



Análisis de Felicidad Mundial

Proyecto de Ciencias de
Datos End-to-End



Jesus David Gil y Tayná Beija Napier



Introducción

Este proyecto analiza los factores socioeconómicos y de bienestar que influyen en la percepción de felicidad a nivel mundial.

Utilizando el dataset “World Happiness Report”, buscamos entender tendencias históricas, identificar variables clave y predecir los niveles futuros de felicidad en los países.



Objetivo Principal:

- Entender qué factores tienen mayor impacto en la felicidad.

Objetivos Específicos:

- Analizar tendencias históricas de felicidad global.
- Identificar variables socioeconómicas clave.
- Implementar modelos predictivos para estimar el puntaje de felicidad.
- Optimizar modelos con ajuste de hiperparámetros.



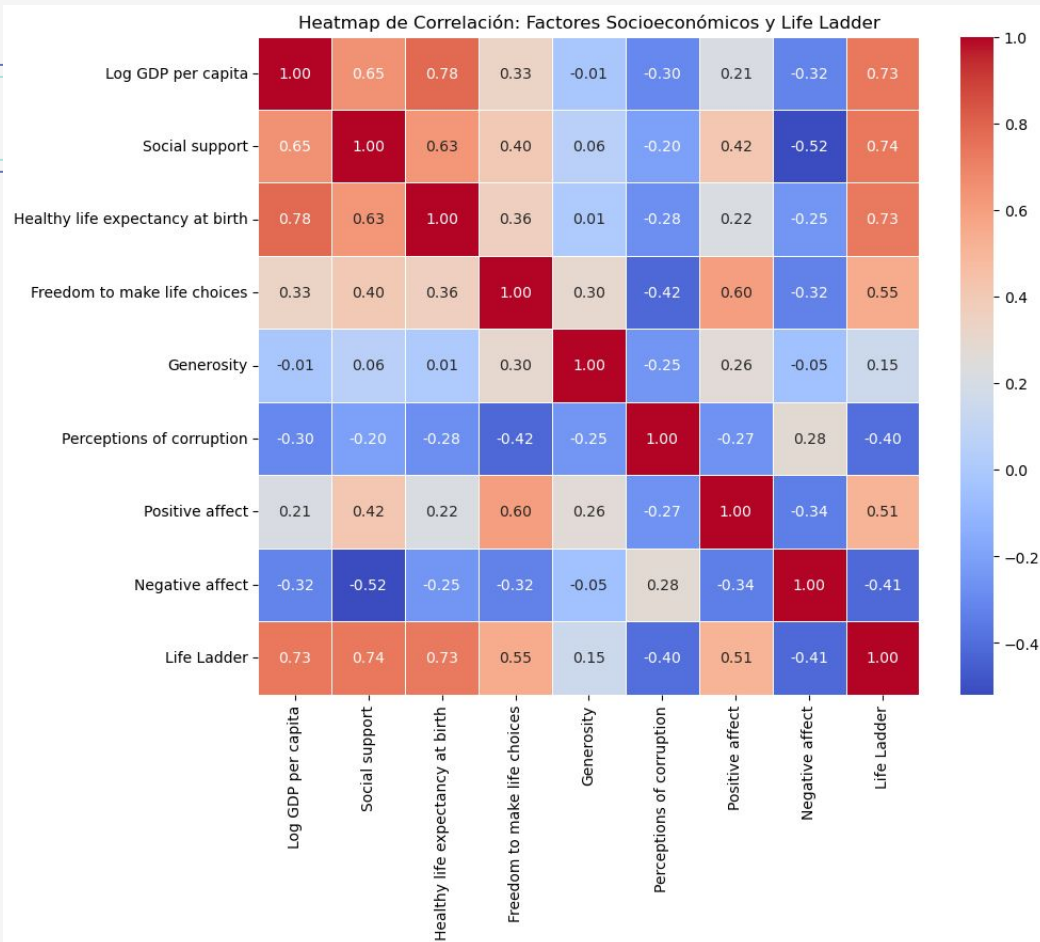
Análisis Exploratório de Datos:

- Identificación de valores nulos y outliers.
- Análisis de correlaciones entre variables como PIB per cápita y Life Ladder.
- Visualización de distribuciones de datos, tendencias y patrones.



Mapa de Calor de Correlación: Factores Socioeconómicos y Life Ladder

En este mapa de calor se observa una alta correlación entre el PIB per cápita y Life Ladder, indicando que son dos variables claves del dataset, y que la economía juega un papel importante en la felicidad.



