



**UNIVERSIDAD DE ANTIOQUIA**  
**FACULTAD DE INGENIERÍA**

Daniel Alexander Basto Moreno – CC. 1030639051  
Kevin Alonso Restrepo García – CC.1216726638  
Angie Viviana Rosales Portilla – CC. 1037661697

*Introducción a la Inteligencia Artificial Para Ciencias e Ingenierías*

**ANÁLISIS DE RASGOS DE PERSONALIDAD MEDIANTE  
COMPORTAMIENTOS ESPECÍFICOS**

**PLANTEAMIENTO DEL PROBLEMA**

Para algunas empresas e instituciones es relevante saber el tipo de personalidad de una persona como parte del proceso de selección, esto con el fin de determinar si es idóneo para desempeñar o no algún cargo específico, por lo que la mayoría de las organizaciones se encargan de diseñar pruebas psicotécnicas con estos mismos fines.

Por otro lado, es de suma importancia para las campañas de marketing tener información acerca de las distintas personalidades de sus potenciales clientes para así poder dirigir dichas campañas a nichos de mercado específicos.

Es sabido que las encuestas o entrevistas pueden llegar a ser costosas y demoradas, es por ello que buscar métodos fiables, ágiles y económicos, es un continuo reto. De lo anteriormente mencionado para las organizaciones es importante mantenerse a la vanguardia en materia tecnológica, por lo tanto, se presume conveniente buscar una alternativa para poder procesar los datos de forma que cumpla los requerimientos mínimos de calidad y que optimice los tiempos en los procesos previamente mencionados.

## DATASET

Se va a utilizar el Dataset de Kaggle [Big Five Personality](#) el cual cuenta con 110 columnas de las cuales existen algunas que se pueden depurar ya que no aportan información al problema que se quiere tratar, como, por ejemplo: fechas, tiempos de respuesta, ubicación geográfica, etc., ya que estas columnas corresponden en su mayoría con respuestas cuantitativas a expresiones de comportamiento de tipo:

- Soy el alma de la fiesta.
- No hablo mucho.
- Me siento cómodo con la gente.
- Cambio mucho mi estado de ánimo.
- Siento poca preocupación por los demás.
- No me interesan los problemas de los demás.
- A menudo me olvido de poner las cosas en su lugar correcto.
- Eludo mis deberes.

Además, se cuenta con 1015341 registros, los cuales se dividirán en 70% y 30% para entrenamiento y validación respectivamente.

## METRICAS DE MACHINE LEARNING

El problema planteado se resolverá a partir de un algoritmo de clasificación, por lo tanto, se usarán métricas de este tipo, las cuales permiten el rendimiento de la clasificación realizada. Se plantean las siguientes métricas:

- **Precisión:** Esta métrica brinda el porcentaje total de elementos clasificados correctamente, lo cual será de suma importancia dado el planteamiento del problema explicado anteriormente. En esta métrica el criterio es la medida más directa de los clasificadores, en cual es un valor entre 0 y 1 donde cuanto más allegado a 1 sea es mejor. Por ejemplo, si se tienen 10 personas con una personalidad específica y el modelo predice que 7 del conjunto total tienen dicha personalidad, pero solo se clasifican bien 5, la precisión entonces es del 0.71.
- **Sensibilidad:** Número de elementos identificados correctamente como positivos del total de verdaderos positivos. Esta es una medida para determinar que tan bien funciona el modelo al momento de detectar verdaderos positivos, con respecto a datos de validación.

Matemáticamente hablando se podría representar de la siguiente forma:

$$S = \frac{TP}{TP + FN}$$

Donde S hace referencia a la sensibilidad, TP a los verdaderos positivos, y FN a los falsos negativos.

- **Matriz de confusión:** Esta matriz es una herramienta para poder evaluar la precisión del modelo, interpretando si se está evaluando de manera correcta cada clase del modelo, en esta matriz tenemos la siguiente distribución:

Verdaderos Positivos	Verdaderos Negativos
Falsos Positivos	Falsos Negativos