

**INTRODUCCIÓN A LA INTELIGENCIA ARTIFICIAL PARA CIENCIAS E
INGENIERÍAS**

ENTREGA 2 DEL PROYECTO

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



**UNIVERSIDAD
DE ANTIOQUIA**

1 8 0 3

Responsables:

DANIEL ALEXANDER BASTO MORENO C.C 1030639051

ANGIE VIVIANA ROSALES PORTILLA C.C 1037661697

KEVIN ALONSO RESTREPO GARCÍA C.C 1216726638

PROFESOR:

RAÚL RAMOS POLLÁN

FACULTAD DE INGENIERÍA
UNIVERSIDAD DE ANTIOQUIA

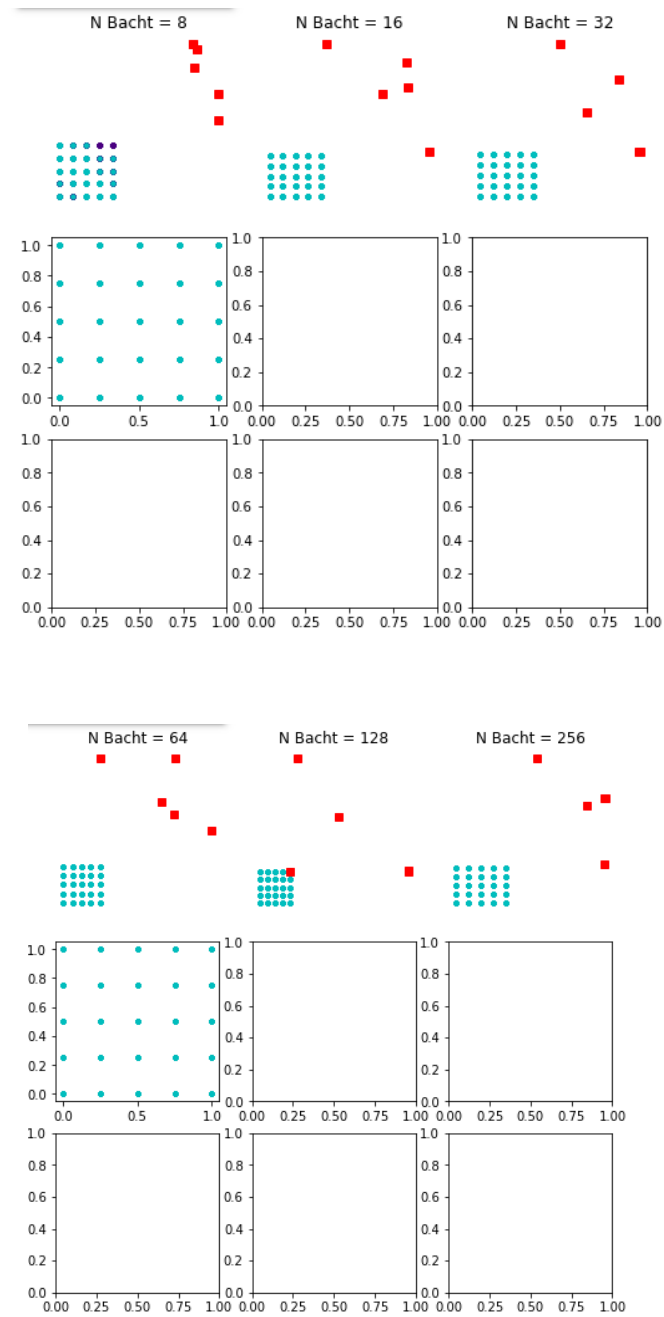
2022

Descripción del avance:

Como primer paso para seguir avanzando en el proyecto se toma los datos usados en el primer avance los cuales era:

