

Tarea 3

Jose Adrian Castillo Sierra

Junio 2022

1. Introducción

La tasa de fertilidad o tasa especial de fecundidad es una variable demográfica que muestra el número promedio de hijos que nacerían por mujer si todas las mujeres vivieran hasta el final de sus años fértiles y dieran a luz de acuerdo con la tasa de fecundidad promedio para cada edad.

La tasa total de fecundidad es una medida más objetiva del nivel de fecundidad que la tasa bruta de natalidad, ya que se refiere a los nacimientos por mujer. Esta variable muestra el potencial de los cambios demográficos en el país.

Un promedio igual a dos hijos por mujer se considera la tasa de sustitución idónea para una población, dando lugar a una relativa estabilidad en términos de cifras totales. Promedios por encima de dos hijos por mujer indican poblaciones en aumento y cuya edad media está disminuyendo. Tasas más elevadas también puede indicar dificultades para las familias, en algunas situaciones, para alimentar y educar a sus hijos y para las mujeres que desean entrar a la fuerza de trabajo. Promedios por debajo de dos hijos por mujer indican una disminución del tamaño de la población y una edad media cada vez más elevada.

En esta tarea se analizarán mediante una técnica de series de tiempo las tasas de fertilidad de los países menos desarrollados junto con los de los menos desarrollados.

2. Desarrollo

2.1. Obtención y Pre-procesamiento de la Información

La información obtenida para la realización de esta actividad proviene del dataset provisto por Our World in Data, una organización no gubernamental la cual transparenta información relacionada a la salud, población, índices de desarrollo, ingresos, etc.

2.2. Series de Tiempo

Las series de tiempo son colecciones de observaciones sobre un determinado fenómeno efectuadas en sucesivos momentos del tiempo, usualmente equiespaciados. Corresponde a una realización de un proceso generador de datos.

En nuestro caso, la serie de tiempo se encuentra descrita por medio de las tasas de fertilidad de cada país desde 1960 hasta el 2019. Principalmente en este análisis nos centraremos en los 10 países con menor calidad de vida y los 5 países con menor calidad de vida.

En la siguiente Figura 1 podemos observar el comportamiento en la tasa de nacimiento de 15 países, diez de estos siendo los que peor índice de desarrollo tuvieron el 2021 y los otros cinco los que mejor desarrollo tuvieron en el 2021, así como mostramos la tasa de nacimiento media a lo largo del tiempo.

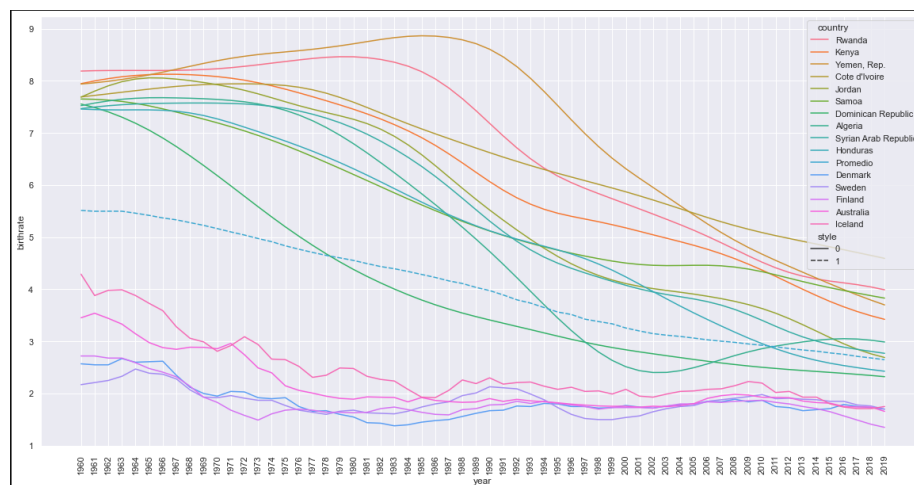


Figura 1: Tasas de Fertilidad

Una de las primeras cosas que podemos notar es que desde 1960 los países con mayor tasa de fertilidad también representan a los países con bajo nivel de desarrollo presentan índices de fertilidad mas altos que los promedio. Países como Rwanda llegando a superar 8 nacimientos por mujer a lo largo de su vida fértil en contraste con países como Suecia donde se ha mantenido cerca de 2 desde 1960.

Un detalle clave es que cada año sin importar el nivel de desarrollo la tasa de fertilidad ha ido disminuyendo, dando a indicar que generación a generación la cantidad de gente joven disminuye año con año.

2.3. Redes Neuronales

Una red neuronal es un modelo simplificado que emula el modo en que el cerebro humano procesa la información: Funciona simultaneando un número elevado de unidades de procesamiento interconectadas que parecen versiones abstractas de neuronas.

Las unidades de procesamiento se organizan en capas. Hay tres partes normalmente en una red neuronal : una capa de entrada, con unidades que representan los campos de entrada; una o varias capas ocultas; y una capa de salida, con una unidad o unidades que representa el campo o los campos de destino. Las unidades se conectan con fuerzas de conexión variables. Los datos de entrada se presentan en la primera capa, y los valores se propagan desde cada neurona hasta cada neurona de la capa siguiente. al final, se envía un resultado desde la capa de salida.

