

RÚBRICA DE CALIFICACIÓN

Esta entrega corresponde al "Reporte técnico de selección e implementación de modelos". Cada proyecto conlleva procesos diferentes de selección de modelos, variables y parámetros. Algunas pautas sobre componentes esperados se describen en la rúbrica de la actividad, aunque los equipos pueden justificar sus respectivos procesos según las características de sus proyectos.

1. Propone el desarrollo de análisis, modelos o algoritmos que responden coherentemente a los requerimientos planteados (y estos a su vez responden a preguntas relevantes para el usuario, negocio, o sector), haciendo explícito cómo los modelos descriptivos, predictivos o prescriptivos se integran para cumplirlos.
2. Plantea un conjunto de alternativas de modelos o técnicas para satisfacer cada requerimiento, señalando ventajas y desventajas para incluirlas en el plan de pruebas.
3. Efectúa pruebas de verificación de los supuestos requeridos para usar los modelos elegidos según los tipos de datos (e.g., booleanos) y relaciones entre ellos (e.g., lineal), asegurando el tratamiento previo de los datos según lo requiera cada tipo de modelo.
4. Reporta procesos de entrenamiento y validación de los modelos y calibración de parámetros, incluyendo: selección de variables, conjuntos de entrenamiento y prueba (o cross-validation), parametrización, y definición de métricas adecuadas de evaluación (e.g., justificar escogencia de RMSE, MAPE, R2, precision, recall, F1, AUC/ROC, etc., según el tipo de modelo y contexto del problema).
5. Presenta un análisis conciso de resultados orientado a: (i) seleccionar aquellas alternativas (e.g., modelos) que se ajustan mejor a cada requerimiento; (ii) identificar e implementar ajustes necesarios a los modelos, al tratamiento de los datos, al esquema general de solución, o a los problemas de negocio o de analítica de datos (anticipando los criterios que se deberán satisfacer en fase de evaluación y validación del prototipo).
6. Identifica los componentes, características, o requerimientos del artefacto que están pendientes por implementar (e.g., automatizaciones, generación de reportes, interfaces, tableros, o mejoramiento de los modelos) y, en función de ello, establece un plan para completar la implementación del prototipo funcional, teniendo en cuenta los plazos para efectuar la evaluación del prototipo, y el alistamiento de entregables y presentación ejecutiva del mismo.

REPORTE TÉCNICO DE SELECCIÓN E IMPLEMENTACIÓN DE MODELOS

GRUPO 21

CASO DE NEGOCIO

Reconociendo la importante oportunidad de expandir la adopción digital en Colombia, con un enfoque especial en la oferta de cursos de modalidad virtual del SENA, se ha solicitado al Centro Nacional de Consultoría de Colombia (CNC), referente en investigación de mercados, la realización de un estudio exhaustivo sobre la dinámica de la adopción digital con fines de Educación y Aprendizaje en el país. El propósito fundamental es obtener conocimientos basados en el análisis de datos mediante diversas metodologías, que sirvan de base para la toma de decisiones estratégicas que impulsen significativamente la adopción de la oferta educativa virtual del SENA en todo el territorio nacional.

Este documento se enfoca en comprender los factores clave que influyen en la adopción de la formación técnica, tecnológica y profesional en modalidad virtual impartida por el SENA, así como a explorar la existencia de distintos grupos o patrones de comportamiento por en relación con esta variable que nos ayuden a estructurar una estrategia que impulse la adopción de la educación virtual impartida por el SENA en el territorio nacional.

Con este propósito, se aplicarán técnicas de analítica descriptiva, técnicas de clustering y técnicas de analítica predictiva a través de algoritmos de aprendizaje supervisado, para realizar la estimación de la variable de respuesta considerando ajustes en los predictores de mayor influencia.

Para la identificación de posibles variables predictoras, se realizó una revisión del estado del arte en torno al uso de plataformas de e-learning impulsadas por sector público, desde esa revisión se identificaron las principales limitaciones o propulsores de adopción digital detectados durante el diseño, implementación y monitoreo de las diferentes plataformas digitales evaluadas. Así mismo se realizó una revisión de las diferentes iniciativas de referentes en la industria a nivel internacional tales como Coursera, EdX, Platzi, entre otros, destacando la tendencia conjunta en mejorar la UX - Experiencia al Usuario.

Para el desarrollo del ejercicio analítico, se cuenta con diferentes fuentes de datos como:

- Encuesta de Apropiación Digital 2023 del CNC
- Base histórica de Indicadores TIC en hogares del Dane
- Base histórica del Speedtest Global Index de Ookla
- Tasa de Rebote del website oficial del SENA proveído por Semrush

- Boletín técnico de educación formal para niños entre 6 y 16 años del DANE 2023 por Departamentos.
- Resumen indicadores Educación Superior del SNIES 2023

