STUDY OF WORLD HAPPINESS REPORT DATASET USING MACHINE LEARNING METHODOLOGIES

FINAL REVIEW

MAHALAKSHMI A (2 3MDT0052)

KAMALESH . K (2 3MDT0056)

HARIHARAN.R .B (2 3MDT0086)

U N D E R TH E G U D IA N C E O F

Dr. KARTHIKEYAN P

TABLE OF CONTENTS

INTRODUCTION

The World Happiness Report, an illustrious annual publication, is highly regarded for its role in providing insights into the happiness levels of nations across the globe. In today's corporate landscape, organizations are increasingly recognizing the profound impact that employee well-being has on overall performance, job satisfaction, and ultimately, the success of their enterprises. Machine learning, including techniques like Simple Linear Regression, Random Forest Regression, and Decision Tree Regression, has emerged as a powerful tool in understanding the intricate link between contentment and workplace efficiency. By employing Machine Learning models, organizations can gain deeper insights into the factors that contribute to employee well-being and, in turn, make more informed decisions to enhance workplace satisfaction. The realization that happy and content employees tend to be more engaged, motivated, and productive has spurred organizations to turn to sources like the World Happiness Report for guidance and leverage data-driven techniques such as Machine Learning to optimize their strategies for fostering a positive work environment and, consequently, improving their overall performance and success.

OBJECTIVE

The World Happiness Report aims to measure and understand subjective well-being and happiness across countries and cultures. Our objective is to uncover the factors that contribute to happiness in different countries.

- **Prediction of Ladder Scores:** Develop machine learning models to predict Ladder scores based on various features and variables present in the World Happiness Report dataset.
- **Identification of Outliers:** Use machine learning algorithms to detect outliers or anomalies in Ladder scores.
- **Cross-Country Comparison:** Explore machine learning techniques to compare and contrast happiness determinants across different countries, highlighting regional variations and similarities.
- **Policy Implications:** Translate the findings of the machine learning analysis into actionable policy recommendations, providing guidance for policymakers on interventions that may positively impact happiness scores.

