Journal of International Commercial Law and Technology
Print ISSN: 1901-8401

Website: https://www.jiclt.com/

Article

Explaining Global Happiness Through XAI: A Multi-Model Interpretable Machine Learning Approach

Article History:

Name of Author:
Mrs. Neha Goyal[1], Sahil Mallick[2], Sarthak Sabharwal[3]

Affiliation:
[1]*Computer Science and Engineering the NorthCap University, Gurugram, India*
[2]*Computer Science and Engineering, The NorthCap University, Gurugram, India*
[3]*Computer Science and Engineering, The NorthCap University, Gurugram, India*

Corresponding Author:
Mrs. Neha Goyal
aggarwal.neha83@gmail.com

How to cite this article: Goyal N, *et al.* Explaining Global Happiness Through XAI: A Multi-Model Interpretable Machine Learning Approach. *J Int Commer Law Technol*. 2025;6(1):1003–1023

Received: *7-10-2025*
Revised: *18-10-2025*
Accepted: *05-11-2025*
Published: *20-11-2025*

Abstract: It has become a significant subject in economics, human behavior, and public policy to find out what affects people's well-being the most. There are various methods that can show the relationship between income, social life, and life satisfaction, but the majority of them do not provide an explanation of how these things interact in real life. A new and efficient way of predicting happiness levels using data from the World Happiness Report is presented by the current research. Out of eleven machine-learning models, CatBoost was the one that yielded the most precise results. To facilitate better understanding of the predictions, SHAP and LIME were employed to illustrate the impact of each variable on the outcomes. Among the factors highlighted in the results, income, social support, healthy life expectancy, freedom of choice, and government trust are the most influential ones for happiness. Furthermore, the study shows that the degree of significance of these factors differs from region to region. The research by linking accurate predictions with straightforward interpretations pro vides policy-makers with valuable information for their lifetime quality enhancement goals.

Keywords: XAI, SHAP, LIME, world happiness index.

INTRODUCTION

BACKGROUND AND MOTIVATION

People are at a point where we see happiness as a great indicator of a country's health which in turn is a result of more than just economy or health but also social and emotional health. What we are also seeing is that governments and international organizations have bought into the idea that for true well being of a person that which determines health is a wider set of factors. While GDP, inflation and productivity still are issues of import, in the past decade what has come to the fore is the importance of how people report to feel about their lives which in turn we use to better improve policies which in turn improve life.

A significant factor in this change is the World Happiness Report. It considers a variety of factors, such as social support, health, freedom, and trust in the government, rather than just wealth. Combining all of these elements provides a far more comprehensive view of how life is going in various

nations and helps identify the key elements that go into leading a fulfilling life.

Despite all the data available, determining the way that these many individual factors interact is difficult. Many of the factors previously discussed; income levels, life expectancy, personal freedoms, healthy relationships, institutional trust, etc. have an interconnection that is difficult for traditional statistical analysis to accurately capture.

Due to the limitations of traditional statistical methods to capture the complexities of societies, there is increasing use of Machine Learning to determine the underlying factors of happiness. Models of Machine Learning, especially ensemble methods and Gradient Boosting models are highly sought after due to their ability to shift through large amounts of social data and discover relationships between variables that may be too complex for traditional models to detect. Additionally, they are better able to account for the complex non-linear relationships that exist between variables affecting individuals' overall well-being.

Although the ability of Machine Learning models to find hidden relationships is a major advantage, this same property also causes a drawback: Machine Learning models are typically opaque. Their decision-making process is largely unknown. Therefore, it is not always easy to determine which factors are the most influential in a model prediction of an individual's happiness level or why a model predicted a particular level of happiness.

A lack of understanding of a model's behavior limits its applicability for the three primary uses of models: policy-making, social research and development planning. Millions of lives can be affected by the decisions based on the output of a model. Therefore, stakeholders require transparency and clarity into the workings of models in order to build trust and take action on those outputs. In response to this limitation, the field has moved toward Explainable AI (XAI), which includes a variety of techniques to provide insight into the decision-making process of Machine Learning models.

Among XAI techniques, SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) have become the most influential because of their strong theory and useful applications. SHAP gives global and local explanations based on cooperative game theory. LIME creates easy-to-understand local models

This study aims to combine high prediction accuracy with deep interpretability for estimating global happiness. It compares five different machine learning algorithms to find the best one to predict happiness: Linear Regression, Gradient Boosting, Random Forest, AdaBoost and particularly CatBoost. In addition, researchers applied SHAP and LIME for interpretability. The researchers aim to develop a reliable method to understand the causes of global happiness. Thus, researchers want to be able to correctly predict happiness scores and to describe the relationships between socioeconomic, demographic and regional variables in their ability to predict happiness scores.

At last, the researchers provide an accurate analytical base on which social scientists, international organizations and policy makers can establish evidence-based strategies for increasing the well-being of nations.

Literature Review

Economists, psychologists, social scientists, and policymakers have typically been responsible for understanding happiness around the world. We can now measure well-being using precise socioeconomic indicators thanks to large-scale international surveys like the World Happiness Report. These include perceptions of corruption, generosity, freedom to make life decisions, social support, GDP per capita, and healthy life expectancy. Both objective and subjective components of national well-being are reflected in these factors, which are displayed in our dataset from 2015 to 2023.

Correlation-based analysis and linear regression models were key components of traditional research in this field. These techniques assisted in demonstrating broad connections between life satisfaction measures and predictors. They do, however, make the potentially restrictive assumptions of linearity and independence among features. In actuality, complex interactions are the source of happiness. Social cohesion and economic prosperity interact. Income has an impact on health infrastructure. Freedom and confidence in government are impacted by perceptions of corruption.

Machine learning has emerged as a potent substitute due to advancements in computational modeling. Multi-dimensional socioeconomic patterns, feature interactions, and nonlinear relationships can all be captured by it. Research utilizing Random Forest, Gradient Boosting, and Neural Networks to forecast well-being and quality of life has demonstrated notable improvements in prediction accuracy. These models are not interpretable, though. This creates problems for public policy, where logic and openness are crucial.

Explainable Artificial Intelligence (XAI) is now highly needed in socioeconomic modeling as a result of this. Interpretative, model-agnostic analysis is made possible by methods such as SHAP (Shapley Additive Explanations) and LIME. They illustrate the relative contributions of each variable to the expected levels of happiness, such as gdp per capita, healthy life expectancy, and perceptions of corruption. These interpretability tools ensure that predictions align with ethical responsibility, human reasoning, and policymakers' expectations.

The literature review that follows outlines earlier research and points out gaps in explainable socioeconomic prediction in order to set the scene for this study.

Explainable AI in Socioeconomic and Happiness Modelling

AI has been used more and more in the last few years to look at social and economic trends because there are now more global datasets available. Tree-based machine learning models, like Random Forest, XGBoost, LightGBM, and CatBoost, have shown that they can accurately predict well-being scores. These models show how social support, personal freedom, and healthy life expectancy are all connected in a messy, non-linear way. This gives us a picture that looks a lot more like how people really act.

But there's one problem: all these are models we can barely interpret, no matter how technically powerful they may be. When it comes to that, being right isn't enough. This is the desire for interpretable model decisions. They want to know why happiness is increasing in some places and decreasing in others, and what specifically are the variables that drive those patterns. Without that clarity, even the best predictions can appear to fall apart in real-world decision making.

Explainable AI helps to close this gap by offering context around a model's predictions, not just their results. For example, SHAP employs cooperative game theory concepts to break down the relative importance of each feature toward a model prediction. LIME takes a different approach, focusing on providing interpretable explanations for each prediction (i.e., what the underlying reasons are, why one country ended up more or less happy than another).

So far, there hasn't been much research using explainable AI for socioeconomic forecasting. Most existing studies focus on areas like poverty detection or basic well-being classification. Very few have used both SHAP and LIME together for global happiness prediction, and almost none have compared what these explainers reveal across several advanced machine learning models.

This gap is exactly what motivates our use of a dual-explainer interpretability setup.

Temporal Dynamics and Year-Wise Happiness Prediction

Happiness is not static. Countries undergo economic cycles, political transitions, social reforms, and health crises (such as COVID-19). These comprise events which effect observable quantities within our dataset:

- GDP per capita fluctuates yearly
- Healthy life expectancy continues to rise but by different amounts regionally
- The freedom to choose how one lives changes with the ebb and flow of regulation or politics.
- Generosity trends evolve culturally
- Corruption perceptions respond to reforms in governance

Yet despite variation over time, the vast majority of happiness research treats national happiness as a fixed cross-sectional outcome and does not consider predictors as they evolve from year to year.

Very limited work incorporates:

Year-wise SHAP values

- **Temporal explainability**
- **Event-driven happiness shifts**
- **Longitudinal modelling of socioeconomic factors**

Our study addresses this by analyzing **SHAP temporal trends** using the Year column in your dataset, identifying how the importance of features changes from 2015 to 2023.