Modelos Predictivos Supervisados

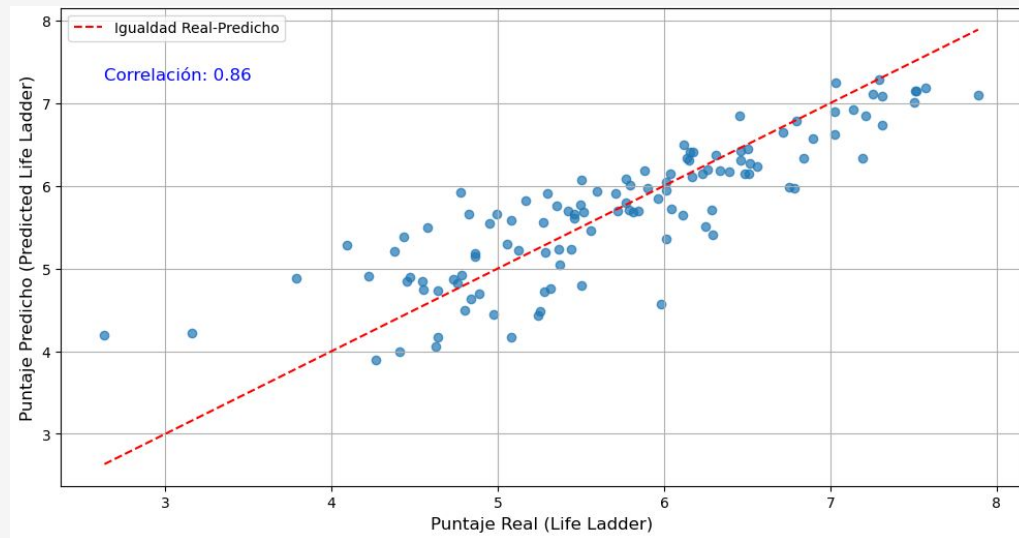
1. Regresión Lineal:

- Usada como modelo de referencia por su simplicidad e interpretabilidad.
- Limitada en su capacidad de capturar relaciones no lineales.

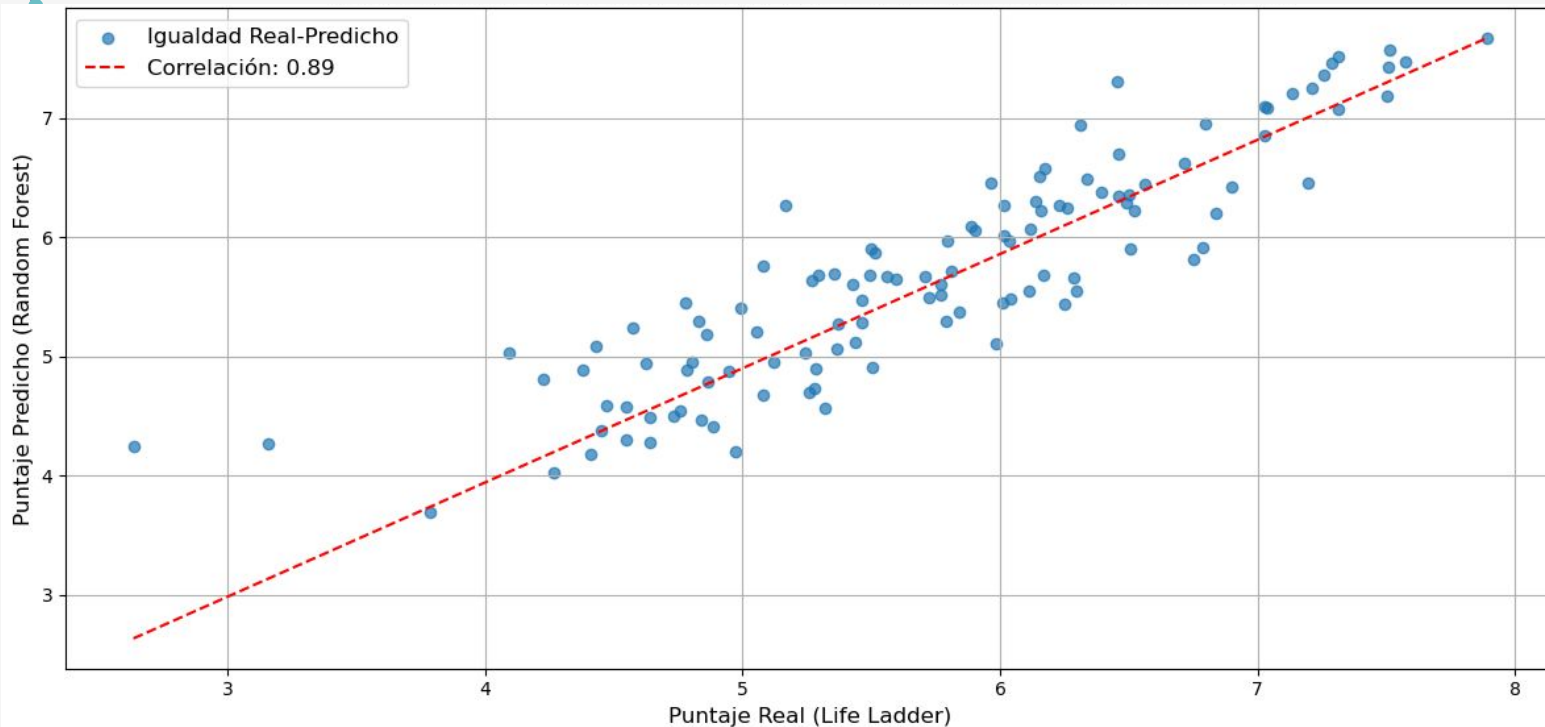
2. Random Forest:

- Modelo más robusto para datos complejos.
- Utilizado por su capacidad de manejar interacciones no lineales y evitar el sobreajuste.