ARCHITECTURE DIAGRAM

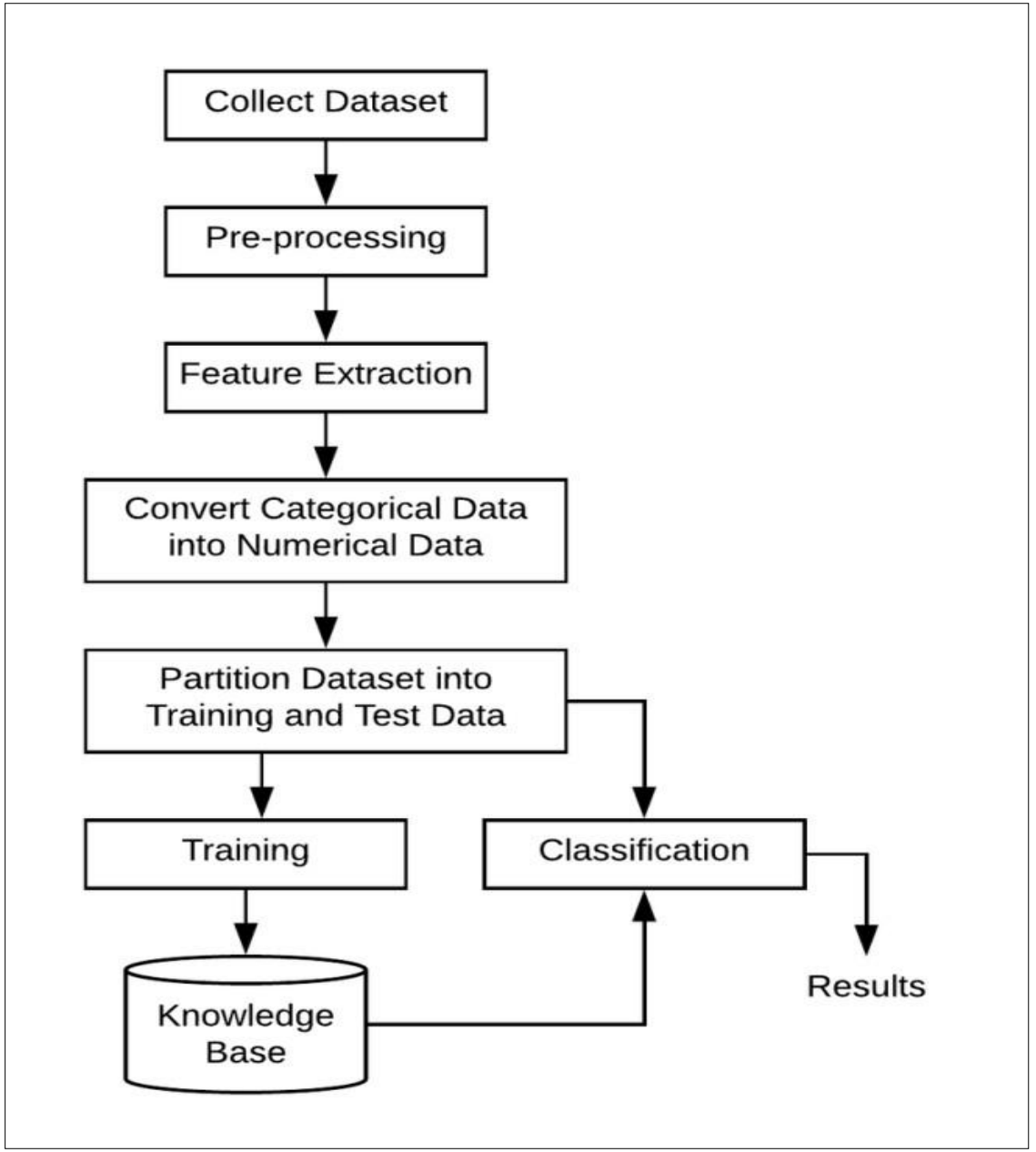

Fig 1 : Architecture Diagram

METHODOLOGY

We have collected data, including metrics like ladder score, upper whisker, lower whisker, logged GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, perceptions of corruption, dystopia, residual from sources like Kaggle to examine factors that contribute to happiness of people in each country.

We must undergo several essential steps in each research study, like collecting data, pre-processing data, selecting a suitable model, implementing it, calculating errors, and creating results.

PROCESSING TECHNIQUES

- **Data Collection**: Gather a comprehensive dataset that includes relevant variables related to happiness from the World Happiness Report and other reputable sources. Ensure the dataset covers a diverse range of countries, demographics, and happiness indicators to capture a representative sample. We used Secondary data from sources like Kaggle to examine people's happiness data. It is a common and valuable approach for research and analysis.

- **Data Preparation**: Missing values, eliminate duplicates, and deal with outliers to clean up the dataset. Perform necessary data transformations, such as normalization or feature scaling, to ensure consistency and comparability across variables.

PROCESSING TECHNIQUES

- **Data Visualization :** The pre-processing of data in order to visualize the World Happiness Report. Among them are cleaning techniques for outliers, inconsistent data, and missing values. For missing values, imputation techniques can be used, and outliers can be corrected. Scaling, also known as normalization, guarantees equitable comparisons across variables with disparate scales. For compatibility, categorical variables might need to be converted to numerical data as the ladder score is in numeric. To produce precise and insightful visualizations of the various happiness indicators included in the World Happiness Report.

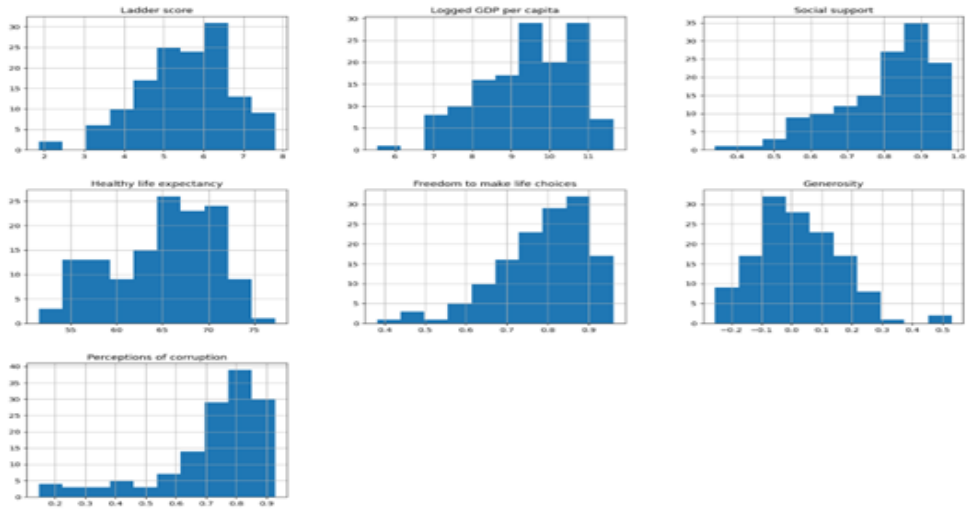

Fig 2 : Univariate analysis visualizations (Histogram)

PROCESSING TECHNIQUES

- **Feature Engineering:** Extract and select meaningful features from the dataset that influence happiness scores. This may include economic indicators, social factors, environmental variables, and cultural dimensions. Consider incorporating additional derived features or aggregating indicators to enhance the predictive power of the model.

