

**Red neuronal**

“Personalidad: Extrovertido o Introvertido”

Alumna: Julieta Sofía Gómez

Profesores: Josefina Bompensieri y Tomás Prudente

Matemática III

1° cuatrimestre 2025

**Parte 1** – Análisis de la base de datos

Para este trabajo final, utilicé la base de datos “Personality” de Rakesh Kapilavayi para tener como objetivo el clasificar entre extrovertido o introvertido dependiendo de distintos factores. La base de datos posee 2900 entradas, sin embargo, en algunas columnas no presenta datos así que es útil para practicar qué realizar en estos casos.

**Descripción columnas:**

* “Time spent alone”: Variable cuantitativa discreta, indica cuántas horas se estuvo solo.
* “Stage fear”: Variable categórica, si se tiene pánico escénico.
* “Social event attendance”: Variable cuantitativa discreta, indica cuán frecuente va a eventos sociables.
* “Going outside”: Variable cuantitativa discreta, indica cuán frecuente sale de su casa.
* “Drained after socializing”: Variable categórica, si se siente exhausto luego de socializar.
* “Friends circle size”: Variable cuantitativa discreta, indica la cantidad de amigos cercanos.
* “Post frequency”: Variable cuantitativa discreta, indica cuán frecuente sube algo a las redes sociales.
* “Personality”: Variable categórica, si es Introvertido o Extrovertido.

Ya que esta base de datos posee algunas columnas con datos NaN en ciertas entradas, decidí probar dos opciones: eliminar estas filas o rellenarlas con la mediana de la columna.

Para la primera opción, se perderían 423 datos lo cual representan un 14,59%. Ya que no es una base de datos muy grande es necesario mantener la mayor cantidad de entradas posibles para que nuestra red neuronal tenga más material para utilizar. Al optar por la segunda opción de rellenar la información, las columnas numéricas se llenaron con la mediana y para las columnas categóricas, la moda. Así obteniendo 2900 entradas en todas las columnas de la base de datos sin datos ‘null’

Sin embargo, todavía poseemos información importante para nuestra red en tipo distinto al numérico. Para solucionar esto, utilicé la función ‘.map()’ y así cambiar “Yes/Extrovert” por 1 y “No/Introvert” por 0. Esto ayudó a tener más material necesario para llegar a nuestro valor objetivo.

Ahora que poseemos los 2900 datos numéricos en todas las columnas, comenzamos a analizar si hay alguno que sea atípico.

Texto

El contenido generado por IA puede ser incorrecto.

Aquí se puede observar que no hay gran varianza y solo “Friends circle size” es la única columna que posee una desviación estándar significativa.

Gráfico, Gráfico de cajas y bigotes

El contenido generado por IA puede ser incorrecto.

Luego de realizar el análisis de los rangos intercuartílicos no se hallaron datos atípicos, entonces no se tuvo que hacer una limpieza de la tabla.

En la normalización de los datos decidí utilizar la cuenta en la que Standard Scaler se basa y así traer todos los valores cercanos a 0. Esto ayuda a la red neuronal ya que es más fácil discriminar los valores entre -1 y 1 que entre céntimas.

Texto

El contenido generado por IA puede ser incorrecto.

Con los datos normalizados podemos analizar la correlación que tienen entre sí las columnas y poder ver cuales influyen más sobre nuestro objetivo.

Imagen que contiene Gráfico

El contenido generado por IA puede ser incorrecto.

En este gráfico se puede observar como “Social event attendance”, “Going outside”, “Friends circle size” y “Post frequency” tienen números altos positivos en correlación con “Personality”. Esto tiene perfecto sentido ya que antes determinamos que alguien extrovertido posee el valor de 1 y si estas categorías aumentan, significa que el valor de “Personality” tiende a aproximarse a 1. Así también significa que si “Time spent alone” aumenta, se disminuye la tendencia a 1 llevándolo a 0 significando en alguien introvertido.