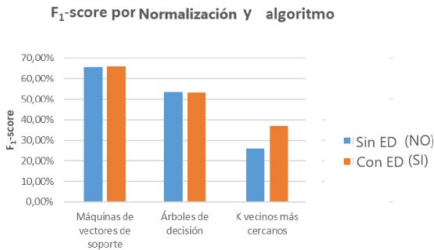


Universidad Nacional de Colombia
Sede Medellín

Figura 1: Casos de uso de las técnicas de aprendizaje frente al preprocesamiento.

Basado en esta tabla generar las siguientes gráficas y realizar un análisis de los resultados mostrados en la figura:

- Diagrama de barras con el F1-Score máximo por algoritmo(general): una grafica con cuatro barras que representan el F1_score de cada algoritmo.
- Diagrama de barras con el F1-Score medio por normalización y algoritmo: se podría ver así (falta el caso de Red Neuronal):



- Diagrama de barras con el F1-Score medio por manejo de datos atípicos y algoritmo
- Diagrama de barras con el F1-Score medio por balanceo y algoritmo

Generar en su informe un análisis de los resultados observados, identificando patrones y posibles causas del comportamiento de cada modelo ante las diferentes condiciones y generar conclusiones de estos resultados.

- Para el caso 8 de la figura 1, obtener las curvas ROC & Precision-Recall donde se comparen las 4 técnicas de aprendizaje de máquinas. Generar en su informe un análisis de los resultados observados y generar conclusiones de estos resultados.

2.4. Visualización de resultados

- Para cada uno de los casos de la figura 1, realizar:
 - Árboles de decisión
 - Obtener la importancia de características
 - Gráfico del árbol de decisión solución usando el criterio Gini
 - K-Vecinos más cercanos
 - El gráfico de los vecinos para KNN una vez entrenado el modelo
 - Máquinas de Vectores de Soporte (SVM).
 - Las regiones de decisión asociadas a su respuesta

3. APRENDIZAJE NO SUPERVISADO

3.1. Dataset y preprocesado.

- Utilizar el mismo dataset del punto anterior y elimine la variable objetivo.
- Preprocesamiento: Generar las diferentes versiones del dataset detalladas en la figura 2.

#	Normalización	OUTLIERS	BALANCED
1	CC(S) y ED(NO)	NO	NO
2	CC(S) y ED(NO)	NO	S
3	CC(S) y ED(NO)	S	NO
4	CC(S) y ED(NO)	S	S
5	CC(S) y ED(S)	NO	NO
6	CC(S) y ED(S)	NO	S
7	CC(S) y ED(S)	S	NO
8	CC(S) y ED(S)	S	S

Figura 2. Datasets para aprendizaje No supervisado.

3.2. Entrenamiento con K-Means y DBSCAN.

- Para cada caso de la figura 2, realizar el entrenamiento y la generación de 3 casos de prueba para cada una de las técnicas de aprendizaje NO supervisado (K-Means y DBSCAN).
- Para cada una de las configuraciones del dataset de la figura 2, entrenar los modelos de aprendizaje No supervisado así:
 - Para K-Means hacer uso de la técnica del codo y el análisis del silhouette score para determinar la cantidad óptima de clusters. Para ello igualmente hacer uso de las métricas vistas en clase para evaluar la calidad del agrupamiento: inercia y silhouette score.
 - Para DBSCAN hacer uso del procedimiento visto en clase para determinar la configuración adecuada del mínimo número de vecinos (MinPts) y la distancia máxima entre dos puntos para que se consideren vecinos (Épsilon ε).

3.3. Evaluación de los modelos generados

- Para cada una de las técnicas de aprendizaje NO supervisado, obtener las métricas: Inercia y Silhouette_Score y elaborar la tabla de la figura 3

#	Normalización	OUTLIERS	BALANCED	K-MEANS		DBSCAN
				INERTIA	SILHOUETTE SCORE	SILHOUETTE SCORE
1	CC(S) y ED(NO)	NO	NO			
2	CC(S) y ED(NO)	NO	S			
3	CC(S) y ED(NO)	S	NO			
4	CC(S) y ED(NO)	S	S			
5	CC(S) y ED(S)	NO	NO			
6	CC(S) y ED(S)	NO	S			
7	CC(S) y ED(S)	S	NO			
8	CC(S) y ED(S)	S	S			

Figura 3: Casos de uso de las técnicas de aprendizaje No supervisado frente al preprocesamiento.

Basado en esta tabla de la figura 3 generar las siguientes gráficas y realizar un análisis de los resultados mostrados en la figura:

- Diagrama de barras con el Silhouette_Score para todos los casos por algoritmo(general): 2 graficas una para k-means y otra para DBScan cada una con 8 barras de resultado.