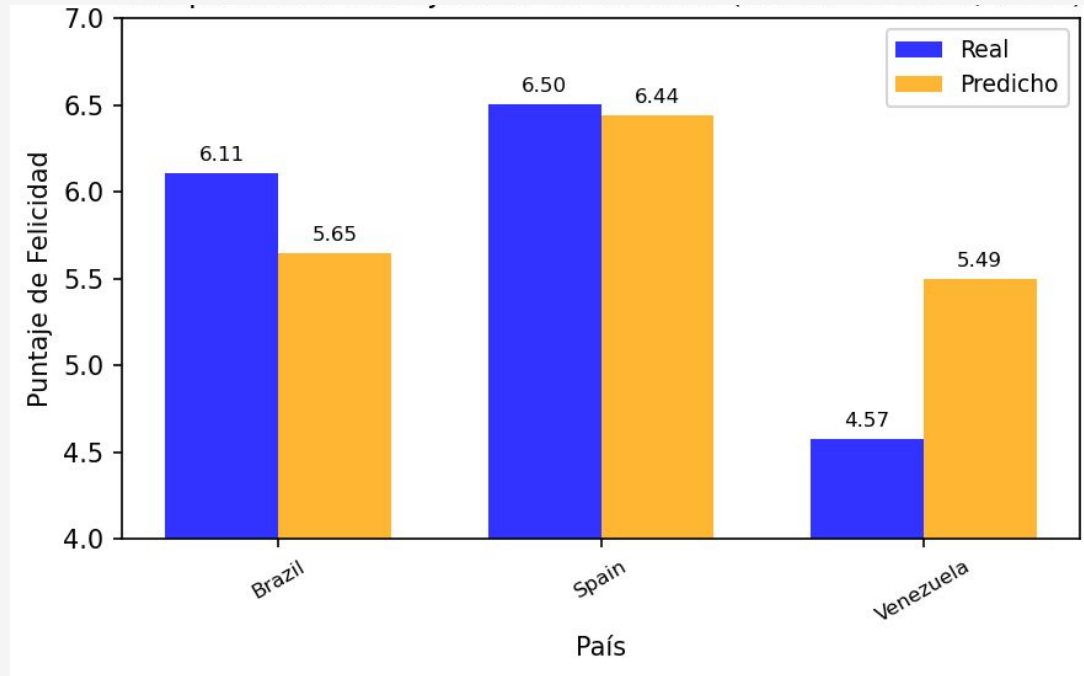
Comparación: Puntaje de Felicidad Real vs Predicha (Regresión Lineal - 2020)



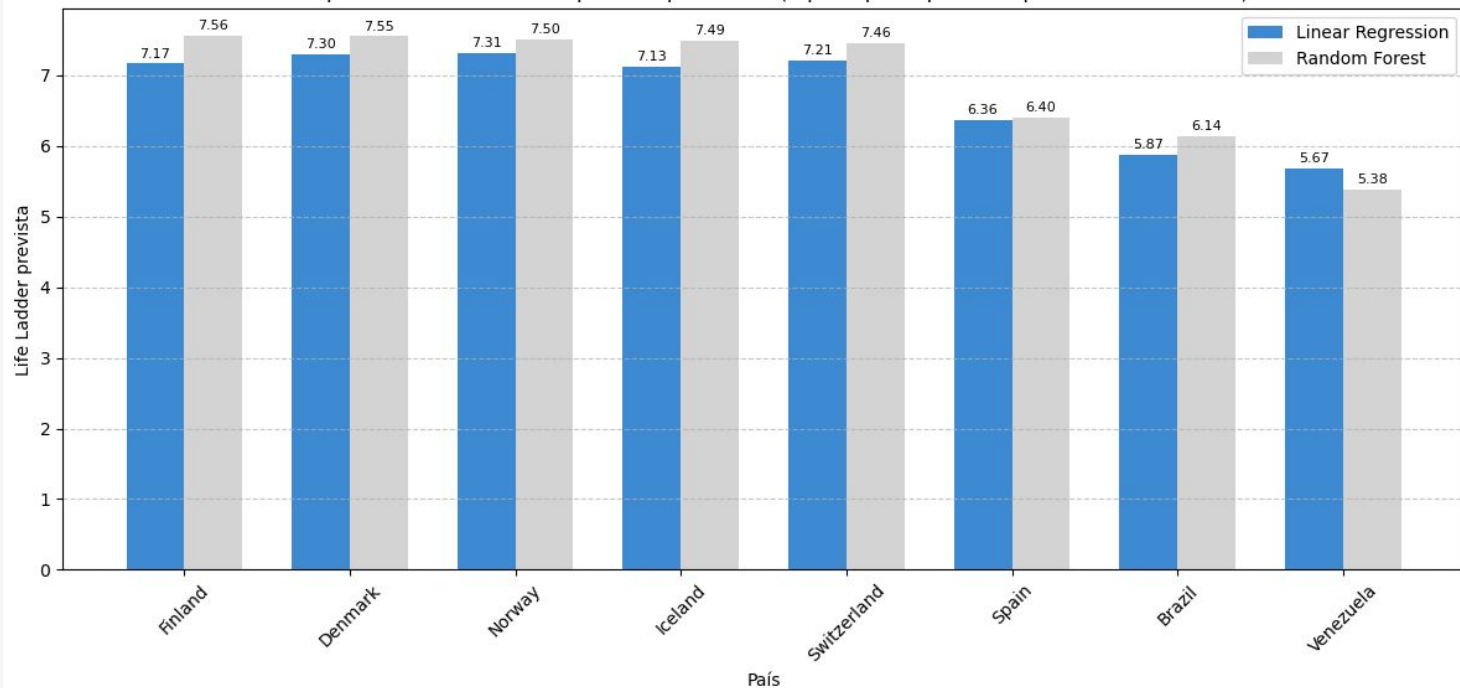
Comparación: Puntaje de Felicidad Real vs Predicha (Random Forest - 2020)