- **Model Selection :** Experiment with various machine learning algorithms suitable for regression tasks, such as linear regression, random forests, gradient boosting, or ridge and lasso. Explore different algorithms to identify the one that yields the best performance in predicting happiness scores.

PROCESSING TECHNIQUES

❖ **Training And Evaluation :** Train the chosen models using the training data after dividing the dataset into training and testing datasets using train_test_split function. Assess the models' ability to predict happiness scores by utilizing suitable assessment metrics, such as R-squared or mean squared error. Cross-validation can be used to verify the robustness of the model.

```
# train test split
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2,random_state=0)
```

Fig 3 : Dividing the dataset into training and testing datasets using train_test_split function

PROCESSING TECHNIQUES

- **Comparative Analysis:** Determine which model predicts happiness scores the best by comparing its performance to that of the others. Examine each model's advantages and disadvantages to learn more about how appropriate and comprehensible it is.
- **Testing on Unseen Data :** Evaluate the selected model on unseen data to assess its generalization ability. This step helps ensure that the model can accurately predict happiness scores for new countries or observations outside the training dataset.
- **Validate and Real-world Application:** Validate the model's performance by applying it to real-world scenarios or deploying it in a controlled environment. Monitor the model's predictions and iterate on improvements if necessary. For example, like a user-friendly website.

RESULT AND FINDINGS

In the analysis of the World Happiness Report data, the effectiveness of several regression models, such as Random Forest, Gradient Boosting, Ridge, and Lasso Regressors, was assessed. The R2 score, a metric that measures the proportion of variance in the dependent variable (ladder scores) that is predictable from the independent factors.

It's interesting to note that out of all the evaluated algorithms, the Random Forest Regressor proved to be the most accurate model. In the test dataset, it performed better than the Gradient Boosting, Ridge, and Lasso Regressors in terms of R2 score. The Random Forest model appears to have fit the underlying patterns in the World Happiness Report data better during the testing phase, based on the higher R2 score.

RESULT AND FINDINGS

The Random Forest algorithm's ensemble nature, which integrates several decision trees to reduce overfitting and improve prediction accuracy, may be the reason for its better performance. Because of its adaptability, the Random Forest model is especially well-suited for forecasting ladder scores, since it can capture relationships within the data. Although Ridge and Lasso Regressors and Gradient Boosting Regressors are strong algorithms on their own, the unique properties of the World Happiness Report data may make the Random Forest method more advantageous. In order to ensure an accurate representation of happiness rankings and contributing factors within the framework of the World Happiness Report, this result highlights the importance of methodically comparing and choosing regression models based on their performance metrics.

	Regression Models	R2_Scores(in %)	Mean_Squared_error	Mean_Absolute_error
0	Linear Regression	67.129383	0.255175	0.385951
1	Random Forest Regression	76.170678	0.184987	0.345507
2	Gradient Boosting Regression	68.243351	0.246527	0.390078
3	Ridge	67.323787	0.253665	0.384666
4	Lasso	-5.750208	0.820939	0.384666