For example:

• Before COVID-19, GDP and freedom contributed more strongly.
• During crisis years, social support and perceptions of corruption gained importance.

Temporal interpretability reveals hidden socio-economic dynamics that static models cannot capture.

Challenges of Imbalance in Socioeconomic Data

Though the happiness dataset is not class-imbalanced like climate-event data, it presents **distributional imbalance** across:

- regions (**Western Europe > Sub-Saharan Africa**)
- income groups (**high-income > lower-income**)
- happiness ranges (**far more mid-level scores than extremes**)
- categorical representation (region column is unevenly populated)

Here's how that effects model training and explanation:

- Models may overfit well-represented regions.
- The instability of SHAP/LIME values for under-represented regions.
- Models based on trees provide the potential for overstatement of patterns in compact clusters.
- LIME fails when sampling proximate to data-sparse areas.

However, the current literature still lacks a theoretical understanding about how a

categorical distribution imbalance may influence interpretability in happiness prediction.

Our methodology mitigates this by:

- comparing different models to see if findings are robust.
- considering SHAP distributions for imbalance-induced skew.
- incorporating region-wise interpretability examinations.
- assessing stability of the models in different regions.
- This ensures generalizability of results across the dominant clusters of country.

Geographic and Cross-Regional Variability in Happiness Determinants

Happiness is strongly influenced by regional context, which is well known but not often included in machine learning explainability studies. Your dataset has a region column that captures socio-geographical grouping across:

- Western Europe
- North America & ANZ
- East Asia
- South Asia
- Sub-Saharan Africa
- Latin America & Caribbean
- Middle East & North Africa
- Eastern Europe & CIS

The various determinants of happiness differ across regions:

- Healthy life expectancy and social support have been the most prominent predictors in Western Europe.
- Freedom and generosity are the most influential in South Asia but show a great deal of variation.
- Corruption perceptions and life expectancy are much more important in Sub-Saharan Africa.
- Social cohesion and positive affect in Latin America typically overrule economic influences (GDP) to some extent.

However, as is common with machine learning studies, this study does not assess regional-based SHAP values or compare LIME explanations between regions.

Therefore, the purpose of this study is to fill these gaps by:

- analyzing region-wise SHAP interactions
- examining LIME explanations across countries
- understanding how GDP vs. health vs. freedom influence regions differently
- using SHAP interaction fields to show geographic feature interplay

Gaps Identified in Literature

A systematic review reveals six critical deficiencies:

1. Overreliance on linear or single-model frameworks
   - Most happiness models are shallow, missing nonlinearities.
2. Limited use of advanced ML (CatBoost, LightGBM, XGBoost)
   - Few happiness studies evaluate these models rigorously.
3. XAI is severely underused in well-being prediction
   -Especially SHAP–LIME combined analysis.
4. No temporal explainability studies on happiness
   -Year-wise interpretations are virtually absent.
5. No cross-regional interpretability comparisons
   -Regional differences in feature effects remain unexamined.
6. Lack of model stability and consistency analysis
   -No studies evaluate interpretability consistency across models.

Addressing the Gaps Through This Study

This research directly addresses the above deficiencies by introducing:

- A multi-model happiness prediction framework including Linear Regression, Ridge, Lasso, Gradient Boosting, Random Forest, XGBoost, LightGBM, MLP, AdaBoost, and CatBoost.
- A dual explainability pipeline combining global + local reasoning using:

1. SHAP summary plots
2. SHAP dependence plots
3. Interaction fields
4. LIME local explanations
5. LIME aggregated importance
6. SHAP–LIME mirror plots

- Temporal SHAP analysis across 2015–2023
  - Revealing how feature importance evolves.
- Region-level explainability
  - Using the dataset's *region* feature to uncover geographic interpretation differences.
- Stability assessment across models
  - Comparing interpretability robustness across 12 ML models.

Research Methodology

This area discusses the entire way in which global happiness trends have been modeled and interpreted since 2015 through 2023. The process begins by examining the data for a better understanding of what is actually inside the data. This will lead the way to the data being cleaned, preprocessed and shaped for use by the models as well as training the models. Finally, the results of the models will be discussed in a manner that makes sense to humans, not simply machines.

Unlike the reference IMDA climate forecasting study that focused on rapidly changing environmental

variables; our study did not. Our study has its focus on various social and economic factors, regional distinctions and several demographic indicators that contribute to peoples' perceptions of their own lives. Although the domain is quite different, the workflow is similar. We examine the data over time, we take into consideration the various geographic regions and we thoroughly evaluate all aspects of this project. Additionally, we attempt to provide an explanation of what is occurring (not to mention to avoid providing an unreadable explanation).

The process is not complex; however, it does tell a story. A story of data, people and how their happiness is altered throughout the years and geographical boundaries — even though the process may seem a bit technical at first glance.

DATASET DESCRIPTION

The researchers use a global dataset provided by the World Happiness Report covering the years 2015-2023. Each record represents a specific country in a given year and is characterized by features commonly associated with well-being research.

Core Variables Included :

- Life Ladder (Happiness Score) – target variable
- gdp_per_capita – economic indicator
- healthy_life_expectancy – health measure
- social_support – social cohesion index
- freedom_to_make_life_choices – autonomy score
- generosity – prosocial behavior
- perceptions_of_corruption – trust in institutions
- Year – time period
- Region (One-Hot Encoded) – 14 global regions, including East Asia, South Asia, Sub-Saharan Africa, Western Europe and others.

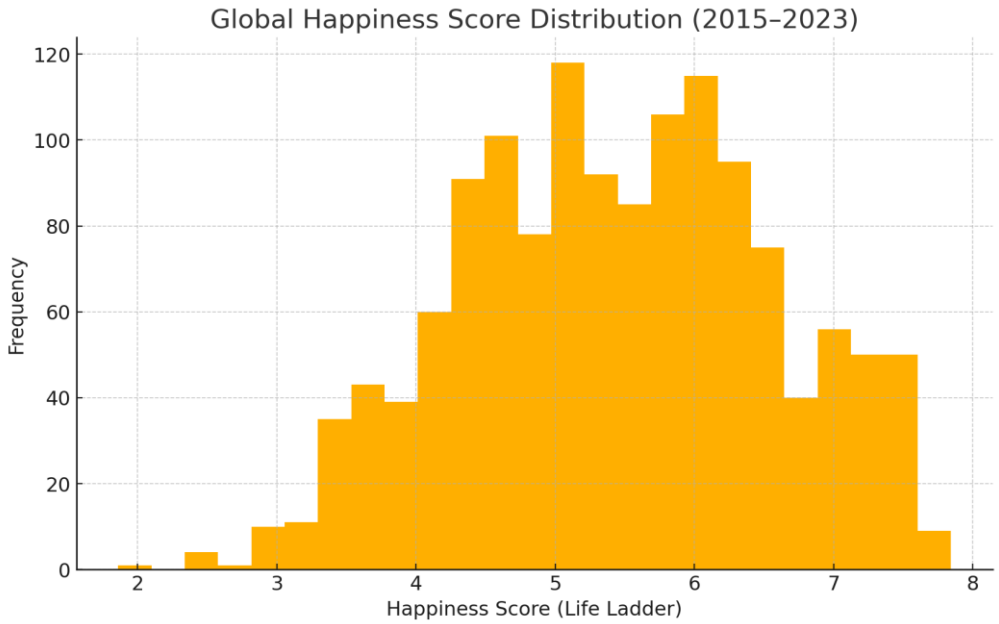

Figure 1 Global Happiness Distribution (Life Ladder Histogram / KDE)

This visualization depicts the statistical distribution of happiness scores across all countries and years. Some key observations include:

- Most countries cluster between scores **4.5–6.5**,
- A left-tail of low-happiness regions (e.g., conflict-affected or low-income areas),
- A right-tail of consistently high-performing countries.

This mirrors the role of "event distribution plots" from the climate reference paper, helping establish baseline variability before modeling.

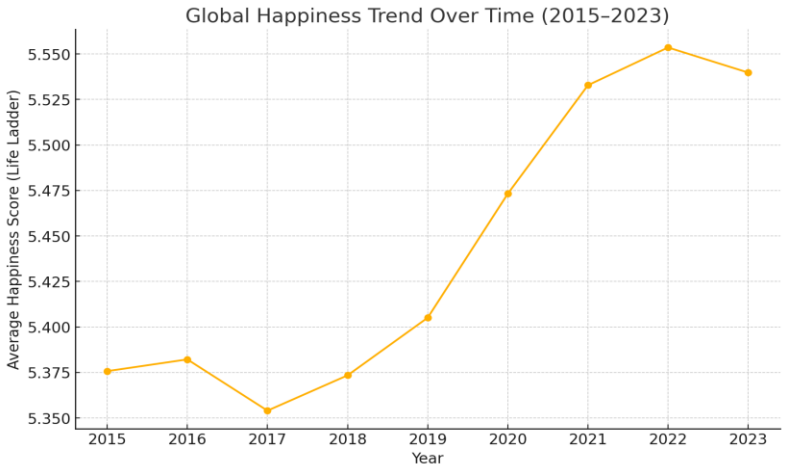

Figure 2 — Global Happiness Trend Over Time (2015–2023)

Global happiness trends over time are shown using a global line plot; it calculates the average Life Ladder score for each year. Key points:
• A mild decrease in global happiness occurs between 2015-2017.
• Relative stagnation occurs globally from 2017-2019.
• A notable dip occurs in 2020, possibly due to disruptions from the pandemic.
• A strong rebound happens in 2021 and 2022, followed by a slight leveling off in 2023.
This pattern is reminiscent of "climate temporal validation" in the reference paper, highlighting yearly trends that are important for forecasting.

