

## Resumen del Trabajo Desarrollado

Este trabajo se centra en el procesamiento y análisis de un **conjunto de datos de imágenes de células** utilizando técnicas de **clustering**. A través de este enfoque, se implementaron métodos de **K-Means** y **MiniBatchKMeans** para realizar el agrupamiento de las imágenes, con el fin de determinar patrones en los datos sin supervisión directa. A continuación, se detalla cómo se abordaron y cumplieron las características clave en relación con los criterios de evaluación.

### 1. Procesamiento y Preparación de los Datos

El primer paso fue **cargar y procesar las imágenes** de células. Las imágenes fueron convertidas a escala de grises, redimensionadas a 64x64 píxeles y luego normalizadas utilizando **StandardScaler** para garantizar que las características tuvieran una escala comparable. Posteriormente, las imágenes procesadas se almacenaron en un archivo CSV para facilitar el análisis y la manipulación posterior de los datos.

### 2. Determinación del Número Óptimo de Clusters

Para elegir el número óptimo de clusters (aaa) en el clustering, se implementaron dos enfoques:

- **Método del codo (SSE):** Se calculó la **Suma de Errores Cuadrados (SSE)** para diferentes valores deaaa(de 1 a 10) y se analizó la gráfica resultante para identificar el "codo", es decir, el punto donde el valor de SSE deja de decrecer significativamente.
- **Coefficiente de silueta:** Se calculó el coeficiente de silueta para diferentes valores deaaa(de 2 a 10). Este coeficiente mide qué tan bien agrupados están los datos dentro de los clusters y ayuda a elegir el número de clusters que maximiza esta métrica.

### 3. Implementación de K-Means y MiniBatchKMeans

Una vez determinado el número óptimo de clusters, se aplican dos variantes de **K-Means** para el clustering:

- **K-Means tradicional:** Utilizado para realizar el agrupamiento con el númeroaaaÓptimo determinado.
- **MiniBatchKMeans:** Una versión más eficiente de K-Means que permite trabajar con grandes volúmenes de datos, donde se entrenó el modelo utilizando minibatches (porciones de datos) en lugar de todo el conjunto de datos a la vez, reduciendo el tiempo de computación.

### 4. Evaluación del modelo

Para evaluar el desempeño del modelo, se utilizó el **coeficiente de silueta** en los datos de prueba. Esta métrica proporciona una medida cuantitativa de la calidad del clustering, indicando qué tan bien se han agrupado los puntos en los clusters. El valor del coeficiente de silueta se comparó con un umbral de 0,8 para determinar si el modelo había alcanzado un rendimiento "excelente". Si el

**NOMBRE:** Rosa Leonor Blanco Moya

**MATERIA:** SIS420

**CARRERA:** Ingeniería de Sistemas

valor era menor, se consideraba que el desempeño era "mejorable". Esta comparación ayudó a establecer la calidad del modelo y la eficacia de la selección del número de clusters.

## 5. Visualización de los resultados

Una vez entrenado el modelo y evaluado su desempeño, se realizaron visualizaciones opcionales para **interpretar los resultados del clustering**. Se utilizó **PCA (Análisis de Componentes Principales)** para reducir la dimensionalidad de los datos y visualizar los clusters en un espacio bidimensional. Esto facilitó la comprensión de cómo se distribuían los puntos en los diferentes clusters y ayudó a identificar visualmente patrones en los datos.

## 6. Comparación con el umbral de rendimiento

Se desarrollará un umbral de rendimiento del **80%** utilizando el coeficiente de silueta. Si el valor era superior a 0.8, el modelo se consideraba de **excelente desempeño**, mientras que los valores más bajos indicaban áreas de mejora. Esta comparación proporcionó una métrica objetiva para medir el éxito del modelo y evaluar su efectividad en el agrupamiento de los datos.

## Cumplimiento de los Criterios de Evaluación

El trabajo ha cumplido con los criterios de evaluación planteados:

1. **Entrenamiento:** Se entrenaron modelos utilizando K-Means y MiniBatchKMeans, y se validaron los resultados con el coeficiente de silueta.
2. **Pruebas:** Se evaluaron los modelos en los datos de prueba, analizando su desempeño con el coeficiente de silueta.
3. **Métricas:** Se utilizaron métricas como el coeficiente de silueta y la SSE para medir la calidad del modelo.
4. **% Desarrollo por encima del 80%:** El modelo alcanzó un rendimiento superior al 80% en términos del coeficiente de silueta, lo que indica que los clusters fueron bien definidos.

## Conclusión

El trabajo desarrollado ha sido exitoso en la implementación de técnicas de clustering no supervisadas para el análisis de un conjunto de datos de imágenes. La evaluación mediante el coeficiente de silueta y el uso de MiniBatchKMeans han permitido obtener un modelo eficiente y de alta calidad. Además, se ha proporcionado un flujo de trabajo que incluye la determinación del número óptimo de clusters, el entrenamiento del modelo, su evaluación y la comparación con un umbral de calidad, asegurando que los resultados estén alineados con los objetivos de rendimiento establecidos.