient of this target line is positive. Thus, with the increase of the family score, the happiness score should increase with the family score which is the same as the GDP. However, the distribution for the family score exhibits more dispersion than the GDP.

The gradient of the linear regression model is α = 0.18470798 and the y-intersection is equal to β = 0.53170677 which will give the equation of the target line:

$$y_i = (0.18470798)x_i + 0.53170677$$

Subsequently, 15% of the data will be used to test the model. The green points will be the test points and give clear information that the test points have evenly distributed on both sides of the line which is also similar to the results of GDP which will be shown in Fig. 11.

Last but not the least, from python the author got *RMSE*, *MSE*, and *MAE* to evaluate the accuracy of the prediction, obtaining the following results:

$$MAE = 0.2140442591$$
$$MSE = 0.0664012612$$
$$RMSE = 0.576844217$$

Compared to the numbers above in the paper, three numbers are all increased which means the prediction for the happiness score is not as accurate as the prediction by GDP. In that way, using GDP gives us the best variable to predict happiness scores.

During the research, the key question is to determine which of the six variables influences the happiness scores most or which of the parameter can predict the happiness score most accurately. Through the methodology, heat maps are drawn to show the correlation between each parameter and the happiness score. Two parameters that

have a higher correlation coefficient will be considered as the target parameters. However, a higher correlation coefficient doesn't mean having the ability to predict the happiness score better. To our results, the family variable has the highest correlation coefficient of 0.78. However, when utilizing this variable to build a linear regression model, the *RMSE* for the family model is greater than the *RMSE* for the GDP.

In conclusion, GDP is the best parameter to predict the happiness score when using a linear regression model.

However, the limitation of linear regression is also obvious. For a given target line, if one country's GDP is greater than another country our model will predict its happiness score must be greater than another country which is not a certain result. Because there are another 5 variables to determine the happiness score it might have some special circumstances for example Singapore and Hong Kong. In 2022, Singapore and Hong Kong are third and ninth rank in GDP but only get twenty-seven and eighty-one in happiness rank. Their dystopia residuals are not high enough to support them to get higher marks though the GDP has contributed a lot.

Thus, a machine learning method that includes more variables should be considered for further research to get a more accurate result.