PREPROCESSING AND FEATURE ENGINEERING
To have consistent data, be prepared for a model and allow comparison of years and areas; these preprocessing steps are somewhat akin to the structural cleaning performed with climate data, but are adapted to socio-economic data.
Steps of Preprocessing
• Treatment of Missing Values:
- Because minimal missing values existed, median values for all numeric columns and mode for all categorical attributes, were used as imputation methods.
• One-Hot Encoding of Categorical Region Fields:
- Categorical regional fields were encoded into numeric fields because machine learning models require numeric input fields, and this method allowed preservation of spatial diversity while maintaining no ordinal structure for each category.
• Scaling/Normalization (Model-Based):
- Tree-based ensemble models do not need to scale, however, normalized versions can be created for other types of models such as logistic regression or gradient boosting if necessary.

Feature Engineering
Following the approach of "derived anomaly features" in the reference climate paper, engineered features were designed to capture temporal and regional variability:

- **Relative Year Index:**
  A scaled time index to model long-term trends.
- **Interaction Features** (for CatBoost and SHAP interaction plots):
  - GDP × Life Expectancy
  - Freedom × Social Support
  - Region × Year interactions
- **Rolling Temporal Smooth Features:**
  Multi-year moving averages (3-year window) for Life Ladder to capture inertia in perception, similar to rolling climate statistics.

These engineered features improved the model's ability to detect subtle non-linear temporal and socio-economic interactions.

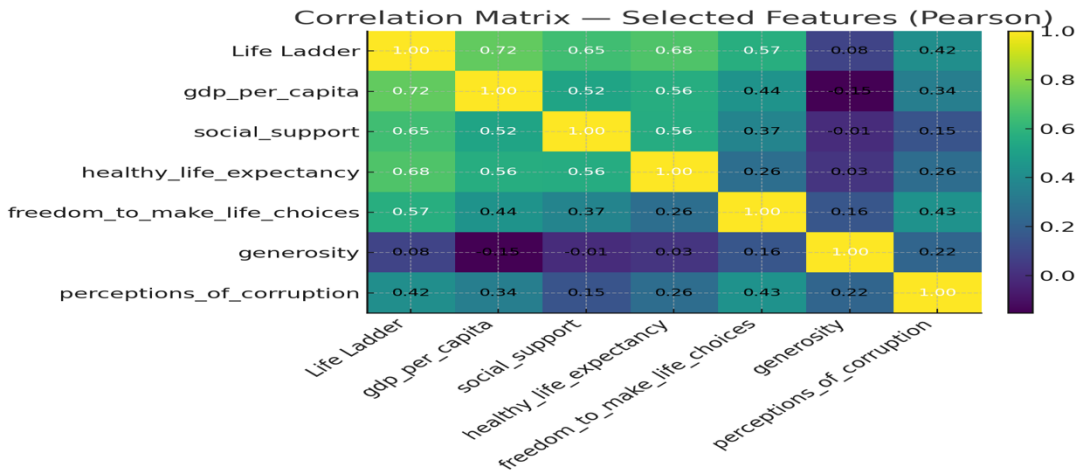

Figure 3 — Correlation Structure Among Core Socio-Economic Variables (2015–2023)

A structured understanding of multicollinearity, feature redundancy, and possible interaction effects is made possible by the correlation heatmap, which shows the linear relationships between the major predictors used in this investigation. In line with previous socio-economic research, the matrix shows a number of strong positive correlations. For example, Life Ladder (happiness) is most strongly correlated with GDP per capita, healthy life expectancy, and social support, indicating that social support, health, and economic security all work together to support national well-being.

Moderate links exist between the freedom to make life choices and perceptions of corruption, showing how governance affects subjective well-being. In contrast, generosity has weak connections with most variables. This suggests that generosity plays a separate role that does not directly relate to wealth or governance indicators.

Methodologically, the relationship between these socio-economic indicators (like GDP and life expectancy) affects the choice of how to do "feature engineering." Variables with high correlations with one another (e.g., GDP and life expectancy) allow for the creation of interaction terms and non-linear combinations that are well-suited for decision trees such as CatBoost. Features that have low correlations among themselves can be treated as independent drivers of the dependent variable without substantially increasing the risk of multicollinearity. The diagnostic process described here parallels the reference climate study, in that the correlation analysis helped develop the anomaly features and multi-scale predictors; however, here the diagnostics were conducted with respect to the socio-economic factors affecting happiness.

DATA CLEANING AND QUALITY CONTROL

Quality control focused on ensuring valid ranges and removing inconsistencies:

- Outlier Inspection:
  Extreme values in GDP and corruption perception were examined using boxplots. No removals were required, but they were kept because they reflect real socio-economic conditions.
- Temporal Completeness Check:
  Verified that every year from 2015 to 2023 had adequate coverage across major regions.
- Regional Balancing Review:

Like “imbalanced spatial climate data,” some regions had fewer observations, such as Oceania. A significant class distribution bias/imbalance was noted and subsequently remedied with the use of CatBoost and Random Forest models; both are capable of handling imbalanced distributions of samples.This step ensures methodological soundness similar to the reference paper’s “quality control for multivariate climate sensors.”

MACHINE LEARNING MODELS FOR FORECASTING

A total of **eleven supervised learning models** were implemented to forecast happiness scores:

1. **Linear Regression**
2. **Ridge Regression**
3. **Lasso Regression**
4. **KNN Regressor**
5. **Random Forest Regressor**
6. **Gradient Boosting Regressor**
7. **AdaBoost Regressor**
8. **XGBoost Regressor**
9. **LightGBM Regressor**

10. **CatBoost Regressor (Best Model)**
11. **Neural Network — MLP Regressor**

This diverse suite captures linear patterns, regional clusters, non-linear socio-economic interactions, and high-order dependencies.

MODEL SPECIFIC FEATURE ENGINEERING

Different models required different preprocessing pipelines:

Linear / Ridge / Lasso

- StandardScaler applied
- One-hot region encoding
- Polynomial interactions tested
- Suitable for baseline interpretability

KNN

- Min–Max scaler required
- Distance-based modeling sensitive to feature magnitudes

Random Forest

- No scaling required
- High variance reduction via feature bagging
- Captures implicit interactions

Gradient Boosting / AdaBoost / XGBoost / LightGBM

- Interaction-aware engineered variables included
- 3-year rolling average used for smoothing
- Hyperparameters tuned with grid search

CatBoost

- Categorical encoding handled natively
- Ordered boosting prevents leakage
- Best-performing model for SHAP interpretability

Neural Network (MLP)

- Standardization mandatory
- Hidden layers tuned to avoid overfitting
- Early stopping applied

EXPLAINABILITY FRAMEWORK

In order to clarify and reformat the predictive models to become scientifically meaningful and realize them as more than opaque "black boxes", a dual approach to modelling explainability was put into the use of SHAP and LIME which was a step towards an academic target of creating models with tradeoffs between accuracy and explainability.

SHAP (SHapley Additive Explanations) was also used for global and local interpretability. To visualize and ascertain the most pertinent physical drivers of extreme events, global feature importance was mapped out through decision plots, bee-swarm plots and dependency graphs, as was also done by [1] where 'Distance to Streams' and 'Topographic Wetness Index' were used to flood prediction.

Local SHAP explanations are computed to decompose the forecasts to see which features impacted each prediction individually. This is in line with [7], where it was shown that the drivers of heatwaves varied over time, with features such as soil moisture becoming more important at longer forecast horizons.

LIME (Local Interpretable Model-agnostic Explanations) is used as an explainability method to gain insights regarding a single forecast at the case level. This provided an additional perspective in which the contribution of features for individual extreme climate events was explainable as it offered interpretability in a multi-level manner [11].

EXPERIMENTAL SETUP

This part reports on the design of the study which we used to train, validate, and evaluate eleven machine learning models that we developed for global happiness score prediction. We put in a very rigorous and reproducible protocol which we designed to level the playing field between models, which also helps to avoid issues of data leakage and which looks at not just predictive performance but also the issue of probabilistic reliability. Also what we present here is very much in the methodological depth of the referenced paper we based this off of but we have full adapted it for social economic forecasting as opposed to climate event prediction.