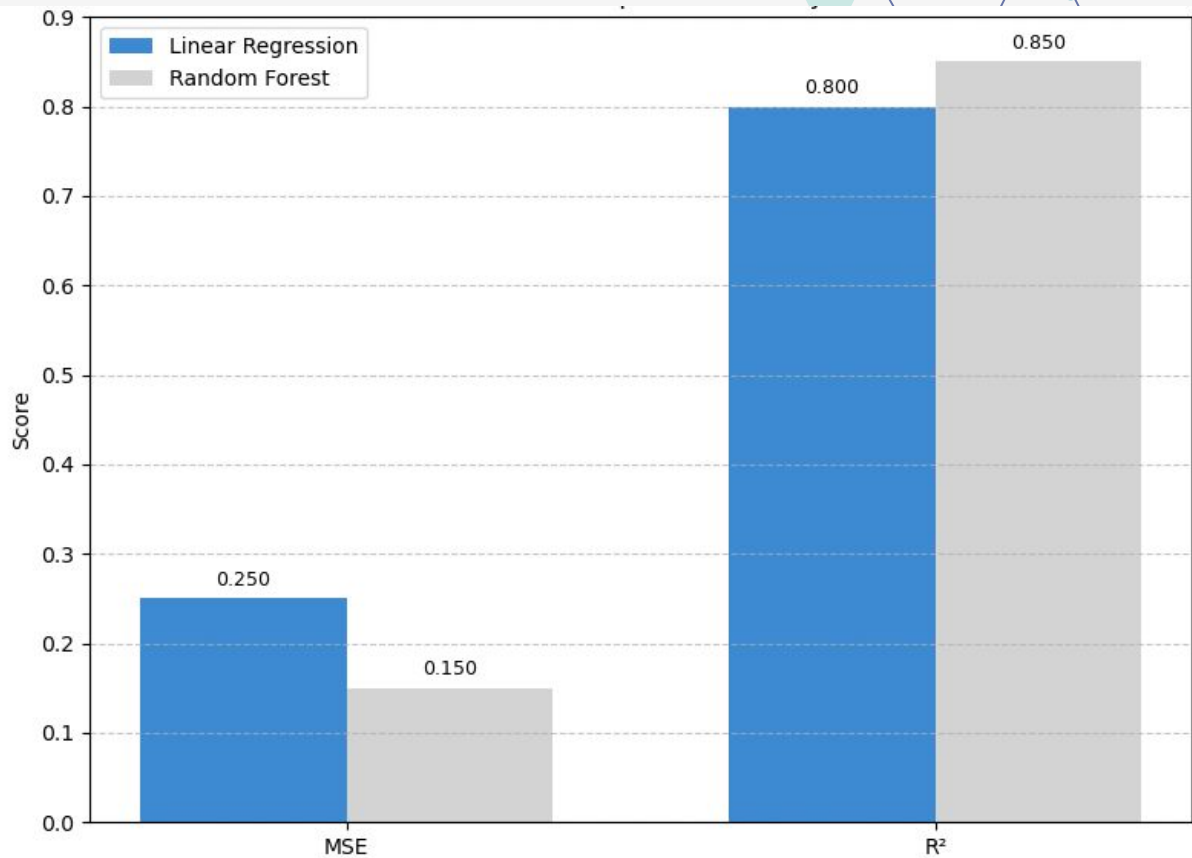


Comparación: Puntaje de Felicidad Real vs Predicha (Random Forest, 2020)