FASE 1: PREPARACIÓN DE LOS DATOS

CONSTRUCCIÓN DEL CONJUNTO DE DATOS

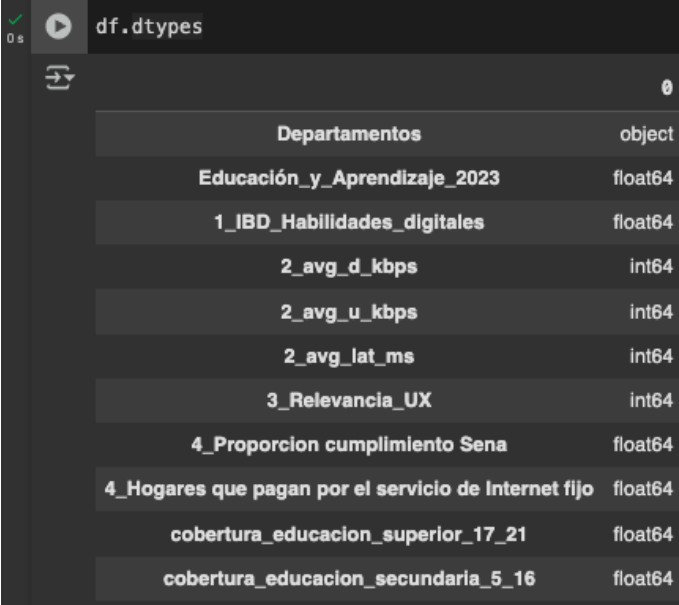
A partir de las fuentes de información mencionadas anteriormente, se construye la base de datos con los siguientes indicadores para cada uno de los departamentos de Colombia:

- **Educación_y_Aprendizaje_2023:** Índice de adopción digital por departamento de internet para actividades educativas y de aprendizaje. (Variable Objetivo)
- **1_IBD_Habilidades_digitales:** Índice departamental de habilidades tecnológicas tomado del Índice de Brecha Digital 2023.
- **2_avg_d_kbps:** Velocidad de descarga promedio por departamento (calidad de conectividad a internet).
- **2_avg_u_kbps:** Velocidad de carga promedio por departamento (calidad de conectividad a internet).
- **2_avg_lat_ms:** Latencia promedio por departamento (calidad de conectividad a internet).
- **3_Relevancia_UX:** Relevancia del UX por departamento. (Tasa de rebote del portal oficial del SENA * Número estudiantes activos en programas virtuales por departamento).
- **4_Proporcion cumplimiento Sena:** Índice de ejecución por departamento en programas de formación virtual (Cupos inscritos / Cupos Meta esperados).
- **4_Hogares que pagan por el servicio de Internet fijo:** Proporción de hogares que pagan una cuota mensual por tener internet fijo en casa.
- **cobertura_educacion_superior_17_21:** Tasa de escolarización de jóvenes entre 17 y 21 años en programas de educación superior.
- **cobertura_educacion_secundaria_5_16:** Tasa de escolarización de niños entre 5 y 16 años en programas de educación primaria, media y secundaria.

Departamentos	1_IBD_Habilidades_digitales	2_avg_d_kbps	2_avg_u_kbps	2_avg_lat_ms	3_Relevancia_UX	4_Proporcion cumplimiento Sena	4_Hogares que pagan por el servicio de Internet fijo	cobertura_educacion_superior_17_21	cobertura_educacion_secundaria_5_16
Vichada	0,868	4225	4934	240	1515	0,8462	0,782	0,0624	0,6763
San Andrés	0,635	20326	9669	69	425	0,8604	0,965	0,2002	0,96
Risaralda	0,509	14555	10103	54	22040	1	0,874	0,614	0,9587
Huila	0,606	13443	7749	59	2692	0,9612	0,849	0,418	0,9052
Guaviare	0,678	10399	5881	239	3284	1,0042	0,833	0,2422	0,7756
Meta	0,555	14196	9235	45	6218	1,0253	0,849	0,3653	0,9243
Caldas	0,542	14749	10008	57	20050	0,8962	0,929	0,5532	0,808
Casanare	0,631	11003	8290	51	790	0,0686	0,867	0,2658	0,9281
Cundinamarca	0,576	12623	8688	48	47773	1,1351	0,912	0,2954	0,9359
Tolima	0,585	15035	8979	48	46953	1,0134	0,9	0,4507	0,9913
Norte de Santander	0,584	14895	10120	66	13652	1	0,835	0,473	0,9116
Antioquia	0,526	15795	9993	59	89957	0,9774	0,939	0,5566	0,9207
Valle del Cauca	0,559	17233	10696	50	24201	1	0,909	0,4521	0,7809

LIMPIEZA DE LOS DATOS

- Garantizamos que todos los campos de las variables descriptivas tengan formato FLOAT o INT según corresponda:



