

Reporte 4

Datos ENIGH

Cluster

Materia: MAA(Aprendizaje Automático)
Profesor: José Anastacio Hernández Saldaña
Grupo: 03
Alumno: José de Jesús Almanza Trejo

Contenido

Introducción.....	3
Clasificación	4
Clúster	4

Introducción

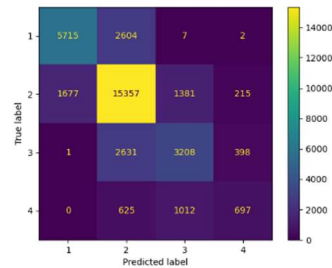
A continuación, se presenta un análisis comparativo entre los resultados de clasificación del mejor modelo obtenido por la tarea anterior donde se utilizará el algoritmo de Clúster KMeans para el mismo número de clústers que la variable objetivo y otro ejercicio donde por medio de la gráfica de codo se seleccionará el mejor valor de clúster que maximice nuestro indicador de accuracy.

Clasificación

El mejor modelo obtenido en la tarea 3 es un modelo de Decision Tree Classifier:

Algoritmo	parámetros	Accuracy Train	Accuracy Test
Decision Tree Classifier	depth=11, min_samples_split=50, min_samples_leaf=50	0.7243	0.7030

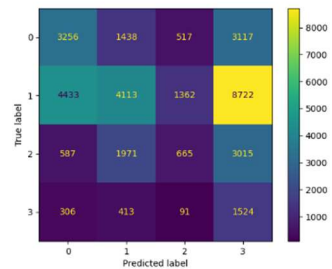
Donde muestra la siguiente tabla de confusión:



Clúster

Se genera un modelo de clúster con n_clusters = 4 para comparar contra los resultados del algoritmo de clasificación:

Algoritmo	parámetros	Accuracy Train	Accuracy Test
KMeans	n_clusters=4	0.3904	0.3925



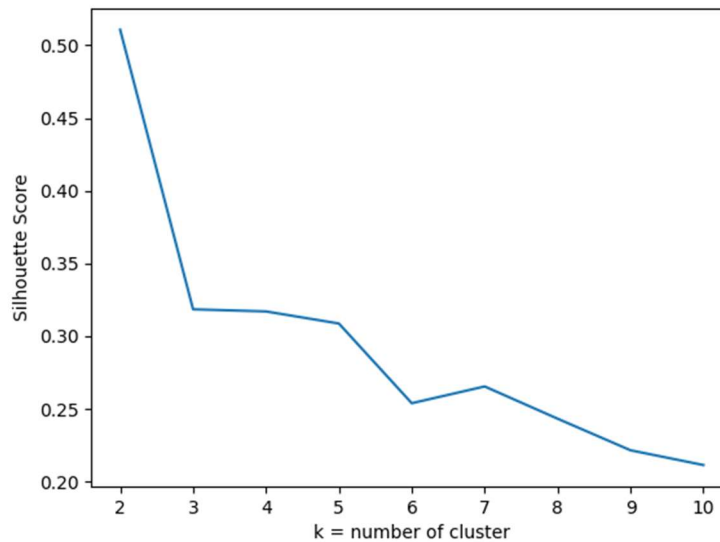
Se reajusta la matriz de confusión dado que los clústeres generados no necesariamente empatan con la definición original, quedando de la siguiente manera:

8722	4433	4113	1362
3117	3256	1438	517
3015	587	1971	665
1524	306	413	91

El accuracy obtenido por este algoritmo no mejora con respecto al modelo anterior.

Mejor Clusterización

Generamos el gráfico de codo donde buscamos el mejor modelo de clusterización de los datos midiendo el score silhouette para $n_clusters = [2,3,4,5,6,7,8,9,10]$



Dado el gráfico de codo con la medida de clasificación interna (Silhouette Score) podemos decir **que la mejor opción para el número de clústers puede ser en 3** al ser el primer punto de inflexión que se observa, sin embargo avanzando en la gráfica observamos otra posible segmentación en 6 clústers donde la medida de clasificación interna tiene una ganancia marginal en los siguientes cortes.