- los cinco rasgos de personalidad (Big Five personality traits), así, una persona puede categorizarse como: Extrovertido, Neurótico, Agradable, Consciente o Abierto. Los rasgos de personalidad Big Five , también conocidos como el modelo de cinco factores (FFM) y el modelo OCEAN, es una taxonomía, o agrupación, para los rasgos de personalidad:
 - Abierto a la experiencia (inventivo/curioso vs. consistente/cauteloso)
 - Consciente (eficiente/organizado vs. tranquilo/descuidado)
 - Extroversión (extrovertido / energético vs. solitario / reservado)
 - Agradable (amigable / compasivo vs. desafiante / desapegado)
 - Neuroticismo (sensible / nervioso vs. seguro / confiado)
- La base de datos Luego de haberle hecho tratamiento de datos: eliminación de la columna 'DATELOAD', eliminación de outliers, categorización de datos y rellenado de datos.
- Para el proceso de clustering se definieron las siguientes clases:
 - 0="Extroversion",
 - 1="Neuroticism",
 - 2="Agreeableness",
 - 3="Conscientiousness",
 - 4="Openness"

Paso siguiente, se hace un análisis de cual tamaño de batch es el más conveniente para construir el modelo, teniendo como posibles tamaños 8, 16, 32, 64, 128 y 256. Para ello se decide usar un gráfico de expansión con el fin de visualizar la ubicación de los centros de los clusters o agrupamientos:



Después de ver los resultados se decide trabajar con el tamaño de 32 y 64, debido a que presentan más dispersión con los 5 cluster. Paso siguiente es crear los modelos con dichos tamaños de batch y luego realizar el predict para cada uno de los modelos (sería dos modelos, uno por cada tamaño de bach):

CLUSTERING USANDO MINIBATCHKMEANS

```
[18] #Clustering con la copia de la BD
      #Los cluster van de 0 hasta 4 ----> 0="Extroversion", 1="Neuroticism", 2="Agreeableness", 3="Conscientiousness", 4="Openness"
      #Utilizamos el minibatchKmeans

[19] #Modelo con batch_size = 32
      Modelo_Kmeans_32 = MiniBatchKMeans(n_clusters=5,random_state=0,batch_size=32, max_iter=1000)
      Modelo_Kmeans_32 = Modelo_Kmeans_32.partial_fit(X_1[0:435000,:])
      Modelo_Kmeans_32 = Modelo_Kmeans_32.partial_fit(X_1[435000:866284,:])
      centros = Modelo_Kmeans_32.cluster_centers_
      personalities = Modelo_Kmeans_32.fit_predict(X_1)
      Y_1['Personalidad'] = personalities

[20] Modelo_Kmeans_32
      print(Y_1.head(60))

[21] #Modelo con batch_size = 64
      Modelo_Kmeans_64 = MiniBatchKMeans(n_clusters=5,random_state=0,batch_size=64, max_iter=1000)
      Modelo_Kmeans_64 = Modelo_Kmeans_64.partial_fit(X_1[0:435000,:])
      Modelo_Kmeans_64 = Modelo_Kmeans_64.partial_fit(X_1[435000:866284,:])
      centros = Modelo_Kmeans_64.cluster_centers_
      personalities = Modelo_Kmeans_64.fit_predict(X_1)
      Y_2['Personalidad'] = personalities
```

Se hace la separación de los datos entre los de entrenamiento y los de prueba, siendo la distribución de la siguiente manera:

- 70% de los datos para la creación del modelo (realizado en el paso anterior).
- 20% para pruebas.
- 10% para validación del modelo.

DEFINICIÓN DE VARIABLE DEPENDIENTE E INDEPENDIENTE

```
[ ] Y = Y_1['Personalidad']
     X = Y_1.drop(Y_1.columns[8], axis=1)
     X = X.astype(np.float64)
     X_train,X_set,y_train,y_test=train_test_split(X,Y,test_size=0.3,random_state=42)
     len(X_train)
     len(X_set)
```

Además, se empiezan a realizar pruebas de entrenamiento con un Perceptron Multi-Capa (MLP), y calculamos varias métricas para tener insumos en la calificación del proceso que estamos realizando.