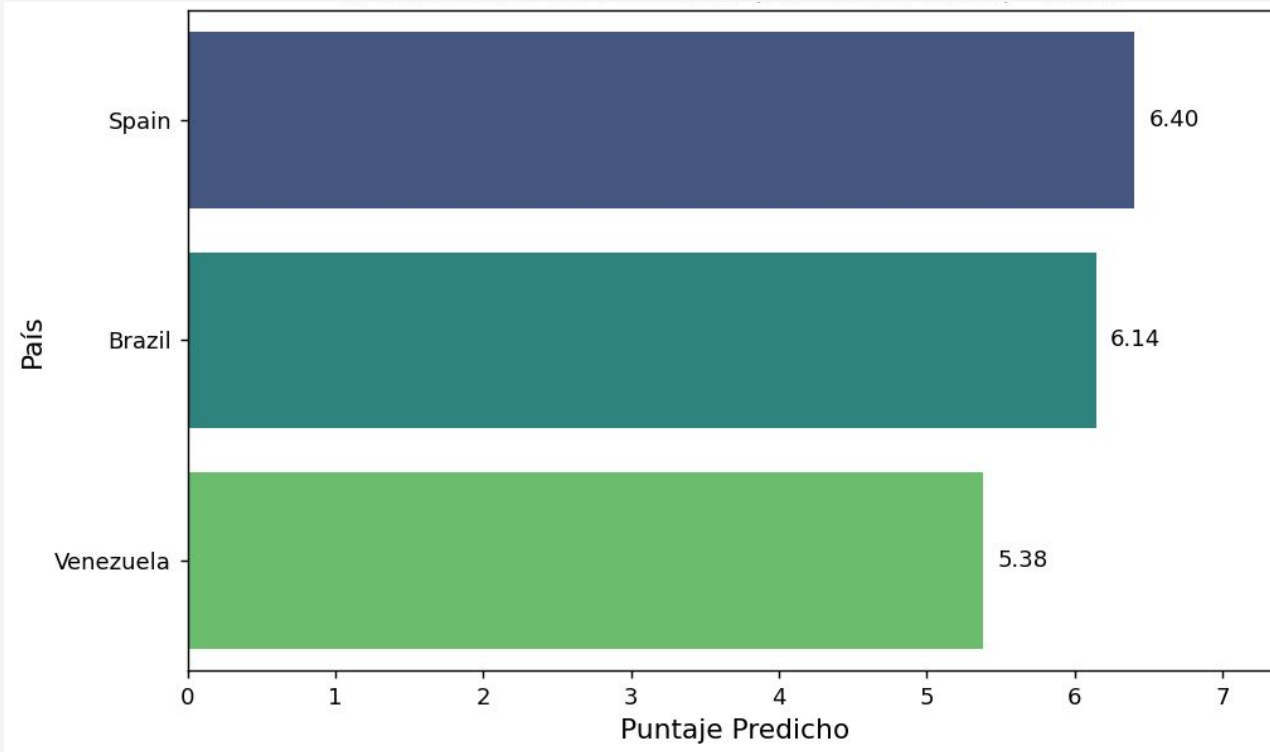
Comparación: Life Ladder prevista para 2025 (5 principales países + países seleccionados: España, Brazil y Venezuela)