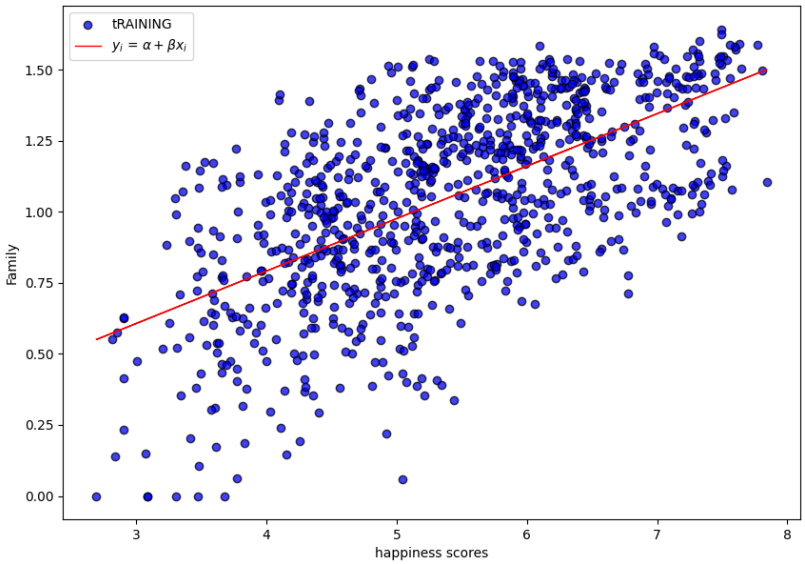

Fig. 10: The training data distribution for family score

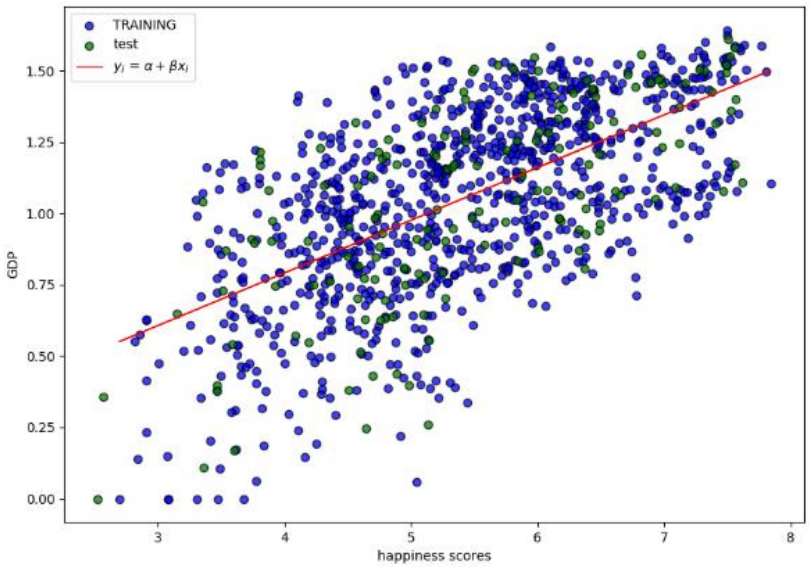

Fig. 11: Test points distribution for test points

Conclusion

Happiness research remains a challenge for researchers concerned with this aspect of societal development. The use of data science tools to model and analyze happiness predictions can be very useful in addressing the challenges associated with this subject matter and aid in future research.

The limitations of using machine learning methods to predict happiness scores also exist. A number of outstanding events have not been taken into account during the process of the predictions, which will undoubtedly have a significant influence on the happiness scores of specific countries.

This article delved into the literature related to this concept from a data science perspective, where machine learning and deep learning algorithms are used. In this study, a brief analysis of the comparison of world happiness reports between 2021-2022 is given. The happiness scores do not have a dramatic change between these two years. An experiment has been approached to explore the potential of linear regression models in predicting happiness scores and then compared their performances. The performance is achieved with an *MAE*: 0.195, *MSE*: 0.056, and *RMSE*: 0.236. A comparison group is constructed by using another important factor Family to predict the happiness score. The result gives that *MAE*: 0.214, *MSE*: 0.066, *RMSE*: 0.258. All three metrics of error are larger in the Family model compared to the GDP model, which asserts that GDP is the best variable to predict the happiness score in a linear regression model.

Using these techniques, readers can study, predict, and model happiness accurately by discovering the most highly correlated variable: GDP.

For future research, additional types of models to predict happiness scores such as ANN and SVM can be used. For further study, the research team can combine the other variables like social support, life expectancy, freedom, absence of corruption, and generosity, to contribute to the prediction more accuracy by adding more layers to reduce the dimension which can be based on the research (Ruggeri *et al*., 2020), the researchers considered the multiplies variables.

Furthermore, more complex linear models can be used to describe the data such as multiple linear regression (Uyanık and Güler, 2013), support vector regression (Gauraha, 2018), and lasso regression (Crone *et al*., 2006).

Acknowledgment

Throughout the writing of this study, I have received a great deal of support and assistance.

I would first like to thank my supervisor, Aida Manzano Kharman, a Ph.D. student at Imperial College London, whose expertise was invaluable in formulating the research questions and methodology. Your insightful feedback pushed me to sharpen my thinking and brought my work to a higher level.

I would also like to thank my tutors, Tanvika Parlikar, and Chloé-Rose Colombero, for their valuable guidance throughout my studies. You provided me with the tools that I needed to choose the right direction and successfully complete my dissertation.

In addition, I would like to thank my parents for their wise counsel and sympathetic ear. You are always there for me.

Finally, I could not have completed this dissertation without the support of my friends, Haoming Guo, and Shiyuan Zhang who provided stimulating discussions as well as happy distractions to rest my mind outside of my research.

Funding Information

The authors have not received any financial support or funding to report.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

Bhattacharyya, S., Burman, R. R., & Paul, S. (2019). The concept of measuring happiness and how India can go the Nordic Way. *Current Science*, *116*(1), 26-28. https://www.jstor.org/stable/27137783

Bouzid Ait-Amir, H., & Pougnet, P. (2020). "6 - Meta-Model Development", *Embed. Mechatron. Syst*., vol. *2*, pp. 157-187, 2020.