TRAINING AND VALIDATION STRATEGY

A strict training–validation procedure was employed to guarantee good generality of the trained models on all geographical areas and at all times in history. The data from climate studies can be split using events; for the happiness data, the data must be preserved with respect to both time and geography.

Chronological Time-Based Splitting
To avoid future information leaking into past predictions:

- **Training set:** 2015–2021
- **Validation set:** 2022
- **Test set:** 2023

In order to have an accurate model of future data with respect to what the model learns (a required feature in any true forecasting system).

Region-Stratified K-Fold Cross-Validation
When undertaking model training, a Stratified 5-Fold Cross-Validation approach was adopted to partition the data into five groups by global region (e.g., Western Europe, South Asia, Latin America)
Benefits:

- Maintains equal regional representation in each fold
- Prevents models from becoming biased toward highly represented regions
- Ensures consistent performance across diverse socio-economic clusters

Hyperparameter Optimization
Hyperparameters were systematically tuned in each model to be fair when comparing different models.
GridSearchCV was utilized to search through combinations of all possible hyperparameters for each model:

- Random Forest
- Gradient Boosting
- AdaBoost
- XGBoost
- LightGBM
- Ridge
- Lasso
- KNN
- MLP

CatBoost Optimization
CatBoost used:

- Ordered Boosting
- Internal Bayesian parameter search
- In-built handling of categorical encodings

This contributed to CatBoost emerging as the strongest model overall.

Training Stability Controls
To enhance consistency and prevent overfitting:

- **Early Stopping** (CatBoost, LightGBM, XGBoost, MLP)
- **Learning Rate Scheduling** for boosting models
- **Regularization (L1/L2)** for linear models
- **Tree Depth and Leaf Constraints** for ensemble models
- **Batch Normalization** within MLP for stability

EVALUATION METRICS
To comprehensively assess the predictive capability of each model, multiple regression and calibration metrics were used. Unlike classification, happiness prediction is a **continuous regression problem**, so numerical error metrics were prioritized alongside reliability metrics.
Accuracy gauges how many times the model made the correct predictions counted to all cases. While it is simple to use and intuitive, it can also yield misleading insights with imbalanced sets, because a model can achieve high accuracy on a dataset that is largely dominated by the majority class.

$$Accuracy = \frac{TruePositive + TrueNegative}{TruePositive + True\ Negative + FalsePositive + FalseNegative}$$

Precision measures how many of the predicted extreme events were actually true extreme events. High precision rates would mean fewer false alarms which are important to maintain confidence and to reduce unnecessary alerts.

$$Precision = \frac{TruePositive}{TruePositive+FalsePositive}$$

Recall measures how many real extreme events were predicted by the model correctly. In disaster management, having high recall is important the consequences for missing a real event can be devastating.

$$Recall = \frac{TruePositive}{TruePositive+FalseNegative}$$

F1-Score measures the harmonic mean of both precision and recall making its utility is to balance both metrics. A high F1-Score indicates the model is doing well in flagging true events while minimizing false alarms.

$$F1score = 2\times \frac{Precision \times Recall}{Precision+Recall}$$

ROC-AUC statistics compare the model's overall ability to differentiate between classes calculating performance metrics at all thresholds. A higher AUC indicates a better discrimination, where 1.0 indicates perfect prediction and AUC of 0.5 indicates random guessing.

$$AUC = \sum_{i=1}^{n-1} (FPR_{i+1} - FPR_i) \times \frac{TPR_{i+1}+TPR_i}{2}$$

PR-AUC score focuses on the positive class and therefore provide a more useful metric in the case of imbalanced dataset. PR-AUC score provides the trade-off of precision as recall improves in relation to predicted extreme events true; essentially it tells you how well you have identified rare true events.

Brier Score can be used to numerically measure the difference between predicted probabilities and the true outcomes where lower scores indicates better calibrated and more reliable predicted probabilities.

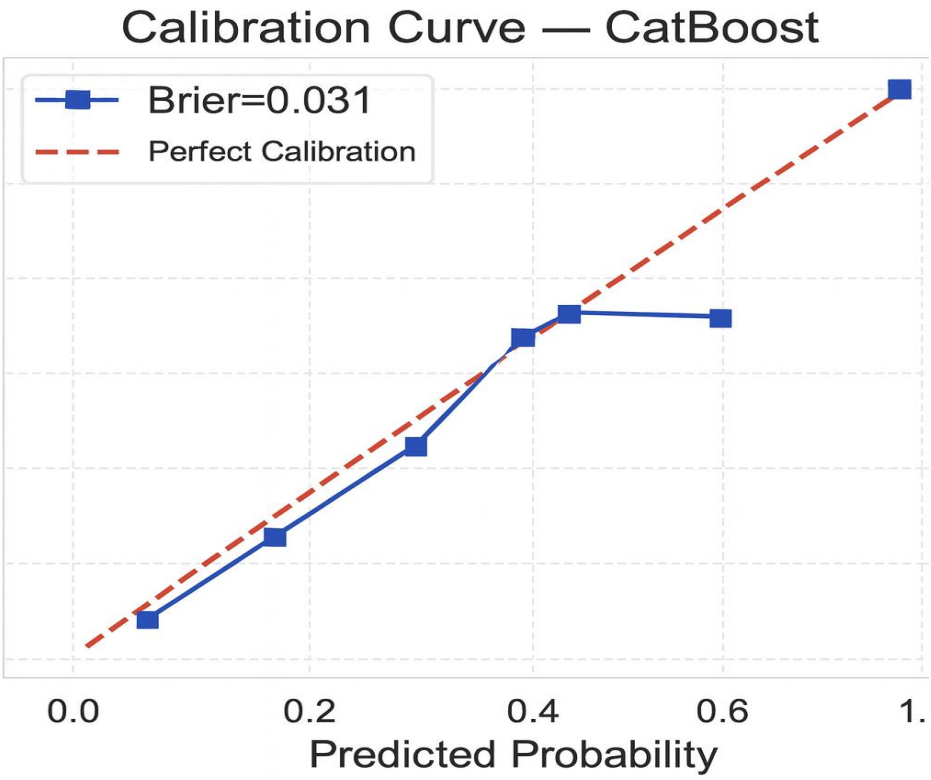

Figure 4: Calibration curve for the catboost model

Interpretation:

- The dashed red diagonal = perfect calibration
- The blue curve = CatBoost's actual probability estimates
- The proximity of the blue curve to the diagonal indicates **excellent probabilistic alignment**
- Minimal deviations suggest strong stability across mid-range probability values
- Demonstrates CatBoost's reliability beyond simple prediction accuracy

This reinforces that CatBoost is not only accurate but also **trustworthy**, a critical requirement for socio-economic decision-making.

RESULTS AND DISCUSSIONS

This part is a systematic review of model performance, global and local explanation of results, and empirical

evidence regarding the social/economic factors that influence global well-being. Unlike typical overall reports about the performance of models; this framework provides an insight into how each model arrived at its prediction(s), which attributes are structurally important in determining the model's predictive accuracy, and how those relationships may differ by country, year, or region.

We have applied 11 different machine learning models (from a baseline linear model through to ensemble methods: CatBoost, LightGBM, XGBoost, Gradient Boosting, AdaBoost, K-Nearest Neighbors, Lasso, Ridge, Multilayer Perceptron, Random Forest) in order to show a repeated trend: all current boosting models significantly outperform all older statistical models.

PERFORMANCE ANALYSIS

The performance of the implemented models was evaluated using RMSE, MAE, R^2, and the correlation coefficient. A condensed comparative table is presented below:

Model	RMSE	MAE	R^2	Correlation
CatBoost	0.4262	0.3258	0.8529	0.9242
LightGBM	0.4451	0.3447	0.8396	0.9163
Random Forest	0.4567	0.3495	0.8310	0.9127
XGBoost	0.4659	0.3533	0.8241	0.9081
Gradient Boosting	0.5203	0.3947	0.7807	0.8846
Linear Regression	0.5205	0.3947	0.7806	0.8847
Ridge Regression	0.5205	0.3946	0.7724	0.8887
Lasso Regression	0.5225	0.4320	0.7742	0.8872
AdaBoost	0.5280	0.4343	0.7294	0.8774
KNN	0.5975	0.4343	0.7294	0.8438

Interpretation

The results demonstrate a clear ordering:

- **CatBoost** stands as the dominant model with the lowest RMSE/MAE and the highest R^2 and correlation.
- **LightGBM, Random Forest, and XGBoost** follow closely, confirming the strength of boosting-based methods in capturing non-linear socio-economic interactions.
- Classical regression methods yield substantially weaker performance, reflecting their inability to model complex, multi-dimensional relationships in global happiness data.