La red aprende examinando los registros individuales, generando una predicción para cada registro y realizando ajustes a las ponderaciones cuando realiza una predicción incorrecta. Este proceso se repite muchas veces y la red sigue mejorando sus predicciones hasta haber alcanzado uno o varios criterios de parada.

En nuestro caso se utilizo una red neuronal para la elaboración de un modelo que se adapte al comportamiento de la información. El uso de una red neuronal para análisis de series de tiempo no es un caso común de uso, ya que existiendo algoritmos menos complejos como el ARIMA pensados específicamente para series de tiempo.

Para que una red neuronal pueda usarse como modelo predictivo tenemos que generar primero una capa de entrada. En nuestro caso se utilizo una capa de entrada de tres neuronas, pero, ¿como obtenemos tres datos que representen un valor en el tiempo? Esto se realiza mediante una descomposición de la serie de tiempo logrando que un valor sea representado por K valores anteriores; generando así un grupo de observaciones con K features. Este procedimiento podemos observarlo en la siguiente Figura 2.

[1, 2, 3, 4, 5, 6, 7, ...]
[1, 2, 3] - 4
[2, 3, 4] - 5
[3, 4, 5] - 6

Figura 2: Descomposición de una Serie de Tiempo

Una vez realizada nuestra capa de entrada en podemos enlazarla con una sola capa oculta y una de salida con una función de activación sigmoïdal. Se utilizara la librería de Keras para realizar el entrenamiento y predicción de la red neuronal.

Cuando el entrenamiento de la red neuronal esta lista podemos ver que tanto se aplica nuestro modelo de red de tiempo a los datos medios de nuestra serie de tiempo. El resultado lo podemos apreciar en la figura 3.

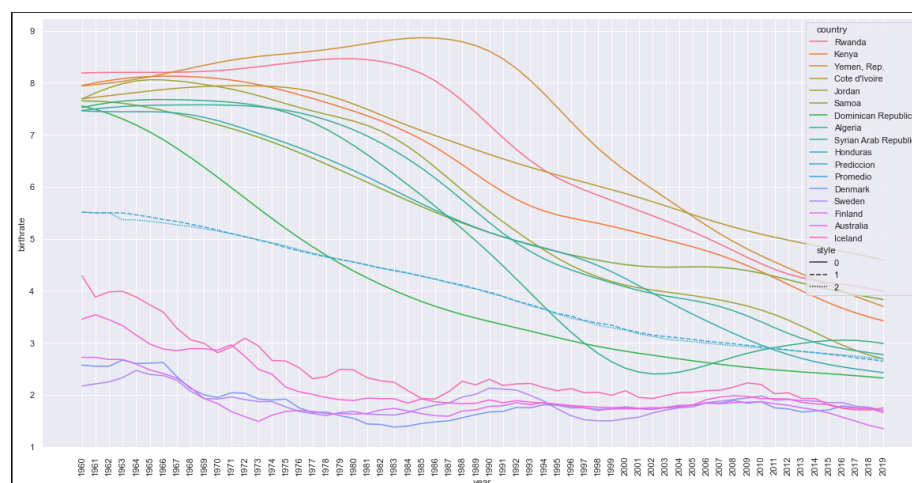


Figura 3: Tasas de Fertilidad Con Predicción

La predicción, representada por la línea punteada, se ajusta mucho a la media de fertilidad de los países. Aunque visualmente se ajuste a nuestro modelo la manera de determinar si nuestro modelo generado tiene un desempeño aceptable es utilizando medidas de desempeño, las utilizadas en este caso son MAE, Mean Square Error, y RMSE, Root Mean Square Error.

El MAE se define como el promedio de la diferencia absoluta entre los valores pronosticados y los reales. El MAE nos muestra cuánta inexactitud debemos esperar del pronóstico en promedio. $MAE = 0$ significa que los valores anticipados son correctos y las estadísticas de error están en las unidades originales de los valores pronosticados. Cuanto menor sea el valor de MAE, mejor será el modelo.

RMSE se define como la raíz cuadrada del promedio de los cuadrados de error. También se conoce como la métrica que evalúa la calidad de un modelo de pronóstico. RMSE también tiene en cuenta la varianza y el sesgo. Cuanto menor sea el valor de RMSE, mejor será el modelo.

Los valores obtenidos para estas medidas siendo una MAE de 0.02404 y un RMSE de 0.035, siendo valores muy pequeños, por lo que podemos concluir

que nuestro modelo de predicción para series de tiempo tiene un desempeño aceptable.

3. Conclusión

El uso de redes neuronales para la predicción de series de tiempo es una alternativa a las técnicas tradicionales como el ARIMA, generando medidas de desempeño aceptables y modelos que se acoplan al comportamiento de los datos.

Referencias

- [1] *Fertility Rate*. URL: <https://data.worldbank.org/indicator/SP.DYN.TFRT.IN?end=2019&start=1960&view=chart%5D>.
- [2] Steelfenix. *Procesamientodatos/Tarea 3 at main · steelfenix/procesamientodatos*. URL: <https://github.com/Steelfenix/ProcesamientoDatos/tree/main/Tarea%203>.
- [3] *Time Series Evaluation Metrics*. URL: <https://analyticsindiamag.com/a-guide-to-different-evaluation-metrics-for-time-series-forecasting-models/>.