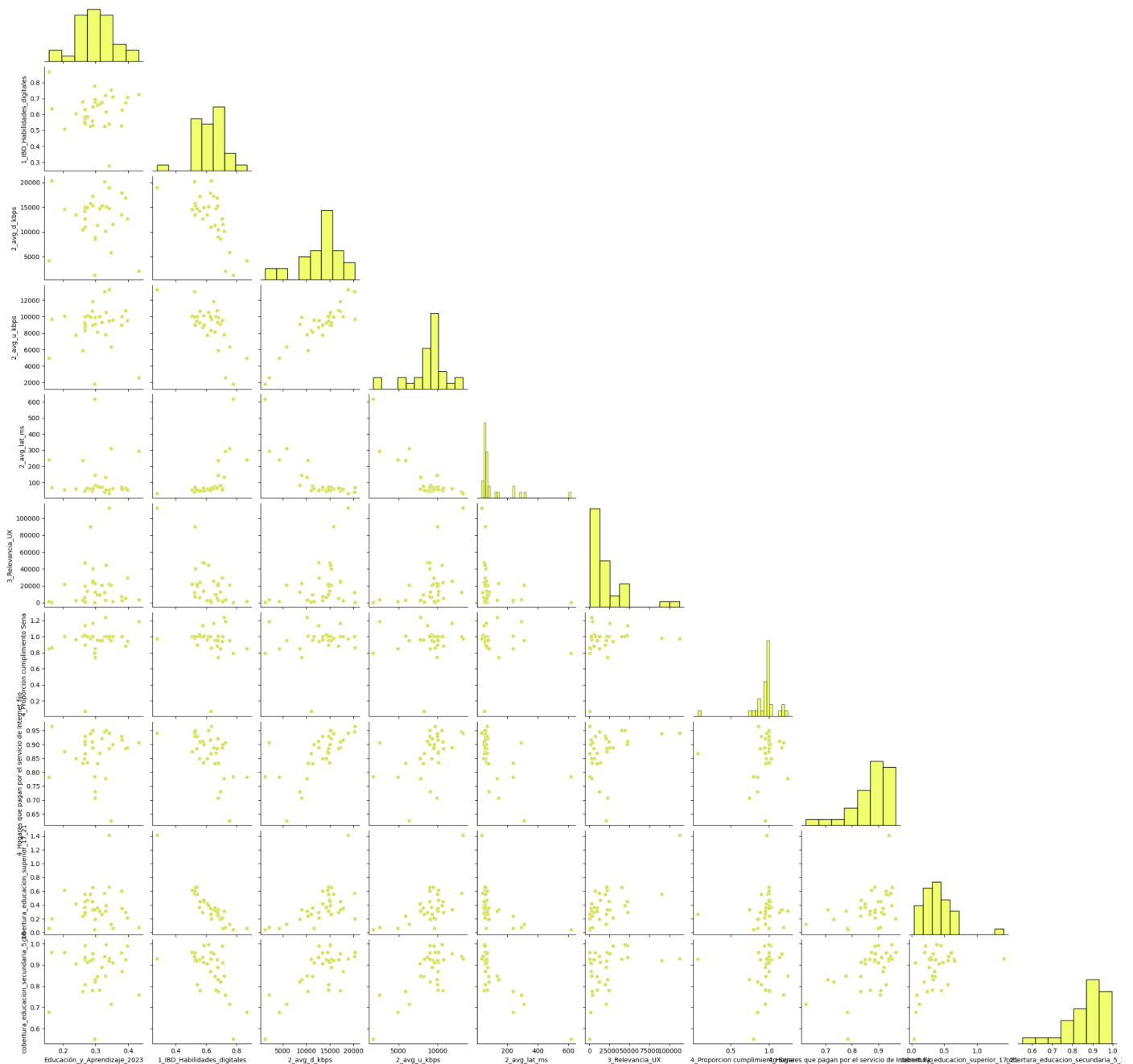
```
df.dtypes
```

Departamentos	object
Educación_y_Aprendizaje_2023	float64
1_IBD_Habilidades_digitales	float64
2_avg_d_kbps	int64
2_avg_u_kbps	int64
2_avg_lat_ms	int64
3_Relevancia_UX	int64
4_Proporcion cumplimiento Sena	float64
4_Hogares que pagan por el servicio de Internet fijo	float64
cobertura_educacion_superior_17_21	float64
cobertura_educacion_secundaria_5_16	float64

- Evidenciamos que a las columnas '3_Relevancia_UX' y '4_Proporción cumplimiento Sena' cuentan con 2 datos faltantes. Para cumplir con el criterio de completitud de la información, imputamos los datos faltantes con el valor de la mediana de la variable correspondiente.

ANÁLISIS DESCRIPTIVO DE LOS DATOS

Distribución de los datos:



En primer lugar, realizamos un análisis gráfico de la información disponible, por medio de la herramienta *sns.pairplot*, la cual nos permite ver de manera rápida cómo se distribuyen las variables individualmente y cómo se relacionan entre sí.

Podemos evidenciar que “Número de latencias”, “Relevancia de la UX”, “Hogares que pagan por servicio de internet fijo”, “Cobertura de la educación superior” y “Cobertura de la educación secundaria” no son variables paramétricas. Así mismo se evidencia posible

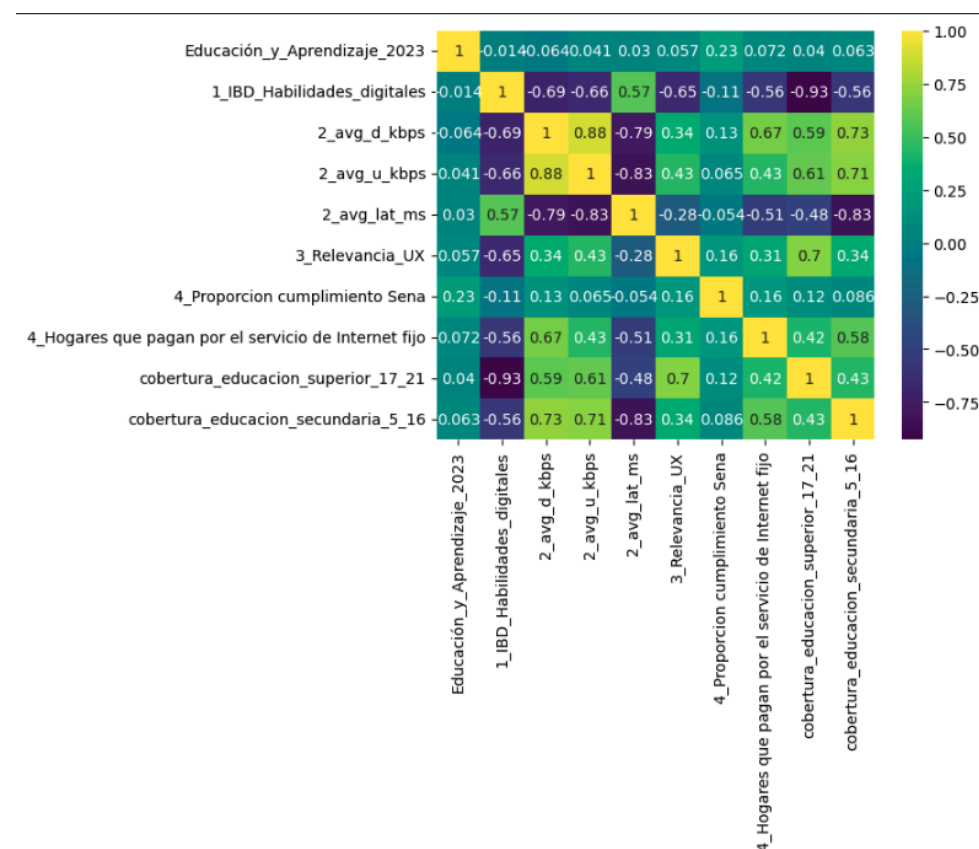
correlación entre algunas de las variables, pero lo confirmaremos a continuación a partir de la matriz de correlación.

