# Happiness Level Prediction - Kaggle Regression Contest 24S1

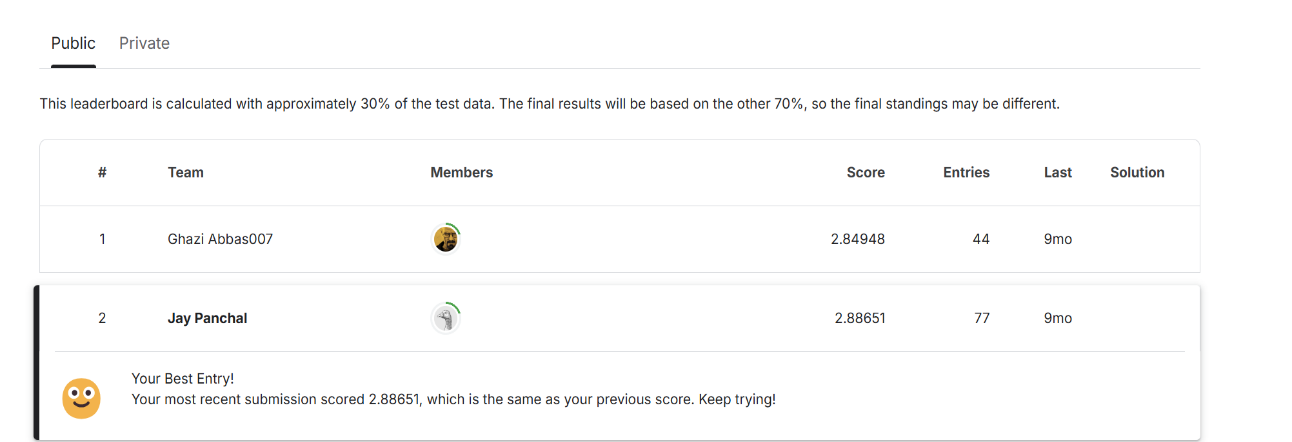
**Project Overview**

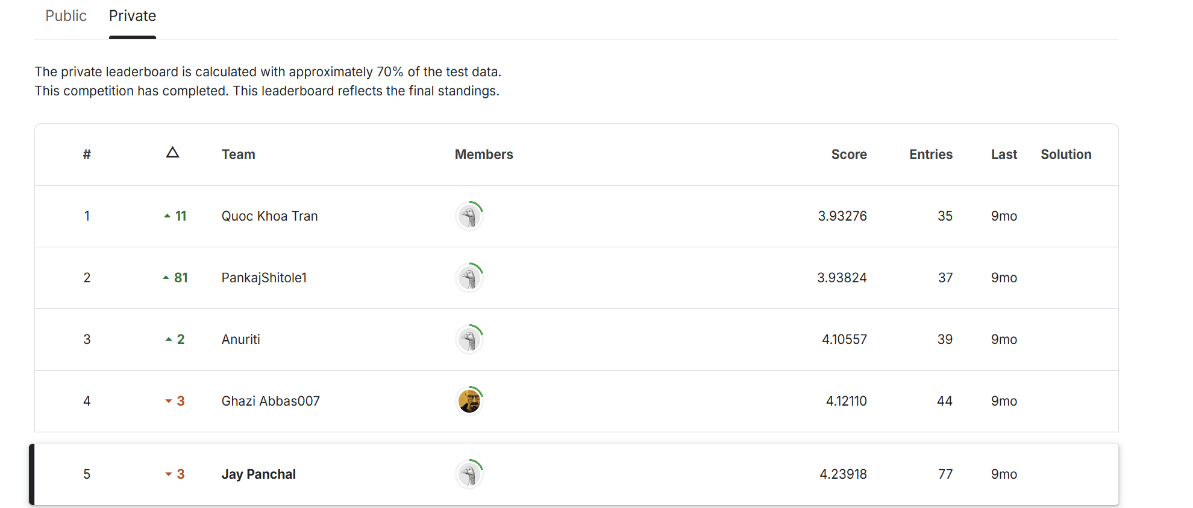
Welcome to the **Happiness Level Prediction** project—a sophisticated predictive analytics solution designed to estimate individual happiness levels based on detailed survey responses related to life satisfaction and overall well-being. Developed for the prestigious **Kaggle Regression Contest 24S1**, this model leverages cutting-edge machine learning techniques, achieving outstanding predictive performance.

**Remarkable Achievements**

This project demonstrated exceptional effectiveness by securing impressive rankings in a highly competitive field:

* 🏆 **Public Leaderboard:** Secured the **2nd place**.



* 🏅 **Private Leaderboard:** Achieved **5th place**.

These rankings underscore the accuracy and reliability of the developed predictive model.

**Comprehensive Dataset**

The dataset utilized encompasses diverse features, including numerical, categorical, and ordinal variables meticulously sourced from comprehensive surveys.

**Files Included:**

* regression\_train.csv: Contains robust training data for model development.
* regression\_test.csv: Features the test dataset for unbiased model evaluation.
* 33564000\_final\_model.R: The master script encapsulating the model-building and prediction generation process.

**Expert Data Preprocessing**

* Rigorous handling of missing data with strategic imputation (-999 values).
* Advanced encoding techniques, including one-hot encoding, optimized for categorical data.
* Precise numerical transformation of ordinal variables.
* Efficient feature normalization via min-max scaling to enhance model performance.

**Innovative Methodology**

This model integrates a powerful **stacked ensemble learning** strategy:

**Base Predictive Models**

* **Linear Regression:** For clear interpretability and baseline performance.
* **Random Forest:** Harnessing ensemble-based decision trees for accurate predictions.
* **Gradient Boosting (GBM):** Leveraging iterative improvement for superior predictive capability.
* **XGBoost:** State-of-the-art boosting algorithm offering exceptional accuracy and speed.
* **Neural Networks:** Utilizing deep learning to uncover intricate patterns in data.

**Robust Model Evaluation**

* Implemented rigorous 10-fold cross-validation to ensure robustness and avoid overfitting.
* Calculated RMSE scores meticulously to benchmark model performance effectively.

**Stacking for Optimal Predictions**

* Employed predictions from individual models as input features for a powerful **meta-model**.
* Utilized a high-performing neural network as the stacking meta-model to produce accurate and reliable final predictions.

**Seamless Model Training and Execution**

**Quick and Easy Setup**

Install all necessary R packages effortlessly:

install.packages(c("data.table", "caret", "randomForest", "gbm", "xgboost", "nnet"))

**Execution Steps:**

Follow these simple steps to replicate our success:

1. Load libraries.
2. Import and preprocess data.
3. Train and cross-validate robust base models.
4. Generate stacked features and train the superior meta-model.
5. Generate precise predictions ready for submission.

Execute effortlessly with:

source("33564000\_final\_model.R")

**Exceptional Results**

* Achieved outstanding predictive accuracy with a **Public RMSE of 2.88651**.
* Demonstrated excellent generalization capabilities, recording a **Private RMSE of 4.23918**.

**Opportunities for Further Excellence**

* Explore hyperparameter tuning with Bayesian Optimization for superior performance.
* Integrate advanced feature selection methods to refine the predictive accuracy.
* Experiment with deeper neural architectures and ensemble diversity.

**Acknowledgments**

This exceptional project was proudly completed for the Kaggle **Regression Contest 24S1**.

**References**

[Regression contest 24S1: Predict 'happiness' level | Kaggle](https://www.kaggle.com/competitions/regression-contest-24-s-1-predict-happiness-level/overview)