

Generación y Proyección del Número de Pensionados en los Estados de México

Javier Horacio Pérez Ricárdez

Febrero del 2025

Descripción de la Proyección Matemática

Este modelo tiene como objetivo proyectar el número de pensionados en los estados de México mediante el uso de diversas variables simuladas y un modelo de Machine Learning basado en **Árboles de Decisión para Regresión**.

Generación de las Variables

Las 9 variables utilizadas para la proyección del número de pensionados son generadas de acuerdo con los siguientes criterios:

- **Población(P)**: La población total del estado se genera de manera aleatoria utilizando un rango determinado por los censos nacionales. Se puede simular con una distribución uniforme o normal dentro de un intervalo representativo para el estado específico.
- **Pensionados Actuales(P_0)**: El número de pensionados actuales se genera en base a un porcentaje estimado de la población que está en edad de recibir pensión. Este porcentaje se puede generar de manera aleatoria entre un rango de valores representativos para el estado. Ejemplo: $P_0 = P \times \text{tasa de pensionados}$, donde la tasa de pensionados es un valor aleatorio dentro de un rango.
- **Monto Pensión(M)**: El monto de la pensión asignada a cada pensionado también se genera de manera aleatoria dentro de un intervalo determinado. Este monto puede estar basado en valores históricos o estimaciones oficiales del sistema de pensiones.
- **Inflación(I)**: La tasa de inflación se genera de acuerdo con las expectativas económicas del estado. Se puede simular como un valor aleatorio en un rango basado en las tasas históricas de inflación, generalmente entre un 2% y un 5%.
- **Esperanza Vida(E)**: La esperanza de vida se genera utilizando datos estadísticos actuales y simulados. Este valor depende de factores demográficos y puede ser asignado aleatoriamente dentro de un rango determinado para el estado o país.
- **Crecimiento Poblacional(G)**: La tasa de crecimiento poblacional es generada con base en datos de crecimiento histórico o proyecciones oficiales. Se puede simular dentro de un rango que refleje un crecimiento positivo moderado, entre un 1% y un 3% anual.
- **Pensionados 5 años(P_5)**: El número de pensionados proyectados a 5 años se calcula utilizando la fórmula de crecimiento poblacional y el número de pensionados actuales. Se genera mediante la fórmula:

$$P_5 = P_0 \times \left(1 + \frac{G}{100}\right)^5$$

donde P_0 es el número de pensionados actuales y G es la tasa de crecimiento poblacional.

- **Pensionados_10años**(P_{10}): Similar a la proyección a 5 años, el número de pensionados a 10 años se calcula mediante la misma fórmula de crecimiento poblacional, pero con $t = 10$:

$$P_{10} = P_0 \times \left(1 + \frac{G}{100}\right)^{10}$$

- **Pensionados_20años**(P_{20}): La proyección a 20 años se calcula con la misma fórmula:

$$P_{20} = P_0 \times \left(1 + \frac{G}{100}\right)^{20}$$

Modelo de Machine Learning: Árbol de Decisión para Regresión

Para mejorar la precisión de las proyecciones a largo plazo, se emplea un *Árbol de Decisión para Regresión*, el cual es utilizado para predecir el número de pensionados en función de variables como el crecimiento poblacional e inflación.

Entrenamiento del Modelo

El modelo de árbol de decisión para regresión se entrena utilizando datos simulados de crecimiento poblacional (G) e inflación (I) como variables de entrada. El número de pensionados proyectados ($P_{\text{proyectados}}$) es la variable objetivo.

La fórmula para calcular el número de pensionados proyectados es similar a la usada en las proyecciones iniciales, pero en este caso se incorpora el árbol de decisión para ajustar las proyecciones basadas en la interacción entre G e I :

$$P_{\text{proyectados}} = f(G, I)$$

donde:

- $P_{\text{proyectados}}$ es el número de pensionados proyectados.
- G es la tasa de crecimiento poblacional.
- I es la tasa de inflación.
- f es el árbol de decisión para regresión, que mapea las entradas a la salida proyectada.

Durante el entrenamiento, el árbol de decisión ajusta su estructura dividiendo el espacio de entrada (G, I) en segmentos que minimizan el error cuadrático medio.

Evaluación del Modelo

La evaluación del modelo se realiza utilizando métricas estándar de regresión, como el error cuadrático medio (MSE) y el coeficiente de determinación (R^2).

El error cuadrático medio (MSE) se define como:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

donde:

- y_i es el valor real de pensionados.
- \hat{y}_i es el valor predicho por el modelo.
- n es el número total de observaciones.

El coeficiente de determinación R^2 es:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

donde \bar{y} es la media de los valores reales.

Fórmulas de Proyección

Las proyecciones del número de pensionados a diferentes plazos son las siguientes:

$$P_5 = P_0 \times \left(1 + \frac{G}{100}\right)^5$$

$$P_{10} = P_0 \times \left(1 + \frac{G}{100}\right)^{10}$$

$$P_{20} = P_0 \times \left(1 + \frac{G}{100}\right)^{20}$$

Estas fórmulas calculan el número de pensionados proyectados, considerando un crecimiento poblacional constante durante el período proyectado.