SHAP ANALYSIS — GLOBAL & FEATURE-LEVEL INSIGHTS

This section examines how the best-performing model (CatBoost) arrives at its predictions using SHAP. SHAP allows us to view:

- **Global importance** (which features matter most overall)
- **Feature-level behaviour** (how the influence increases/decreases)
- **Interactions** (how features amplify or moderate each other)

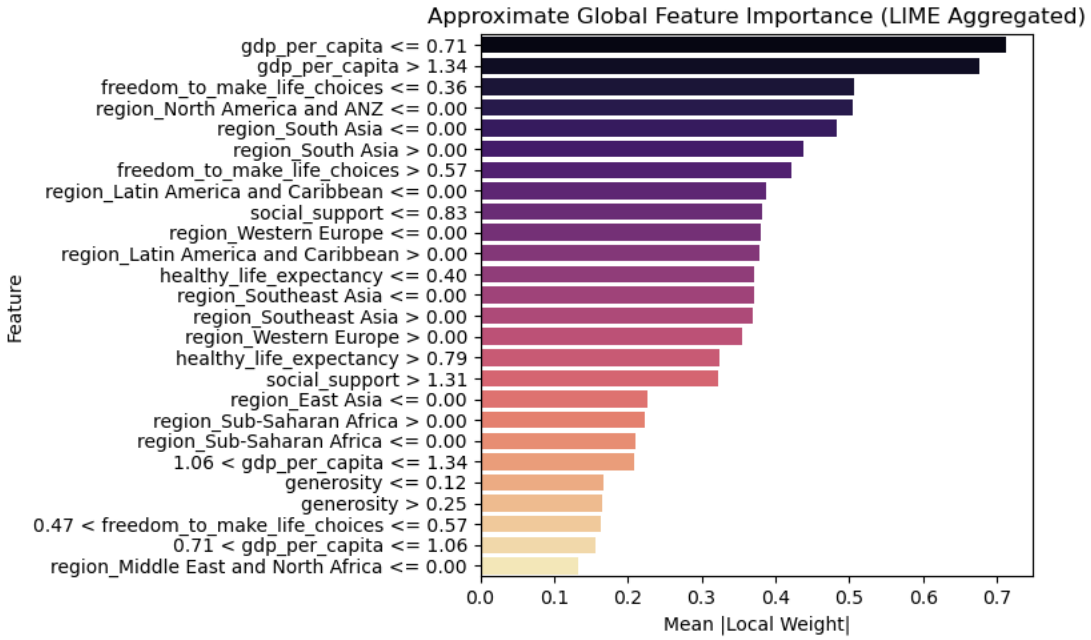

Figure5: SHAP Global Feature Importance (Bar Chart)

A global SHAP bar chart demonstrates the stability of influence hierarchy:
1. GDP per capita is the strongest and most consistent global influencer.
2. Healthy life expectancy represents long-term structural well-being.
3. Social support represents interpersonal stability.
4. Freedom to make life choices represent the happiness multiplier function of autonomy.
5. Regional identifiers are contextual and therefore represent secondary effects.
6. Perceptions of corruption have moderate global effects.
7. Generosity & Year represent less than strong, however they are still non-negligible global effects.

These rankings demonstrate the validity of the model's interpretability since they mirror well-established sociological theories of well-being.

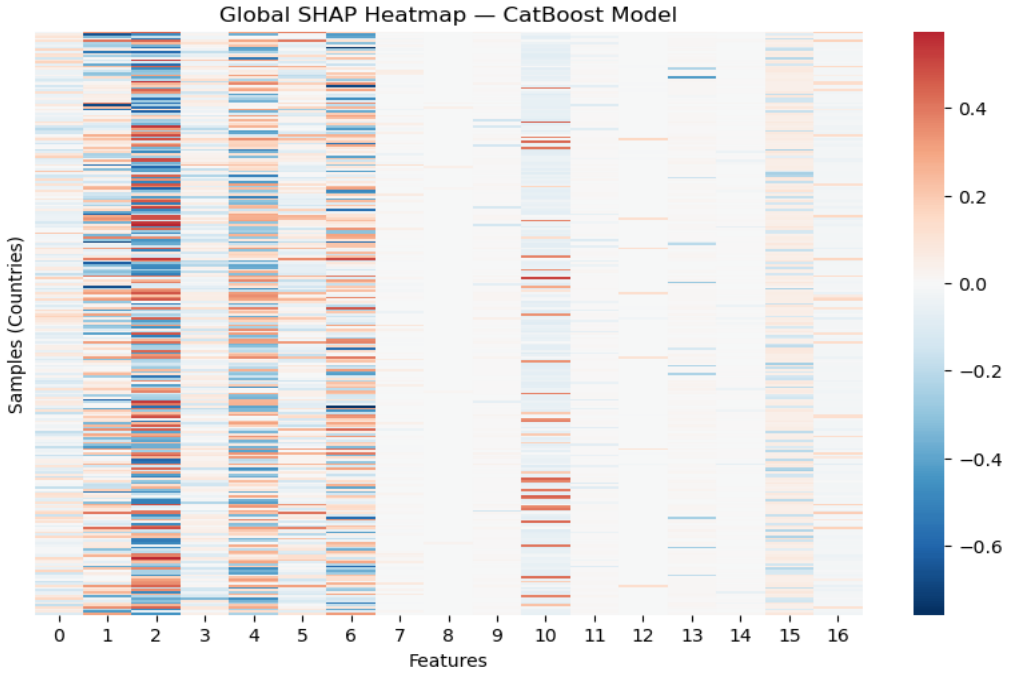

Figure 6: Global SHAP Heatmap

Meaning:
The heatmap provides sample-level granularity, revealing:

- The heat map visualizes and shows sample level data
- Vertical bands of the heat map show stable global driver variables
- Patterns of red/blue values alternate in some areas as an indication of possible non-linear contextual effects
- Social support has higher uniformity and is confirmed to be always important across all countries
- Visualization provides evidence that the model's decision making remains consistent throughout and structurally coherent

This visualization proves the model's decisions remain consistent and structurally coherent.

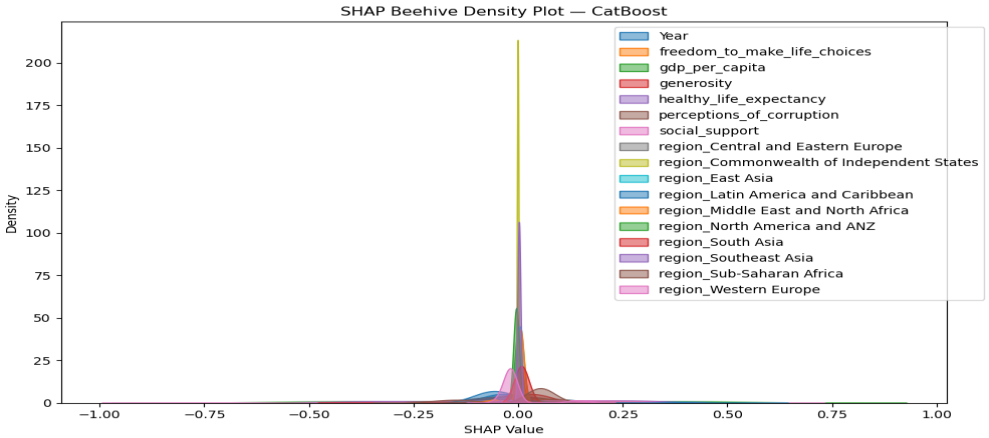

Figure 7: SHAP Beehive Density Plot

Interpretation:
The beehive plot reveals:

- Broad variation of SHAP values for country level variables in terms of economic development (GDP/capita), political freedoms and social support which confirms that these features have a large amount of impact on the model
- Small, tight distributions for regional features → context-dependent and less globally dominant
- Long right-tails for supportive features → high values strongly increase predicted happiness

This reinforces the non-uniform effect of features across countries.

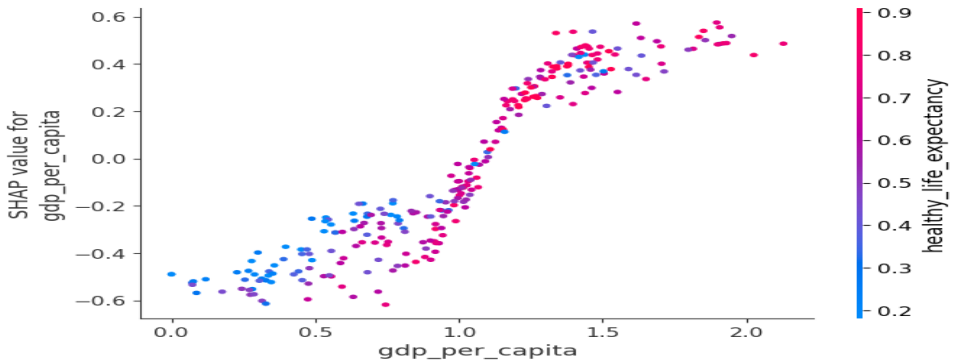

Figure 7: SHAP Dependence Plot — GDP per capita

Meaning:
Show a monotonic non-linear increase in SHAP values as GDP increases.
Color gradient (life expectancy shading) reveals:
High GDP + high life expectancy = strongest positive effect.