Análisis de correlación:

Es posible observar que se cuenta con una diversidad de mediciones que describen diversos aspectos relacionados con la adopción digital (infraestructura, habilidades técnicas, etc); sin embargo consideramos prudente verificar si existen relaciones lineales significativas entre estos atributos para descartar posibles problemas de autocorrelación que puedan afectar o sesgar los análisis futuros.

Para evitar estos inconvenientes, generamos una matriz de correlación y verificamos sus resultados.

Partiendo de un umbral de tolerancia para los coeficientes de correlación de 0.8, se encuentran 3 variables altamente correlacionadas (1_avg_d_kbps, 1_avg_u_kbps, 1_avg_lat_ms) todas asociadas a la medición de la calidad de la conectividad por departamento y las variables cobertura_educacion_superior_17_21, 1_IBD_Habilidades_digitales. Si bien el resultado no es sorpresivo, deja claro que es recomendable eliminar al menos una de estas variables del análisis futuro para evitar redundancia de información y así maximizar la eficiencia de los métodos utilizados.



CONSTRUCCIÓN Y ELECCIÓN DE LOS ALGORITMOS DE REGRESIÓN Y CLASIFICACIÓN

MODELOS DE CLASIFICACIÓN

Estandarización

El último paso del proceso de preparación de los datos es configurar apropiadamente los datos para utilizarlos en los algoritmos implementados. En esta fase se considera necesario aplicar una estandarización sobre las variables explicativas de forma que se logren unificar las unidades de medida y se eliminen posibles imperfecciones o sesgos que puedan influenciar de forma negativa la clasificación.

Para completar este proceso se realiza una estandarización simple de la forma:

$$Z = \frac{x - m}{s}$$

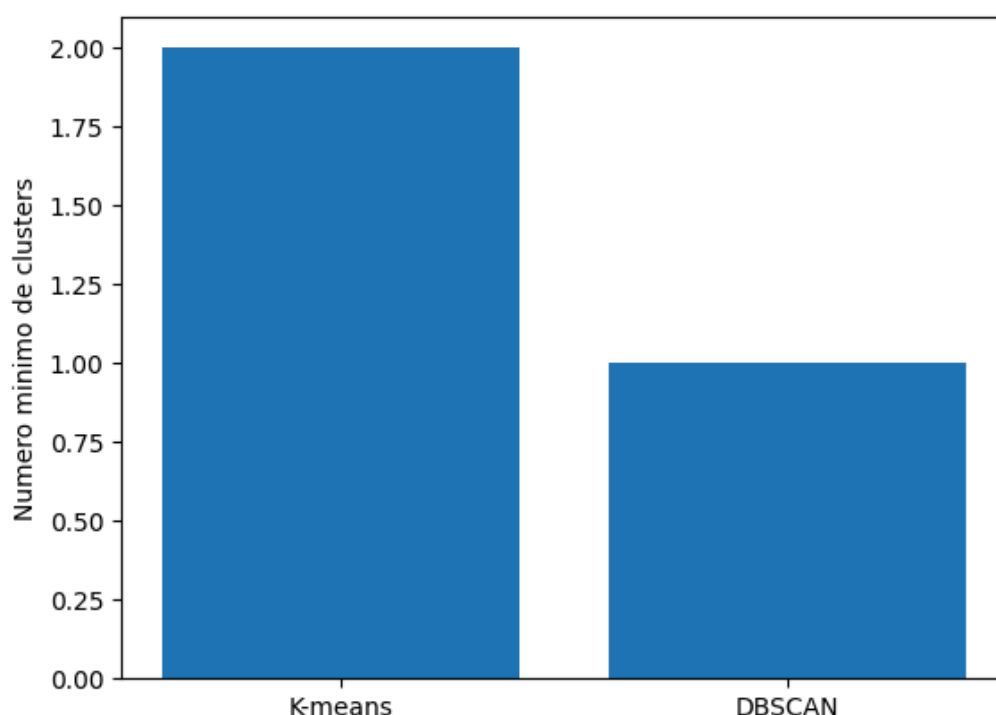
Donde x es cada valor de cada variable, m es la media de la variable y s la desviación estándar. Una vez se ha implementado esta transformación, se procede a la estimación y calibración del modelo predictivo.

Fase 2: Elección del algoritmo de clasificación

Dado que se busca generar una segmentación departamental basada en los índices disponibles, se consideran dos algoritmos de clustering como candidatos principales para desarrollar la clasificación:

- **K-means:** Algoritmo de clustering que clasifica las observaciones de acuerdo a su distancia de la media del cluster (centroide). Ofrece ventajas en términos de eficiencia computacional e interpretabilidad; sin embargo puede ser sensible a datos inusuales (outliers).
- **DBSCAN:** Algoritmo de clustering que clasifica las observaciones de acuerdo a su proximidad (densidad). Ofrece ventajas en términos de adaptabilidad a diversos tipos de datos y al manejo de outliers; pero puede requerir de recursos computacionales avanzados y de procesos exhaustivos de calibración para generar resultados útiles.