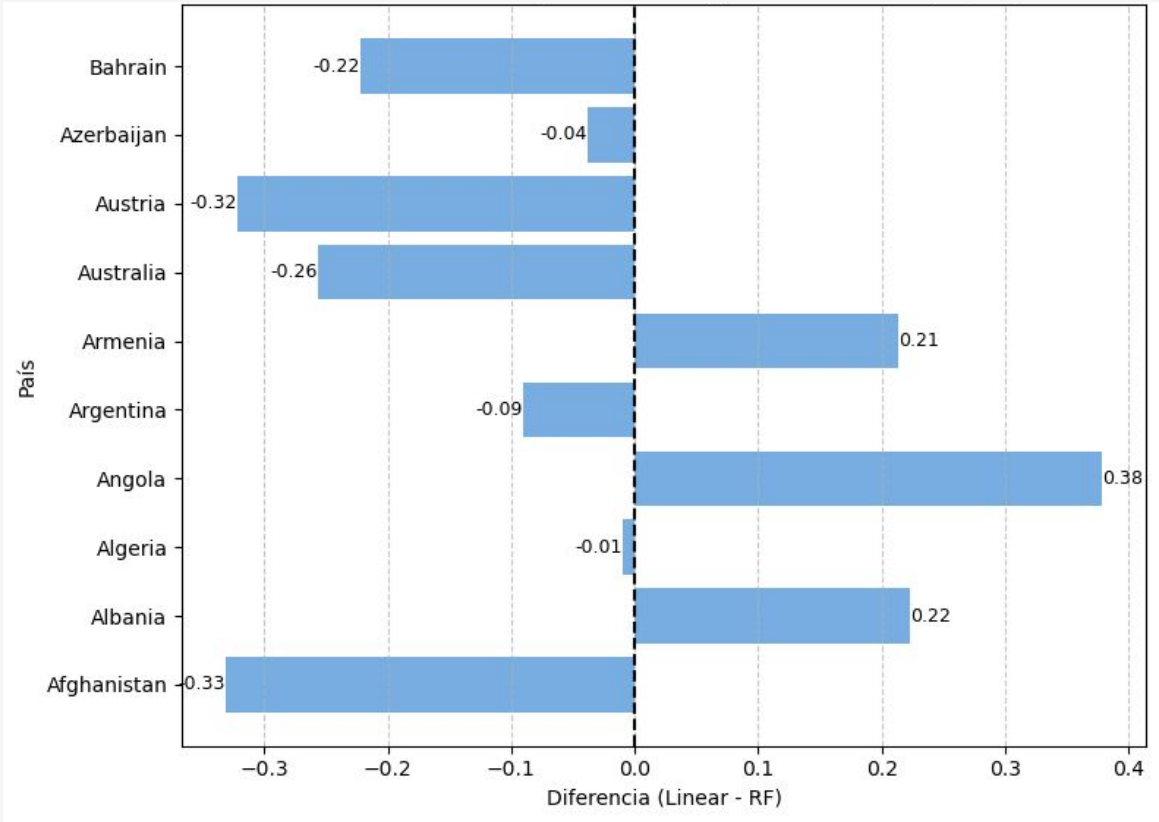


Modelo de Comparación MSE y R2

Predicciones de Felicidad Random Forest 2025 - España, Brazil y Venezuela



Diferencia entre Regresión Lineal y Random Forest (Top 10)



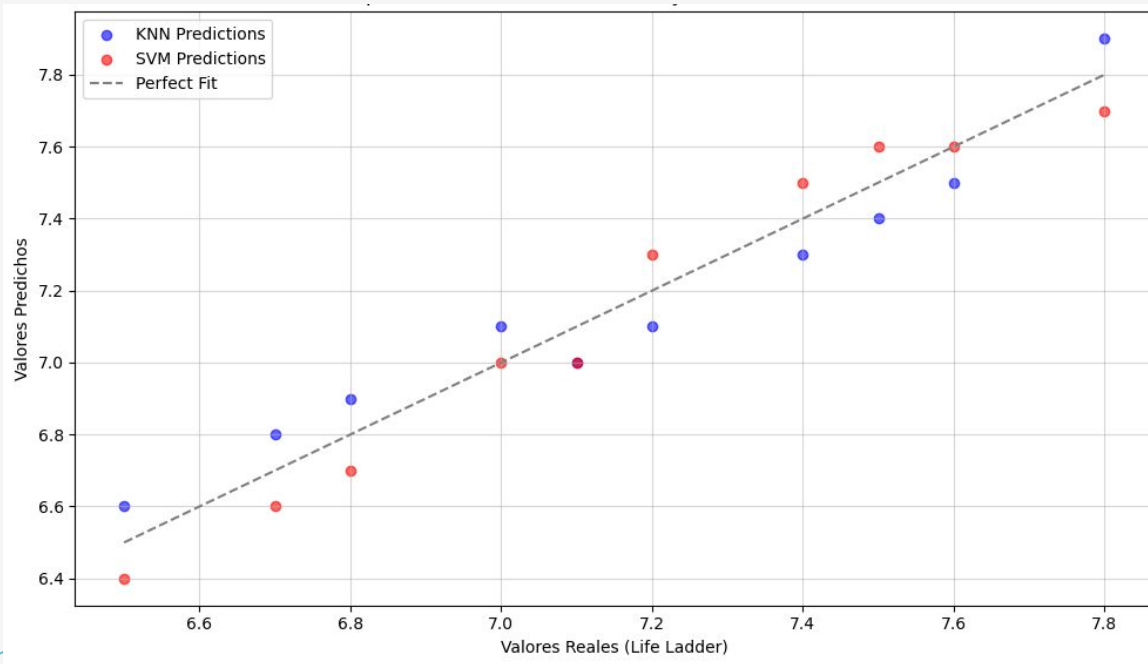
Otros Modelos Supervisados

3. KNN:

- Modelo basado en vecinos más cercanos.
- Ventaja: Captura relaciones locales entre puntos de datos.
- Limitación: Puede no funcionar bien en datasets complejos o con muchas dimensiones.

4. SVM:

- Ventaja: Eficiente para manejar alta dimensionalidad y relaciones no lineales.