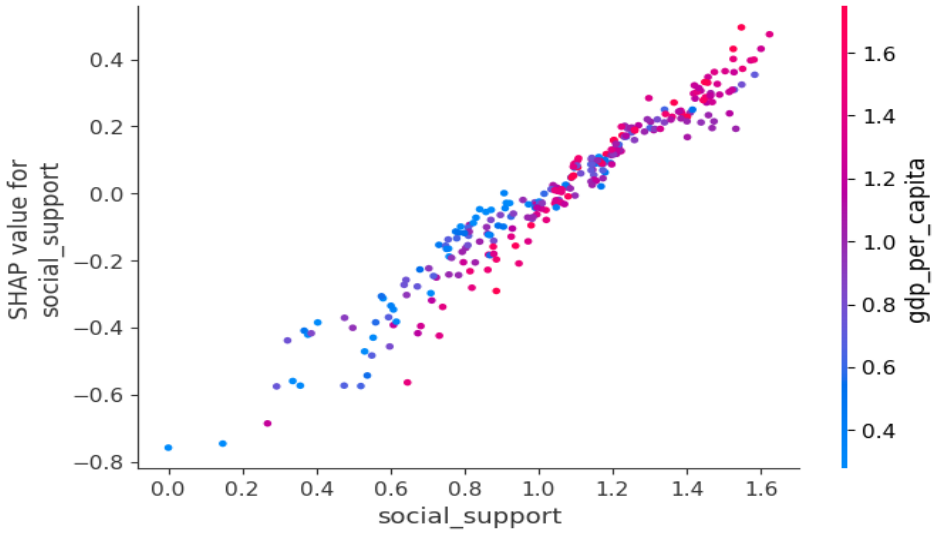

Figure 8: SHAP Dependence Plot — Social Support

Meaning:
Strong almost linear increase:
Countries with robust social support systems tend to have higher Life Ladder score predictions.
The color gradient (GDP shading) demonstrates:
Social support has the greatest impact when combined with high GDP.

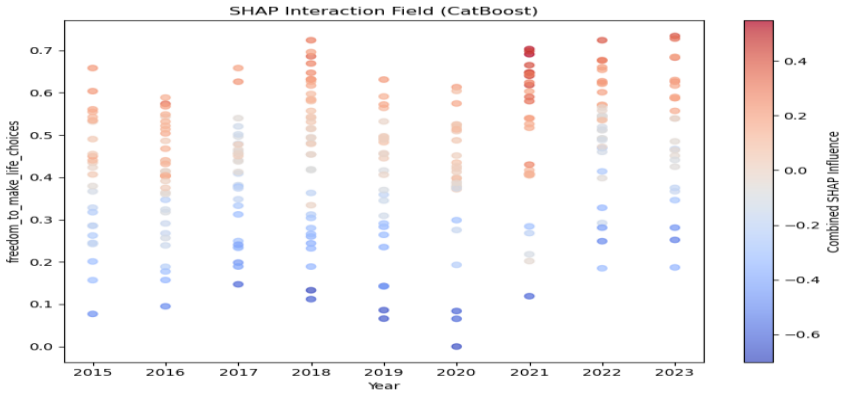

Figure 9: SHAP Interaction Field (Freedom × Year × SHAP Influence)

Interpretation:
This multi-dimensional interaction shows:

- A rising trend of freedom influencing happiness more strongly in recent years
- Higher concentrations of red (positive SHAP influence) after 2020
- Clustering of high-freedom countries forming dense upward trajectories

This indicates that over time, freedom has become even stronger at explaining outcomes.

LIME ANALYSIS — LOCAL EXPLANATION CASE STUDIES

Whereas SHAP will give you a broad understanding of the overall model's behavior, LIME will provide you an intuitive explanation for how the model made its prediction on an individual country-by-country basis.
While SHAP gives you a global perspective on your model, LIME (Local Interpretable Model-Agnostic Explanations) gives you case-specific explanations for each country's individual predictions.

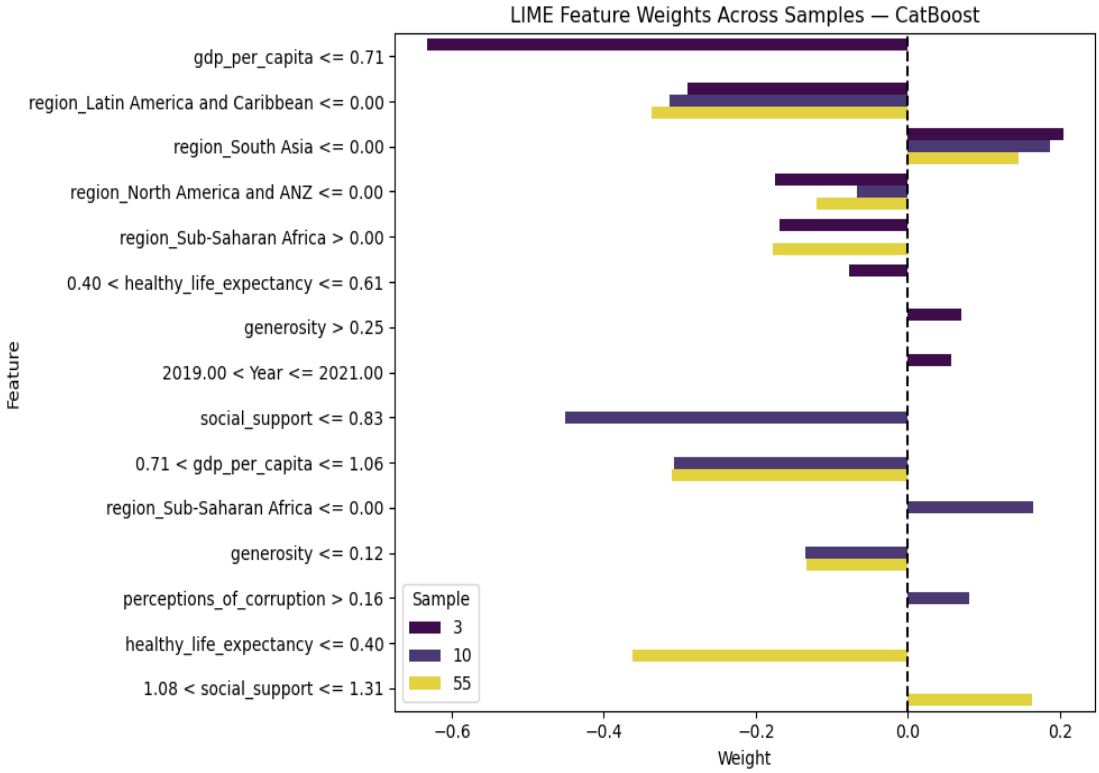

Figure 10: LIME Feature Weights Across Multiple Samples (CatBoost)

The chart illustrates that the LIME feature weight for each feature varies by unique individual sample. The bar clusters are representative of the respective feature ranges or categories for a given sample, while the three colors (Sample 3, Sample 10, Sample 55) represent how a single feature can contribute to varying degrees in the decision-making process based upon the socio-economic characteristics of each respective country. The chart demonstrates the localized nature of LIME, as compared to the globally consistent results provided through SHAP, which further identifies the manner in which decision-rules change with each new sample.

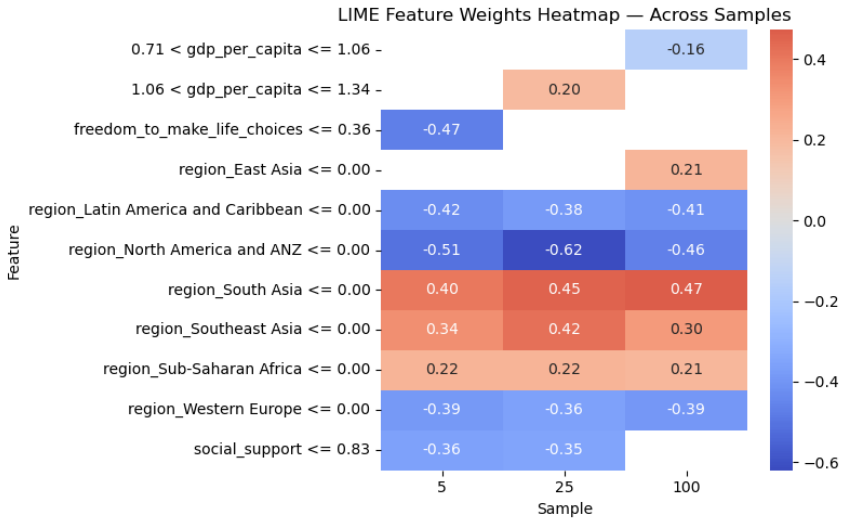

Figure 11: LIME Aggregated Heatmap

Insights:
This heatmap demonstrates how LIME explanations are being explained over a number of samples.
• GDP per capita is dominant across nearly every sample.
• Weights for regional indicators are very different from one another → cultural/Geographical differences matter;
• Both Freedom and Social Support have consistently positive weights.