El principal criterio de selección será la capacidad de cada algoritmo para adaptarse a los datos disponibles y generar agrupaciones relevantes. Esto permitirá observar el comportamiento de cada algoritmo bajo condiciones similares (sin calibración de parámetros) y analizar su potencial de aplicación para el proyecto.



Es posible observar que sin ningún tipo de condicionamiento previo, el algoritmo de K-means logra segmentar la información mucho mejor que DBSCAN, lo que puede ser un indicio de la capacidad de procesamiento de los datos disponibles de cada algoritmo.

Por todo lo anterior se decide implementar K-means como el método de segmentación principal para encontrar las subdivisiones territoriales de adopción digital en Colombia.

Fase 3: Optimización del algoritmo de clasificación

Para evaluar el desempeño del modelo de segmentación estimado se tendrá como parámetro principal al Silhouette Score, una medida que evalúa el nivel de similitud de cada punto con aquellos que pertenecen a su mismo cluster (valores cercanos a -1 indican una asignación ineficiente de observaciones entre clusters similares; por el contrario valores cercanos a 1 son un indicio de una segmentación exitosa con observaciones asignadas a clusters definidos eficientemente).

Los valores de Silhouette score que guiarán el proceso de estimación y análisis serán los siguientes:

Menor a 0 (Inaceptable)	Entre 0 - 0.5 (Aceptable)	Mayor a 0.5 (Ideal)
-------------------------	---------------------------	---------------------

Con el objetivo de alcanzar los mejores resultados posibles se implementa un proceso de calibración de hiper parámetros que permita comparar múltiples iteraciones del algoritmo de K-means implementando diversas combinaciones paramétricas, de forma que se obtenga

una que maximice la métrica de Silhouette mientras genera una segmentación eficiente para los datos disponibles.

Los hiper parámetros a calibrar serán:

- **Número de clusters:** La cantidad de segmentaciones territoriales que generará el modelo.
- **Método de inicialización:** El método inicial de segmentación de los datos (aleatorio o a partir de la evaluación de una distribución de probabilidad estimada).
- **N_init:** Número de inicializaciones del algoritmo hasta encontrar convergencia
- **Max_iter:** Número de iteraciones máximo por inicialización.
- **Tol:** Nivel de tolerancia máximo que se acepta en las diferencias entre medias (centroides) por cluster.
- **Algoritmo:** Tipo de algoritmo de K-means a implementar

La técnica de calibración y optimización del modelo utilizada para encontrar la mejor combinación posible es GridSearch, que a partir de un conjunto de hiperparametros establecido, genera un proceso iterativo de comparación capaz de encontrar el mejor modelo posible para al algoritmo propuesto y los datos implementados.

Modelo óptimo

El resultado de la experimentación e implementación de los métodos de calibración arrojó un modelo de clustering que segmenta la información en 9 agrupaciones departamentales de acuerdo a sus atributos con un Silhouette score de 0.16 (zona de aceptación).

A continuación una breve descripción de los resultados:

Cluster	Miembros (Dpto)	Descripción
1	Bogotá	<ul style="list-style-type: none">● Nivel más bajo de habilidades digitales.● Altos niveles de calidad de conexión● Alto nivel de Relevancia del UX● Alto nivel de Ejecución del SENA● Nivel más alto de pago por servicio de internet● Tasa de escolaridad en pregrado alta
2	Casanare	<ul style="list-style-type: none">● Nivel medio de habilidades digitales● Niveles de calidad de conexión bajos

		<ul style="list-style-type: none"> Nivel mínimo de Relevancia del UX Bajo nivel de Ejecución del SENA Nivel medio de pago de servicio de internet Tasa de escolaridad en pregrado baja.
3	Amazonas Chocó	<ul style="list-style-type: none"> Alto nivel de habilidades digitales Niveles de calidad de conexión bajos Nivel medio de Relevancia de UX Alto nivel de Ejecución del SENA Nivel medio de pago por servicio de internet Tasa de escolaridad en pregrado baja.
4	Guainía Vaupés	<ul style="list-style-type: none"> Mayor índice de habilidades digitales Peores niveles de calidad de conexión Alto nivel de Relevancia del UX Nivel medio de proporción de cumplimiento del SENA Nivel más bajo de pago por servicio de internet Tasa de escolaridad en pregrado baja.
5	Caldas Huila Meta N. de Santander Risaralda San Andrés	<ul style="list-style-type: none"> Nivel bajo de habilidades digitales Niveles de calidad de conexión altos Alto nivel de Relevancia de la UX Alto nivel de Proporción de cumplimiento del SENA Nivel medio de pago por servicio de internet. Tasa de escolaridad en pregrado cercanos a la tasa Nacional (55%).
6	Arauca Atlántico Boyacá Cesar Córdoba Guajira Nariño Santander Sucre	<ul style="list-style-type: none"> Nivel promedio de habilidades digitales Niveles de calidad de conexión altos Alto nivel de Relevancia de la UX Alto nivel de Proporción de cumplimiento del SENA Nivel alto de pago por servicio de internet Tasa de escolaridad en pregrado cercanos a la tasa Nacional (55%).
7	Caquetá Cauca	<ul style="list-style-type: none"> Nivel alto de habilidades digitales