Comparación de Predicciones: KNN y SVM vs Valores Reales

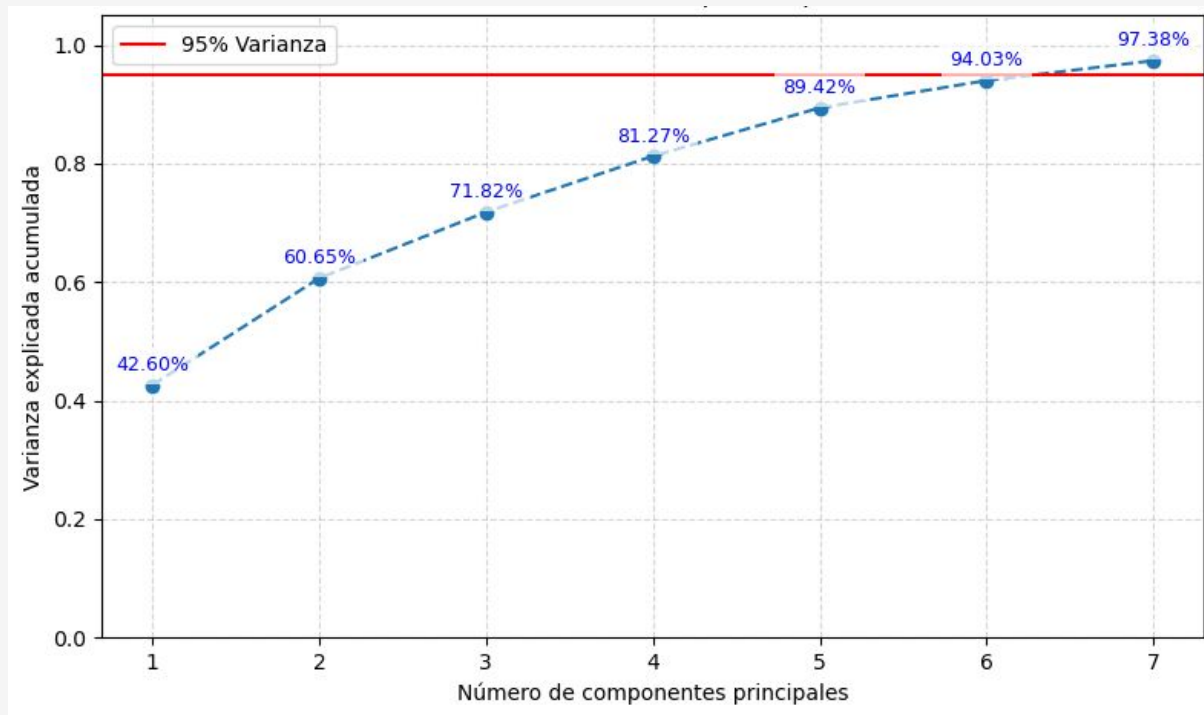
Este gráfico muestra cómo las predicciones de KNN y SVM se comparan con los valores reales, una vez que ambos modelos se acercan a la línea de ajuste perfecto (gris), con ligeras diferencias en precisión, como lo indican las métricas de evaluación.

Reducción de Dimensionalidad

5. PCA (Principal Component Analysis):

- **Ventaja:** Simplifica los datos manteniendo la mayor cantidad de información posible, facilitando la eficiencia de los modelos.
- **Aplicación:** Preprocesamiento previo a modelos supervisados para mejorar el rendimiento y la velocidad de cómputo.

Varianza acumulada explicada por PCA







Arima y Análisis de Series Temporales

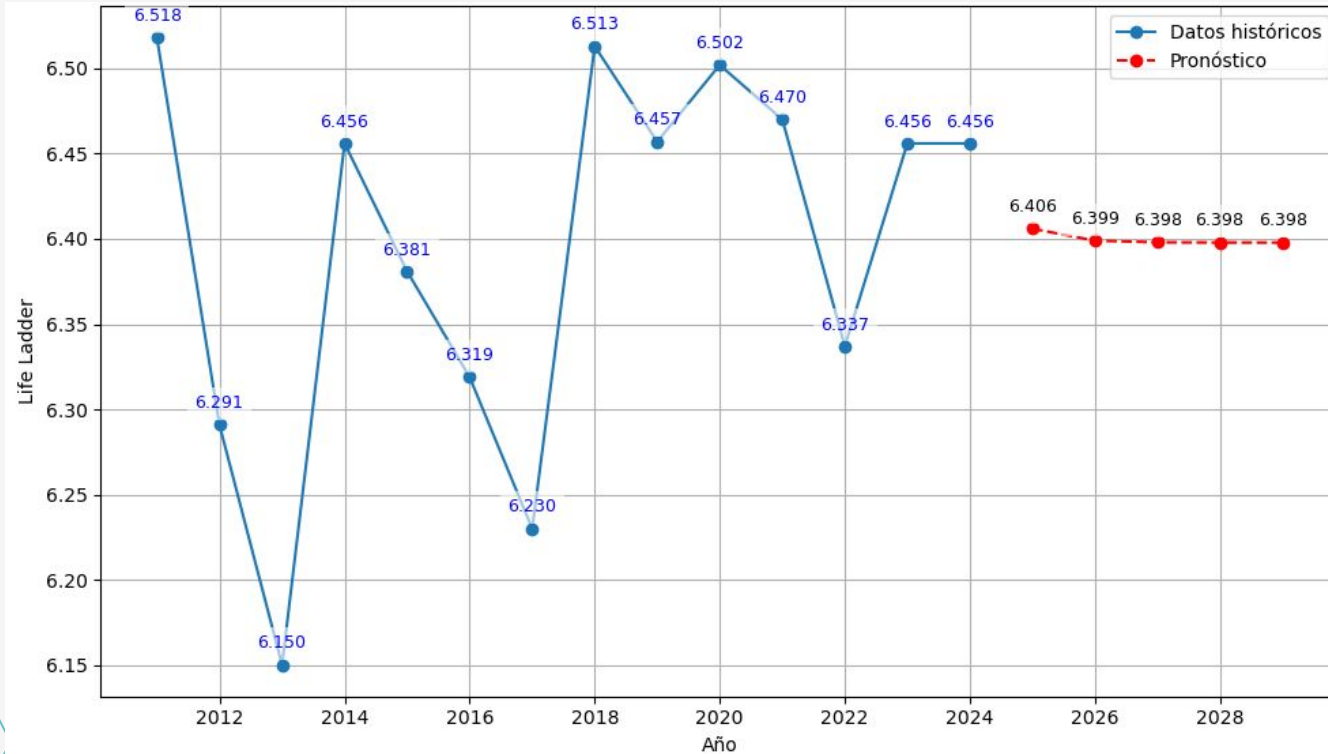
5. ARIMA:

- Usado para modelar tendencias temporales.
- Predicción para los próximos 5 años basada en datos históricos.