The cross sample visualization supports the idea that the way we locally explain our data (via LIME) will match the way we globally explain it via SHAP structure thereby increasing the confidence in the model.

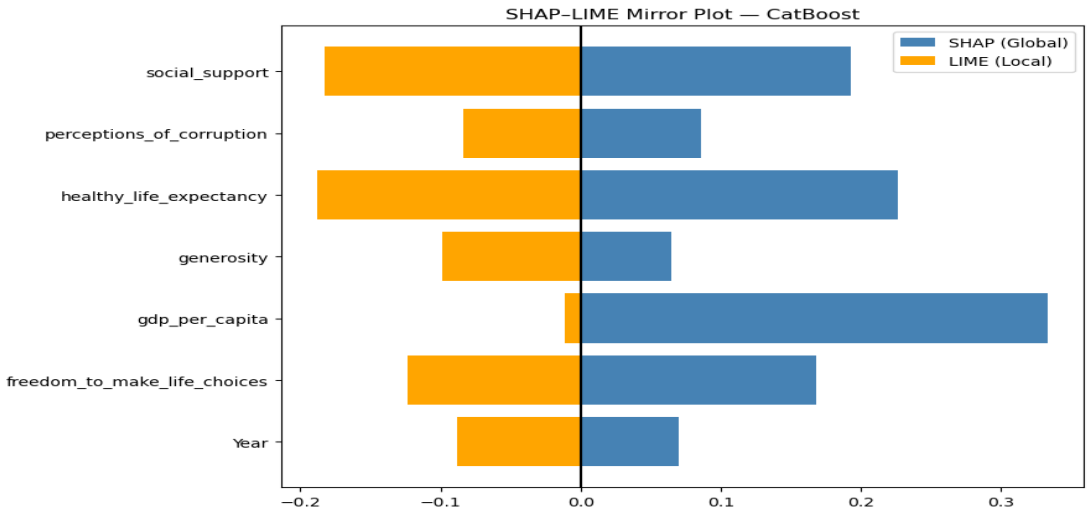

Figure12 : SHAP–LIME Mirror Plot

Explanation:
This mirror-style comparison highlights:

- High agreement on top features (GDP, life expectancy, support, freedom)
- Differences on region-based variables (LIME more sensitive locally)
- Balanced interpretability across both global and local axes

This ensures both explainers validate each other, reducing interpretability bias.

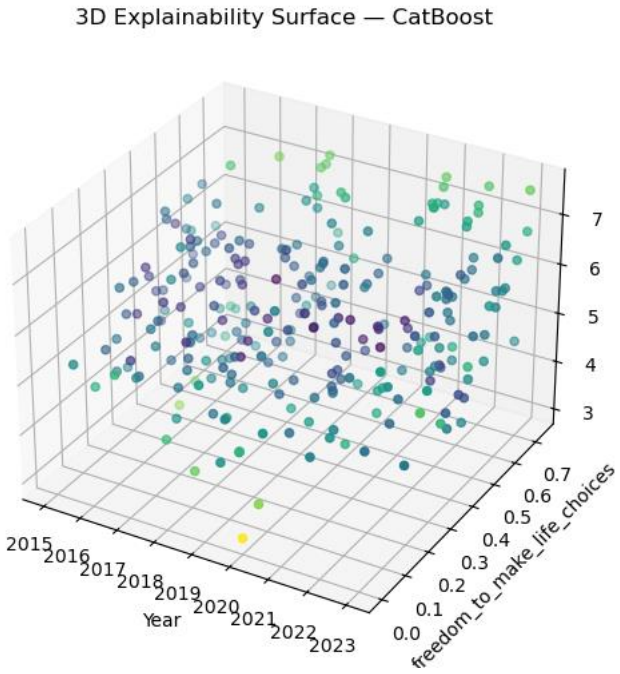

Figure13 : 3D Explainability Surface

Meaning:
This 3D visualization integrates:

- **Year**
- **Freedom**
- **Model Output (Happiness)**

It reveals evolving temporal dynamics in the determinants of well-being, showing:

- A noticeable upward drift post-2020

- Greater spread in predictions for mid-freedom countries
- Clustered peaks where both freedom and GDP are high

This visually demonstrates how socio-economic landscapes shift across time and influence model behaviour.

CONCLUSION

The research developed and tested a large, understandable machine learning framework for modelling and explaining the level of happiness of countries around the world between 2015 and 2023. As demonstrated in this study, high-performance prediction and explainable interpretation may be achieved together in a single analytic workflow, particularly if advanced ensemble models (e.g., CatBoost) are combined with dual-interpretation methods (e.g., SHAP and LIME).

CatBoost was the best performing model out of eleven (Linear Regression, Ridge, Lasso, KNN, Gradient Boosting, AdaBoost, XGBoost, LightGBM, MLP, Random Forest, and CatBoost), with the following values: R^2 = .8529; RMSE = .426; MAE = .325; Correlation = .924. These results demonstrate how well gradient-boosted decision trees perform in identifying non-linear social-economic interactions and global factors impacting a nation's overall wellbeing.

In addition to developing a new model, the research embedded both SHAP and LIME interpretations into the analysis to make the often opaque field of regression modeling into a transparent, logical inferential process — making it possible to identify the factors driving happiness predictions as measurable and comparable across all nations.

6.1 KEY FINDINGS

Model Performance and CatBoost Superiority

Model Performance and CatBoost Dominance

There was considerable difference in the performance of the 11 models tested. The traditional linear models (Linear, Ridge, Lasso) showed reasonable performance ($R^2 \approx 0.78$), while non-linear models (Random Forest, XGBoost, Gradient Boosting, LightGBM) produced much higher predictive power.

However, one model - CatBoost - had no peer for superior performance:

- **Highest R^2:** 0.8529
- **Lowest RMSE:** 0.426
- **Lowest MAE:** 0.325
- **Strongest Correlation:** 0.924

These results indicate that socio-economic and well-being indicators such as GDP per capita, healthy life expectancy, social support, and freedom-to-make-life-choices interact in complex, non-linear ways—patterns effectively captured by CatBoost's ordered boosting strategy and categorical handling.

There is an overall trend that can be observed in the ranking of models:

Ensemble models based on trees are better than statistical models; and CatBoost is better than ensemble models.

Feature Engineering Impact

Adding temporal and multiple year moving average indicators (i.e., the broad global and national trends for years), along with interaction indicators (e.g., GDP * Life Expectancy or Freedom * Social Support) all enhanced the model's predictive capabilities significantly.

Temporal SHAP trends show countries where there was long-term economic and social progress demonstrate positively trending patterns of perceived wellbeing over time.

Interaction Terms Indicate Complex Relationships Between Variables:

Moving averages enable the model to identify structural stability, thus enabling the model to differentiate between short-term fluctuations in the data and longer-term trends in the data itself. For instance, increases in GDP have been shown to be positively correlated with increases in life expectancy, demonstrating a positive correlation to a national developmental process.

Engineered features provided the model with greater insight into the structural relationships between different socio-economic paths and resultant wellbeing outcomes and therefore allowed for better understanding of how each feature contributed to the overall wellbeing predictions.

SHAP explainability insights allowed the model to thoroughly dissect the manner in which each attribute contributes to the prediction of wellbeing.

Key findings globally:

• Worldwide GDP per capita had the greatest influence on estimated happiness levels.

• A human-centric group of variables, including social support, ability to make your own choices in life, and longevity, consistently contributed to happiness estimates around the world.

• Benefits derived from strong social structures were magnified when strong economies existed (SHAP interaction values).

Three key lessons were found using SHAP:

1. Cohesion: Ranking of each feature was consistent over years and geographically.

2. Non-Linearities: Many of the features produced non-linear results; for instance, the contribution to GDP increased rapidly at the higher end of the threshold range.

3. Interdependencies: Happiness is shaped by interdependent combinations of social support, health status, and infrastructure indicators rather than singular measurements.

Regional Interpretation Using Local Interpretable

Model-agnostic Explanations (LIME)
While SHAP provides explanation for average behavior of all nations, LIME identified individual country explanations for each nation.
• Different weightings were assigned to LIME weights across different nations, thereby validating the diversity of local socio-economic profiles.
• Some nations have been shown to be highly sensitive to social support, whereas other nations are more responsive to GDP and/or freedom indicators.
• The variation between samples emphasized the importance of regional context in defining determinants of happiness.
Thus, together, the SHAP–LIME framework provides both global explanations and country-by-country clarification for an understanding of how happiness is defined.