Fig 4: Comparison of the models

RESULT AND FINDINGS

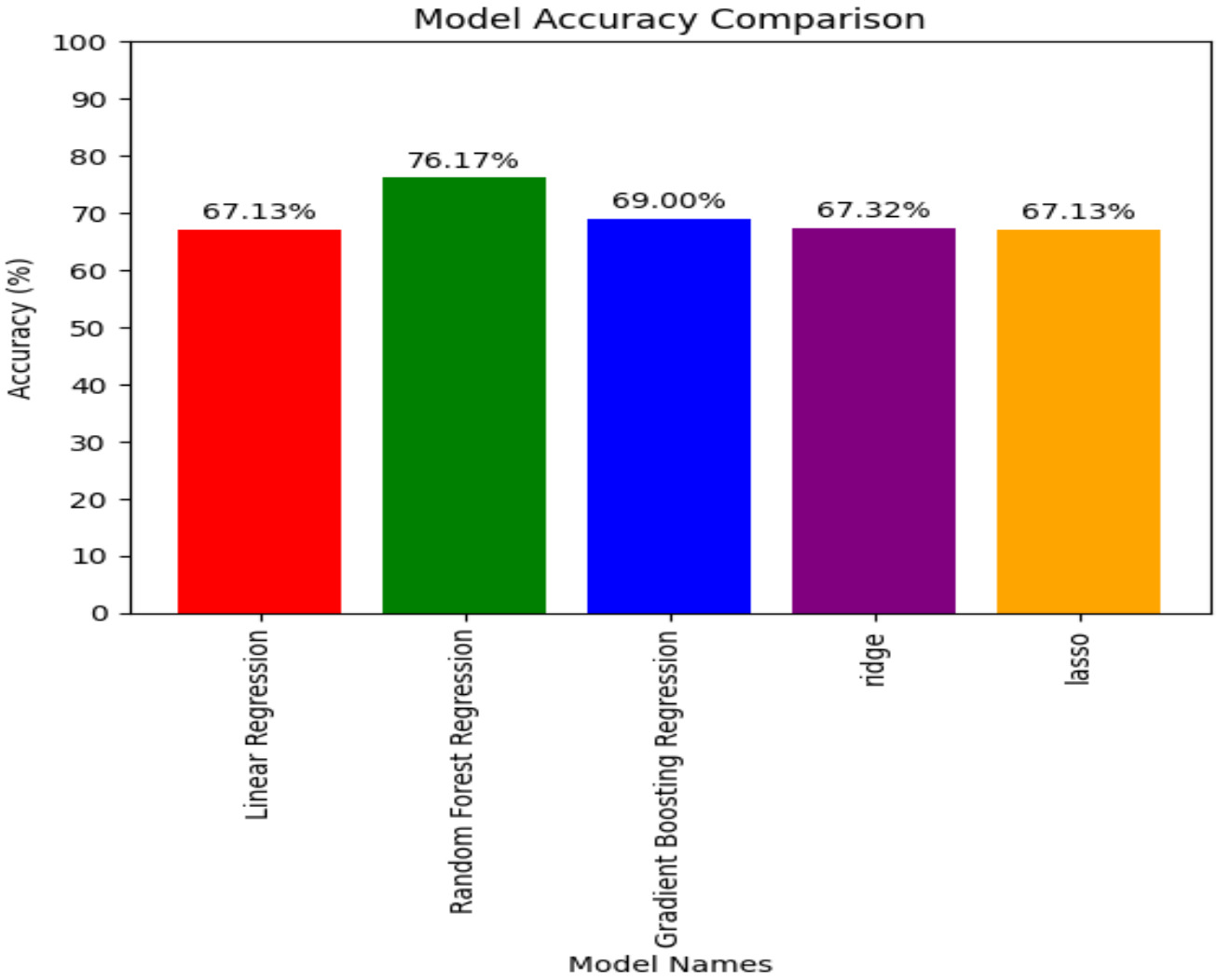

Fig 5: R2score on test data of various models

Fig 6: Ladder Score of different countries in the world

CONCLUSION

In conclusion, the World Happiness Report project provides valuable insights into the factors influencing happiness at a global and country level. By leveraging the comprehensive analysis and rankings provided in the report, the project aims to predict happiness scores based on user inputs, empowering individuals to understand and explore the happiness levels of their respective countries.

https://gitfront.io/r/user-5847750/N6PVRRfN15jY/happinesspredictor-repo/

FUTURE WORK

Website Development

Through the development of a user-friendly website and the integration of data from the World Happiness Report, the project will offer users the opportunity to input their country's data and personal information, resulting in a predicted happiness score. This allows individuals to obtain a better comprehension of the elements that go into happiness and provides a platform for exploring potential areas for improvement.

FUTURE WORK

Fig 7: Integration of this model as backend by creating a user-friendly website

REFRENCES

1. Andrew N. Rowan, Well Being International (WellBeing News) ,“World Happiness Report 2023”.
2. Charles Alba, ”A Data Analysis Of The World Happiness Index And Its Relation To The North-South Divide 2019”, Pennsylvania State University - Main Campus.
3. Francesco Sarracino HSE-LCSR; STATEC, Kelsey O'Connor, STATEC Research; IZA Institute of Labor Economics, “A Measure of Well Being Efficiency Based on the World Happiness Report, IZA Discussion Paper No. 15669 (2022)” .
4. Francesco Sarracino, Kelsey O'Connor, “Analysing happiness index as a measure along with its parameters and strategies for improving India’s rank in world happiness report (2020)”.
5. John Helliwell, Richard Layard and Jeffrey Sachs” World happiness report(2012)” . The Earth Institute, Columbia University, New York, USA.

THANK YOU