	Guaviare Putumayo Vichada	<ul style="list-style-type: none"> Niveles de calidad de conexión medio-bajos Alto nivel de Relevancia de la UX Bajo nivel de Proporción de cumplimiento del SENA Nivel medio-bajo de pago por servicio de internet Tasa de escolaridad en pregrado baja.
8	Bolívar Cundinamarca Magdalena Quindío Tolima Valle	<ul style="list-style-type: none"> Nivel promedio de habilidades digitales Niveles de calidad de conexión medio-altos Nivel alto de Relevancia del UX Nivel alto de la proporción del cumplimiento del SENA Nivel alto de pago por servicio de internet Tasa de escolaridad en pregrado cercanos a la tasa Nacional (55%).
9	Antioquia	<ul style="list-style-type: none"> Nivel bajo de habilidades digitales Niveles de calidad de conexión medio-altos Nivel alto de Relevancia del UX Nivel alto de proporción del cumplimiento del SENA Nivel más alto de pago por servicio de internet Tasa de escolaridad en pregrado alta.

De acuerdo con la caracterización de los clústeres, se seleccionaron tres (marcados con gris) que se destacan por su bajo nivel de cumplimiento por parte del SENA. En estos clústeres, la estrategia que se implemente tendrá un mayor impacto, ya que el limitado nivel de ejecución actual representa una oportunidad más amplia en comparación con aquellos donde la meta ya se está cumpliendo.

MODELOS DE PREDICCIÓN - DEFINICIÓN DE LOS FACTORES CON MAYOR RELEVANCIA

Como parte del caso de negocio, buscamos desarrollar un modelo predictivo con alto grado de interpretabilidad con el fin de determinar los factores más relevantes a intervenir como parte de la estrategia. Para ello, se construyeron diferentes modelos aplicando diferentes metodologías analíticas con el fin de compararlas a partir de indicadores como R^2 , MAE (Error Absoluto Medio), RAE (Error Relativo Absoluto), RSE (Error Relativo Cuadrático).

Generación de muestras de entrenamiento y prueba

- En primer lugar eliminamos las variables con alta correlación.
- Separamos las variables predictivas (X) de la variable objetivo (y)

- Generamos las muestras de entrenamiento (Train) del 70% y prueba (Test) del 30%.

```
X =
df2.drop(columns=['Educación_y_Aprendizaje_2023','2_avg_lat_ms','cobertura_educacion_superior_17_21'], axis=1)

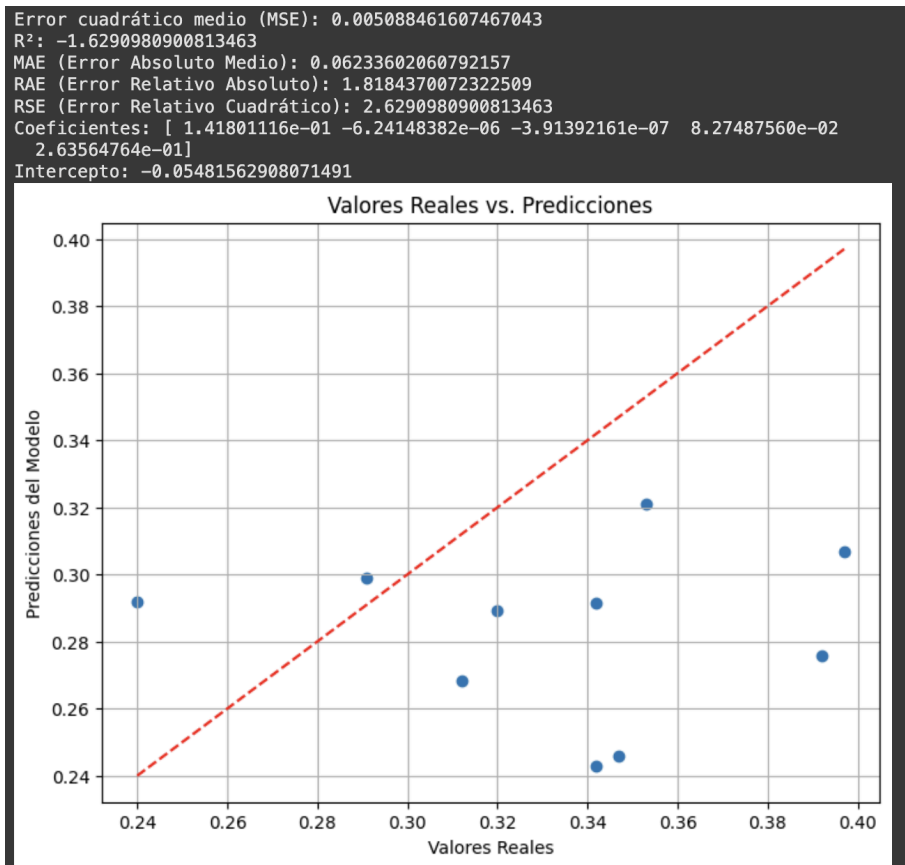
y = df2['Educación_y_Aprendizaje_2023']

X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.3, random_state=42)
```