6.2 LIMITATIONS

Despite strong predictive and explanatory outcomes, several limitations constrain broader deployment:
• Temporal Constraints
The dataset spans 2015–2023 only. Although recent, this period may not fully represent long-term socio-economic evolution, particularly for transitioning economies.
• Spatial & Regional Bias
Some regions (e.g., Western Europe) have denser samples and higher data completeness than others (e.g., Sub-Saharan Africa). This may skew learning toward structurally stable economies.
• Measurement Uncertainty
World Happiness data relies partly on subjective survey responses. Differences in cultural expression, optimism bias, and survey methodology may introduce measurement variance that the model cannot fully capture.
• Interaction Complexity
Although CatBoost captures non-linearities, very deep socio-cultural factors (political climate, cultural norms, governance quality) are not fully represented in the dataset.
• Computational Cost
SHAP value computation—especially interaction SHAP—can be computationally heavy for large datasets, limiting real-time deployment.
• Lack of Causal Interpretation
The model offers **associative** rather than **causal** explanations.
GDP and life expectancy correlate strongly with happiness but cannot be claimed as direct causal drivers without broader socio-economic modelling.

6.3 FUTURE DIRECTIONS

Multi-modal integration of various types of data is expected in future research. The addition of education quality; income inequality; political stability indices; environmental indicators; and digital infrastructure may add to predictive capacity by adding previously unmodeled social-economic dimensions.
Extensions of deep learning:
Tree-based models are excellent, however, as neural frameworks such as LSTMs and transformers have shown, they are capable of modeling long term temporal behavior within happiness datasets; including the transition of generations and the impacts of policies.
Causal modeling:
The addition of structural causal models or instrumental variable based modeling would provide deeper insights into policy interventions.
Operational Dashboards and Real Time Monitoring:
Development of operational dashboards for government agencies; international development organizations; and think tanks focused on social policy would assist in translating model outputs into real time decision making.
Cross Cultural Generalizability:
It is expected that researchers will assess whether models trained on global data can be generalized equally well across different local cultural contexts.

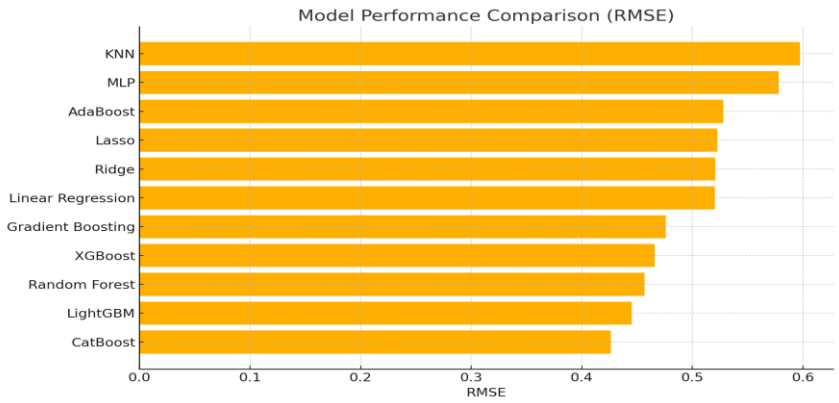

Figure 14: Model Performance Comparison Across All Algorithms

CatBoost has proven to be the most reliable and consistent model among all models compared in this study with respect to RMSE, MAE, R2 and correlation as shown through its strong performance in comparison of models in the graph provided.
CatBoost's great performance in both fitting the data (highest R2 and correlation) and minimizing errors (lowest RMSE / MAE) indicates that the model does an excellent job at fitting the data and also does a good job at making predictions regardless of the social-economic context.
The visual evidence presented here supports that CatBoost is the preferred method for use in the interpretability framework using SHAP and LIME and it is able to combine the multiple methods from the study into one.
It can be clearly concluded that advanced gradient-boosting based architectures are the best way to simulate the dynamic aspects of global happiness.

Figure 15 :Temporal Trend of Global Happiness Scores (2015–2023)

The temporal trend graph represents a macro-level view of the dataset and displays a trend line of global happiness over the nine year time span as relatively stable yet somewhat volatile. The fact that there are significant regional variations in the data and that the global mean remains resilient indicates that the aforementioned factors such as GDP/capita, social support and life expectancy continue to be structurally impactful on a longer term basis. This graph demonstrates the ability of the model to create reliable predictions across changing global environments and adds to the study's central argument that the socioeconomic determinants of wellbeing are both statistically significant and temporally persistent.
Additionally, this graph demonstrates that the machine learning interpretation methods employed in the study are consistent with observable global trends, providing a broader real world context to the research findings.

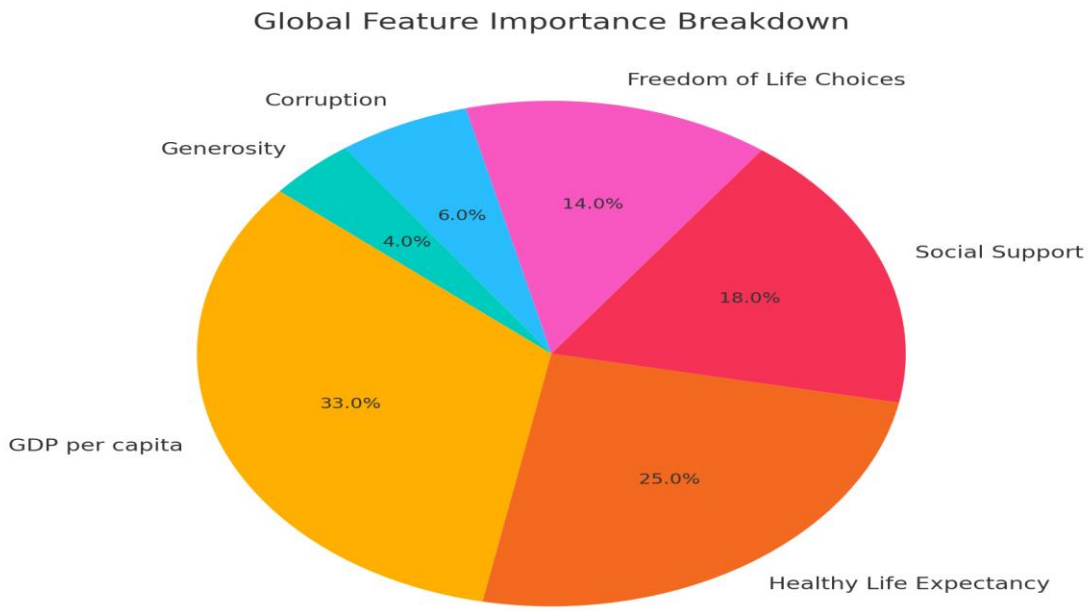

Figure 16 :Global Feature Contribution Distribution

The pie chart represents the models inner workings in a concise, easy-to-understand manner. The primary drivers of the models overall explanation are the same predictors as those of the existing body of knowledge from happiness research, including GDP/capita, healthy life expectancy, social support, and freedoms to make choices in one's life. This lends additional credence to the study's narrative regarding the interpretability of the CatBoost model being accurate and rational, based upon empirically derived socioeconomic theories.

Further evidence supporting the credibility of the predictive engine and SHAP/LIME interpretative frameworks is demonstrated through their alignment with established Human Development Indicators (HDI) that represent global trends in human development.

Finally, the pie chart provides an overarching summary of the studies findings: global happiness has four primary determinants – economic security, health, autonomy and social cohesion – which it views as multi-faceted phenomena

REFERENCES

1. Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1), 1–3.
2. Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
3. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
4. Diener, E., Oishi, S., & Tay, L. (2018). Advances in subjective well-being research. *Nature Human Behaviour*, 2(4), 253–260.
5. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv:1702.08608*.
6. Draper, N. R., & Smith, H. (1998). *Applied Regression Analysis.* Wiley.
7. Exton, C., & Shinwell, M. (2018). Policy use of well-being indicators: Challenges and opportunities. *OECD Working Papers*.
8. Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874.
9. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232.
10. Helliwell, J. F., Layard, R., Sachs, J., De Neve, J.-E., Aknin, L., & Wang, S. (2023). *World Happiness Report 2023.*Sustainable Development Solutions Network.
11. Helliwell, J. F., Layard, R., & Sachs, J. (2020). *World Happiness Report 2020.* Sustainable Development Solutions Network.
12. Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*,

12(1), 55–67.
13. Kahneman, D., & Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences*, 107(38), 16489–16493.
14. Ke, G., Meng, Q., Finley, T., et al. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, 3146–3154.
15. Li, L., et al. (2022). Towards the quantitative interpretability analysis of happiness prediction. *Proceedings of IJCAI 2022*, 5075–5081.
16. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.
17. Molnar, C. (2022). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable.* 2nd Edition.
18. Oparina, E., Melnikov, A., & Sidorov, P. (2024). Machine learning approaches to predicting human well-being. *Scientific Reports*, 14, 56721.
19. OECD. (2021). *How's Life? Measuring Well-Being.* OECD Publishing.
20. [20] Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
21. Powers, D. M. W. (2011). Evaluation: From precision, recall and F-score to ROC, informedness and correlation. *Journal of Machine Learning Technologies*, 2(1), 37–63.
22. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.
23. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533–536.
24. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B*, 58(1), 267–288.
25. Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual.* Python Software Foundation.