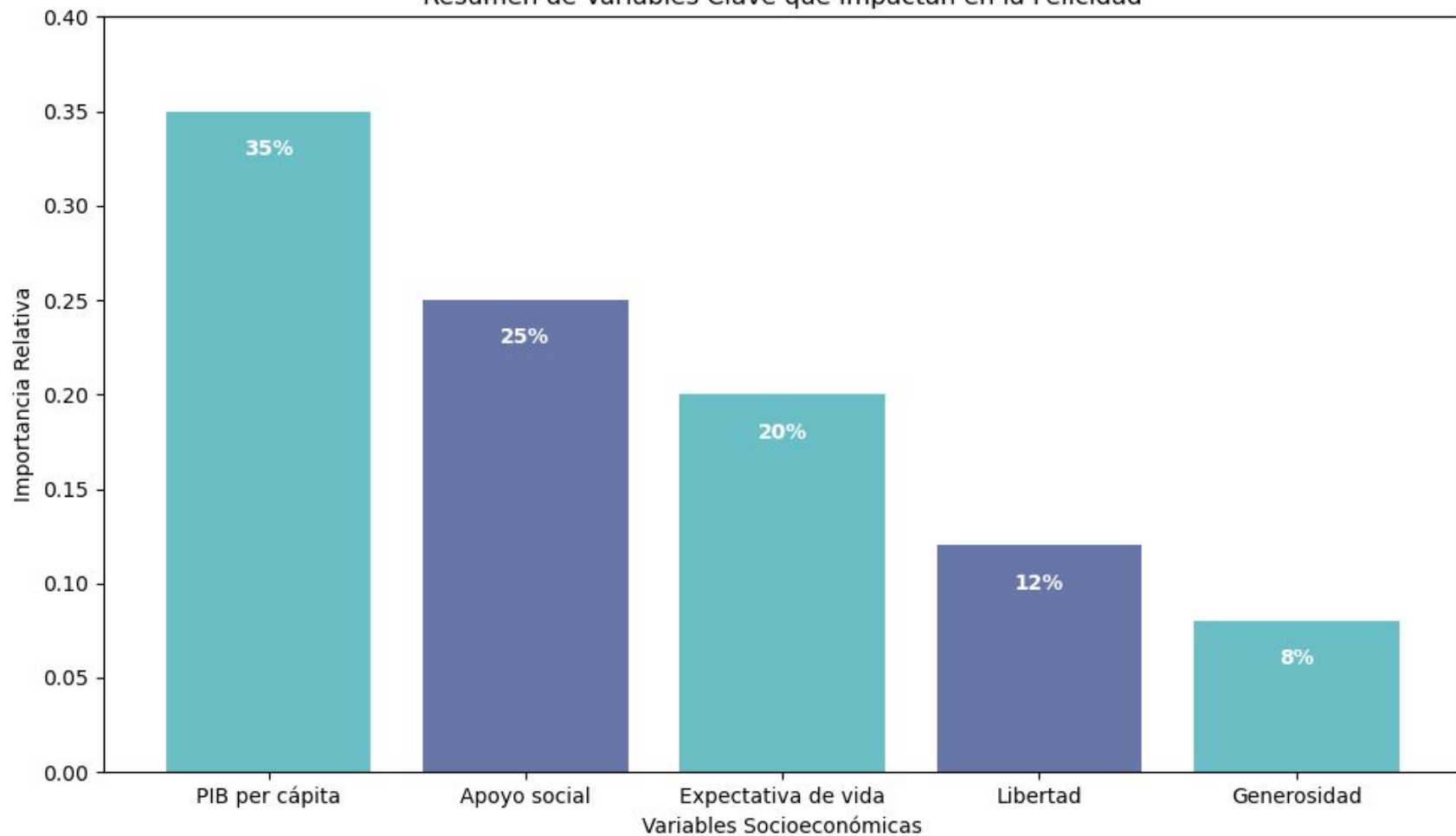
6. Técnicas empleadas:

- Identificación de patrones cíclicos y estacionales.
 - Uso de interpolación para manejar valores ausentes.
- 
- 

Pronóstico de Life Ladder España



Resumen de Variables Clave que Impactan en la Felicidad



Conclusión

Identificación de variables clave:

- *El PIB per cápita, apoyo social y expectativa de vida saludable* son los mayores determinantes de la felicidad.

Tendencias y predicciones:

- Diferencias regionales: estabilidad en países desarrollados y variabilidad en regiones en desarrollo.
- Predicciones de **ARIMA** muestran estabilidad general a lo largo de los años.

Desempeño de modelos:

- **Random Forest** tuvo el mejor desempeño, explicando con precisión las variaciones en los puntajes de felicidad.

Aplicabilidad y próximos pasos:

- Insights útiles para políticas públicas y planificación estratégica.
- Incorporar nuevas variables y explorar redes neuronales profundas.