Fase 2: Elección del algoritmo de predicción

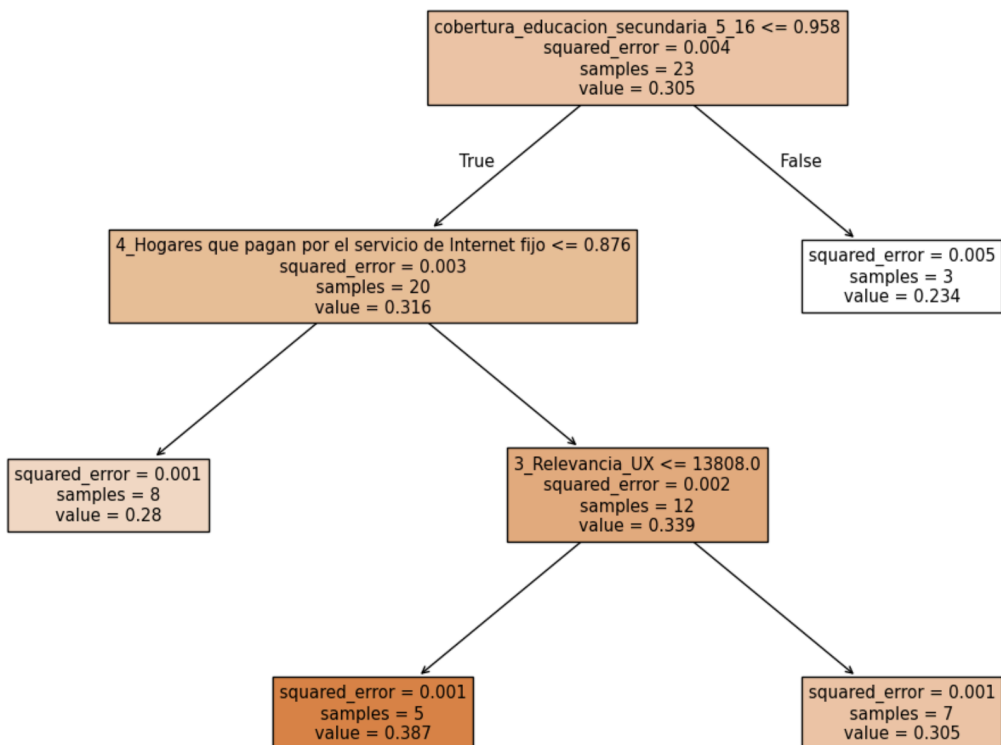
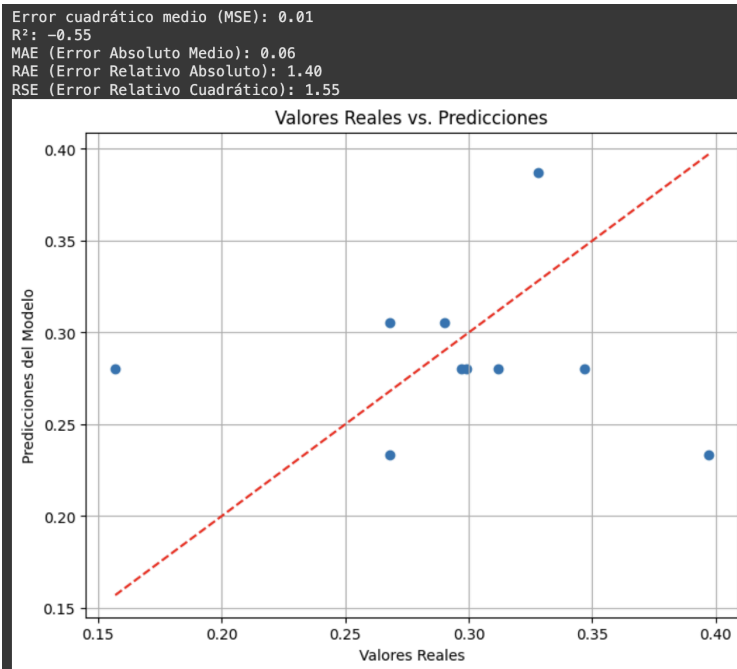
- **Regresión lineal:**

1. A partir de la matriz de correlación, seleccionamos las 5 variables principales.
 - a. 4_Proporcion cumplimiento Sena
 - b. 4_Hogares que pagan por el servicio de Internet fijo
 - c. 2_avg_d_kbps
 - d. cobertura_educacion_secundaria_5_16
 - e. 3_Relevancia_UX
2. Construcción del modelo con el conjunto de entrenamiento.
3. Evalúa el modelo en el conjunto de prueba.
4. Resultados:



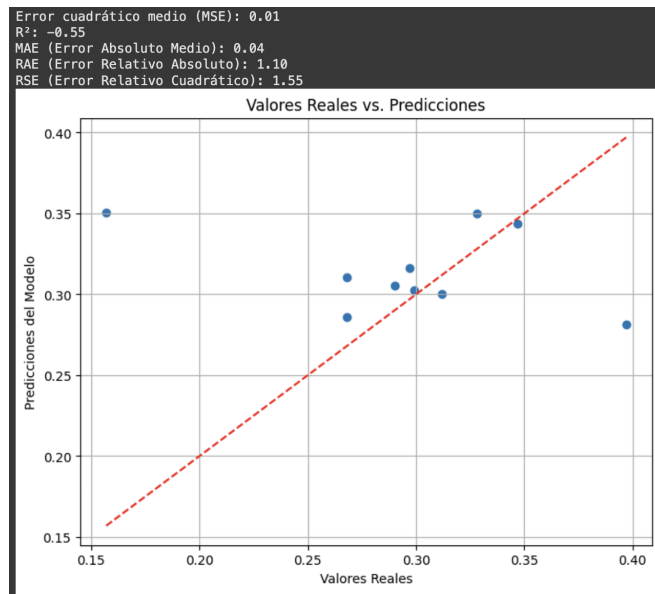
- **Árbol de decisión:**

1. Determinar los mejores hiperparámetros: Mejores hiperparámetros encontrados: 'min_samples_leaf': 3 y 'min_samples_split': 2
2. Construcción del modelo con el conjunto de entrenamiento.
3. Evalúa el modelo en el conjunto de prueba.
4. Resultados:



