

Explorations on the Levels of Happiness across Countries

02433236

Compiled: May 13, 2024

Github Repo: TAGGED-RELEASE-02433236

1 Project Description (approx. 250 words)

The project researches the levels of happiness across various countries from 2022 to 2024, using datasets sourced from the World Happiness Report in collaboration with Gallup Poll. The main objective of this report is to figure out the relationship between the happiness score of different countries and six social and economic factors (GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, perceptions of corruption). It is started with an exploratory data analysis to identify the correlations between variables. As there can be some features showing minimal correlation with the happiness scores, the first model used is the Lasso Regression by adding a regularization term. To further evaluate the predicting results, the Random Forest model and the K-Nearest Neighbours model are also implemented in the report. After tuning the hyperparameters of different models, the KNN model shows a much higher accuracy in predictions than other two models. It can be learned from the report that the social support offered to the population and the extent of freedom that people have in their daily life are highly related to the happiness scores of a country. Additionally, the entire happiness levels of each region show small fluctuations from year to year and remain at a relatively stable range. The predictions of the happiness scores can be vital for policy-making and improving the life quality of people. It also establishes a basis for future findings with more features.

2 Assessment Criteria

Technical Competence: Proficiency in data collection, processing, analysis, and coding.

- Data collection and process: The datasets are collected from official website and are preprocessed by deleting missing values and aligning the data structures over three years.
- Coding Competence: Each plot is altered with ggplot and effective model packages are used for my three models.

User Interface: Design, functionality, and usability of the final data product.

- Usability: It allows users to find useful information through visualisations and find happiness levels across regions.
- Design: The data product is clearly and logically organized offering different audience to find their target information quickly.

Analysis and Interpretation: Depth of analysis, appropriate use of statistical methods, and meaningful interpretation.

- Model techniques: Lasso Regression, Random Forest and KNN are implemented and hyperparameters of models are tuned to find the best method.
- Meaningful interpretations: It gives insight that the social support and the freedom to make life choices are highly correlated to the happiness levels.

Presentation and Communication: Clarity, organisation and effectiveness of written and visual communication.

- Well-structured: The data report is presented with introduction, data analysis, statistical method implementations and conclusion and future work in order.
- Effective visualisations: This report uses effective and suitable figures for results analysis, such as correlation plots, histograms and scatterplots that makes the results much interpretable.

Reproducibility: Clarity and completeness of documentation for result reproducibility.

- Documentation: README file includes the detailed documentation.
- Code: Codes are made available in a public GitHub repository including explanations of each block of code.

Version Control: Effective use of version control systems.

- Properly using tags and informative title for releasing in the GitHub platform
- The files are well-organized and split, making the project reproducible.

3 Project Reflection

Reflect on the experience of creating your data product. In 6 bullet points and at most 1 page total, summarise the following.

- *3 things you have learned as part of this process,*
- *2 aspects of the project that you found challenging or would approach differently with hindsight,*
- *1 aspect of the project that you would like to learn more about in the future.*

You may delete this italicised text when filling in the template.

Learnings:

- The data science project requires careful planning, structuring in both code and results presentations. This improves my data collection techniques and project managing skill.
- Analysing the model results and making right statistical choices are important to prediction tasks and further give appropriate recommendations to audience. It enhances my statistical thinking and interpretation ability of statistical results.
- I get more familiar with the usage of GitHub in conjunction with handling data science tasks, enabling me to work collaboratively and research on a project simultaneously with multiple team members.

Challenges:

- There are various number of countries each year and the prediction for several countries has to be deleted with incomplete dataset. In addition to the six main features including in my report, it can be really hard to collect other social, educational or geographical features for all countries.
- The correlation matrix show multicollinearity in various features and it can be challenging to handle multicollinearity and find the suitable model.

Further Development:

- For further study, we can improve the understanding of the happiness levels by including more features like geographic positions and the medical treatment and resources available to people. Additionally, there can be some interactions between variables, a more complex feature map $\phi(\mathbf{X})$ can be used for more accurate predictions.