Bullen, E. (2022). "Ed Bullen. https://www.nitj.ac.in/nitj_files/links/SEHCM_Final_Proceeding_compressed_52305.pdf

Crone, S. F., Guajardo, J., & Weber, R. (2006). A study on the ability of support vector regression and neural networks to forecast basic time series patterns. In *Artificial Intelligence in Theory and Practice: IFIP 19th World Computer Congress, TC 12: IFIP AI 2006 Stream, August 21-24, 2006, Santiago, Chile 1* (pp. 149-158). Springer US. https://doi.org/10.1007/978-0-387-34747-9_16

Gauraha, N. (2018). Introduction to the LASSO. Reson, 23, 439-464. https://doi.org/10.1007/s12045-018-0635-x

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... & Berg, S. (2020). Smith 474 nj. *Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, del R'ıo JF, Wiebe M, Peterson P, G'erard-475 Marchant P, et al. Array programming with NumPy. Nature*, *585*(7825), 357-362.

Helliwell, J. F., Huang, H., Wang, S., & Norton, M. (2021). World happiness, trust and deaths under COVID-19. *World happiness report*, *2021*, 13-57.

Helliwell, J. F., Layard, R., & Sachs, J. D. (2016). *World happiness report 2016-update* (Vol. *1*). Sustainable Development Solutions Network. https://apo.org.au/sites/default/files/resource-files/2016-03/apo-nid226706.pdf

Jannani, A., Sael, N., & Benabbou, F. (2021, December). Predicting Quality of Life using Machine Learning: Case of World Happiness Index. In *2021 4th International Symposium on Advanced Electrical and Communication Technologies (ISAECT)* (pp. 1-6). IEEE. https://ieeexplore.ieee.org/abstract/document/9668429

Khder, M. A., Sayfi, M. A., & Fujo, S. W. (2022). Analysis of World Happiness Report Dataset Using Machine Learning Approaches. *International Journal of Advances in Soft Computing & Its Applications*, *14*(1). http://www.i-csrs.org/Volumes/ijasca/2022.1.2.pdf

Khemraj, S., Thepa, P. C. A., Chi, H., Wu, W. Y., Samanta, S., & Prakash, J. (2021). Prediction of world happiness scenario effective in the period of COVID-19 pandemic, by artificial neuron network (ANN), support vector machine (SVM), and regression tree (RT). *Nveo Natural Volatiles & Essential Oils Journal| NVEO*, 13944-13959. http://www.nveo.org/index.php/journal/article/view/2995

McKinney, W. (2010, June). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference* (Vol. *445*, No. 1, pp. 51-56). https://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf

Millard, L. (2011). "Data Mining and Analysis of Global Happiness: A Machine Learning Approach," University of Bristol: Department of Computer Science, Bristol, 2011.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *the Journal of Machine Learning Research*, *12*, 2825-2830. https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?ref=https:/

Ruggeri, K., Garcia-Garzon, E., Maguire, Á., Matz, S., & Huppert, F. A. (2020). Well-being is more than happiness and life satisfaction: a multidimensional analysis of 21 countries. *Health and Quality of life Outcomes*, *18*(1), 1-16. https://doi.org/10.1186/s12955-020-01423-y

Salma, U., Hasan, M. M., & Sultana, M. I. (2020, February). A regression model-based approach to identifying determining factors for GDP growth in Bangladesh. In *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)* (pp. 1-6). IEEE. https://ieeexplore.ieee.org/abstract/document/9077851

Sandvik, E., Diener, E., & Seidlitz, L. (1993). Subjective well-being: The convergence and stability of self-report and non-self-report measures. *Journal of Personality*, *61*(3), 317-342. https://doi.org/10.1111/j.1467-6494.1993.tb00283.x

Tan, Y., Singhapreecha, C., & Yamaka, W. (2020). Applying Machine Learning to Predict Happiness: A case study of 20 Countries. *Mukht Shabd Journal*, *9*(6), 3433-3437.

Uyanık, G. K., & Güler, N. (2013). A study on multiple linear regression analysis. *Procedia-Social and Behavioral Sciences*, *106*, 234-240. https://doi.org/10.1016/j.sbspro.2013.12.027

Waskom, M. L., (2021). seaborn: statistical data visualization. *Journal of Open-Source Software*, *6*(60), 3021, https://doi.org/10.21105/joss.03021

WHR. (2021). Kaggle. https://www.kaggle.com/datasets/ajaypalsinghlo/world-happiness-report-2021