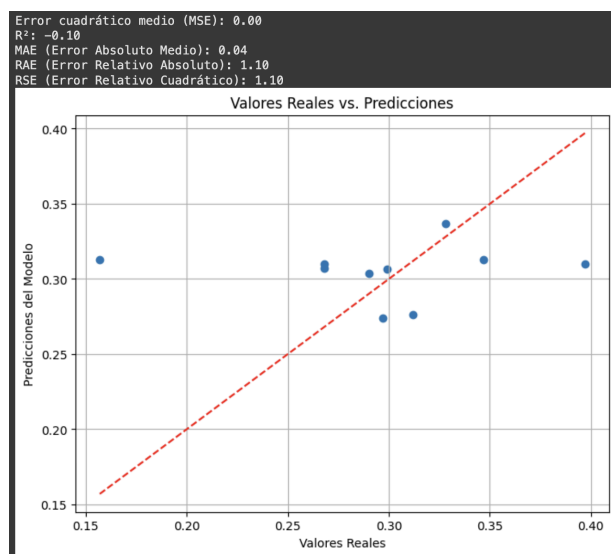
- **Random forest:**

1. Seleccionar las variables más relevantes:
 - 4_Hogares que pagan por el servicio de Internet fijo 0.298783
 - 3_Relevancia_UX 0.286328
 - cobertura_educacion_secundaria_5_16 0.262951
 - 1_IBD_Habilidades_digitales 0.151939
2. Construcción del modelo con el conjunto de entrenamiento.
3. Evalúa el modelo en el conjunto de prueba.
4. Resultados:



- **Support Vector Machine:**

1. Determinar los mejores hiperparámetros: Mejores hiperparámetros encontrados: 'learning_rate': 0.1, 'max_depth': 4, 'min_samples_leaf': 1, 'min_samples_split': 4, 'n_estimators': 10, 'subsample': 1.0
2. Construcción del modelo con el conjunto de entrenamiento.
3. Evalúa el modelo en el conjunto de prueba.
4. Resultados:



Acorde al desempeño de los modelos ambos modelos (Regresión Lineal y Random Forest) arrojan un R2 negativo, lo cual indica que ninguno de los dos se está ajustando bien a los datos. Sin embargo, el modelo Random Forest presenta mejor desempeño de las métricas de evaluación lo cual muestra claramente que éste último capta mejor las no linealidades de nuestro dataframe.

	Random Forest	Regresión Lineal	Mejor modelo
MSE	0.01	0.0051	Reg. Lineal
R2	-0.55	-1.63	Random Forest
MAE	0.04	0.0623	Random Forest
RAE	1.10	1.82	Random Forest
RSE	1.55	2.63	Random Forest

En este sentido, proponemos estructurar una estrategia focalizada en los departamentos previamente seleccionados mediante el análisis de clústeres. Cabe destacar que en estas regiones el pago por servicios de Internet suele ser bajo, lo que se traduce en una calidad de conexión limitada. Este constituye uno de los principales retos a resolver. Por ello, sugerimos evaluar la implementación de una estrategia similar a **Colombia Aprende Móvil**, iniciativa en la cual se estableció un convenio con los proveedores de servicios de Internet para permitir el acceso gratuito al portal educativo. Este enfoque podría resultar altamente efectivo en los territorios priorizados.


Por otro lado, la experiencia de usuario (UX) se posiciona como la segunda variable más relevante según el modelo Random Forest. Por tanto, se recomienda que, especialmente en la prueba piloto en estos departamentos, se diseñe una interfaz responsiva y amigable para dispositivos móviles, que facilite la consulta de contenidos académicos. Asimismo, se sugiere replicar estrategias basadas en microcontenidos de hasta 20 minutos de duración, presentados en formatos diversos (videos, PDFs, tableros interactivos, infografías), con textos de alto contraste y diseño accesible. En general, se busca ofrecer una experiencia atractiva, considerando que los usuarios de estos clústeres se caracterizan por contar con un alto nivel de habilidades digitales.


PRÓXIMOS PASOS

1. Se definirán los municipios y caracterización de los mismos para terminar de darle forma a la prueba piloto de la estrategia propuesta.
2. Desarrollo de Tablero para la visualización de la información recolectada:
 - a. Para responder a los requerimientos y necesidades analíticas, se utilizará un reporte de resultados interactivo desarrollado en Looker. Con esta herramienta, buscamos ofrecer al SENA un recurso clave para optimizar los procesos de toma de decisión en torno a la adopción digital
 - b. Considerando que se busca proporcionar un contexto claro acerca del estado actual de la apropiación digital en Colombia, recurrir a un dashboard proporciona múltiples ventajas:

- i. Visualizaciones intuitivas capaces de proveer información clara y precisa.
 - ii. Diseño simplificado
 - iii. Interacciones eficientes entre usuario y herramienta
 - iv. Manejo de datos eficiente y centralizado.
- 3. Monitoreo y recalibración de los modelos construidos:
La recalibración de los modelos construidos se llevará a cabo de forma continua, incorporando la información más reciente proveniente tanto de las actualizaciones de los informes gubernamentales (nuestras fuentes de información actuales) como de nuevas fuentes relevantes que se identifiquen. El objetivo principal será preservar sus propiedades estadísticas y optimizar su rendimiento.
- 4. Plan de implementación:
 - a. Tablero de visualización: 2 trimestre del 2025.
 - b. Monitoreo y recalibración: 1 trimestre del 2026.

ANEXOS

 20250420 Modelo Predictivo proyecto Final.ipynb

 Clustering.ipynb