Generar en su informe un análisis de los resultados observados, identificando patrones y posibles causas del comportamiento de cada modelo ante las diferentes condiciones y generar conclusiones de estos resultados.

3.4. Visualización de resultados

- Visualizar los resultados de los clusters mediante **gráficas en 2D**.
 - Analizar e interpretar los clusters obtenidos, discutiendo si tienen sentido desde el punto de vista del dominio del problema.

4. ENTREGABLES

4.1. Documento de diseño (PDF):

- Descripción detallada de cada componente: Aprendizaje Supervisado y Aprendizaje No Supervisado.
- Particularmente:
 - Para cada uno de los componentes se debe incluir los análisis de los resultados solicitados en cada una de las secciones y generar unas conclusiones generales sobre los mismos.
 - Diagrama Solicitados en el análisis de cada técnica de aprendizaje supervisado y no supervisado.
 - Enlace del video tipo pitch de la sustentación que se detalla en el numeral 5 (sustentación).

4.2. Código fuente documentado en Python (Google Colab):

- Código fuente separado por módulos: aprendizaje supervisado y aprendizaje no supervisado.
- Comentarios por función y explicación del flujo general

5. SUSTENTACIÓN

-
- Se realizará de manera virtual.
- Se deberá implementar un video tipo pitch (duración máx. 5 minutos) con la demostración del sistema funcional, la explicación clara del problema y la solución híbrida implementada.
 - Todos los integrantes deben participar en el video.
 - El video se presentará durante la Sustentación, para lo cual deberán tener un buen equipo con internet para su presentación virtual.
 - El video deberá ser subido a YouTube y en el documento de diseño (PDF) a entregar incluir el enlace de este (ver numeral 3.1).
 - Este video deberá estar disponible desde la fecha de entrega del trabajo indicada al inicio de este documento.
- Se realizará una sesión de preguntas (5 min el pitch y 5 min de preguntas).
 - Todos los integrantes deben estar presentes en la sustentación y se podrá seleccionar uno cualquiera para responder preguntas.

6. METODOLOGÍA DE EVALUACION

- **Código y documentación: 70%**
 - Módulo de aprendizaje supervisado: 45%
 - Módulo de aprendizaje no supervisado: 25%
- **Sustentación: 30%**
 - Video tipo pitch: 20%
 - Atención a preguntas: 10 %

7. MATERIAL A ENTREGAR

Se deberá entregar en Google Classroom un archivo ZIP (con el siguiente nombre del archivo: practica1-grupo-**XX**-equipo-**YY**.zip, donde **XX** representa el número del grupo (1 o 2) y **YY** el número del equipo) que contenga lo siguiente:

- **Documento de diseño:** Descripción detallada de cada módulo (Aprendizaje Supervisado y Redes Neuronales, Aprendizaje No Supervisado), junto con la justificación de las decisiones tomadas en cada parte del programa y **el enlace al video de sustentación tipo pitch y al conjunto de datos original**.
- **Código Fuente:** Todo el programa en Python, debidamente organizado y documentado.

ANEXO 1

INSTRUCCIONES GENERALES PARA LA PRESENTACION DE LOS TRABAJOS

- Los trabajos serán presentados en grupos de 3 personas definidos al inicio del semestre (ver grupos en el material de la clase 2).
- No se admiten trabajos similares (incluyendo el planteamiento del dominio), en caso de presentarse programas con códigos similares serán anulados. La totalidad del trabajo solicitado al alumno, en esta evaluación práctica, deberá ser original y propio del autor que lo presenta. El estudiante es responsable de evitar que su material evaluable (código, solución al problema, memorias, etc.) sea accesible a estudiantes de otros grupos. En caso de que se detecten copias por incumplimiento de estas reglas, por acción o inacción, la sanción afectará todos los estudiantes involucrados: quienes copien y quienes hayan sido copiados.
- Los integrantes de cada grupo deberán ser capaces de explicar y responder preguntas sobre el trabajo realizado durante la sustentación del trabajo.
- La evaluación de la práctica está asociada con la sustentación respectiva programada. Para tal fin, la persona que no asista a la sustentación, se considerará que no presentó la práctica en su totalidad (“**no se tendrá en cuenta el código, ni la documentación, ni el video tipo pitch para la evaluación**”).
- El plazo de entrega del material asociado a la práctica es estricto y se indica al inicio de este documento. No se admiten trabajos después de esta fecha. No se considera como entrega aquella que solo contiene código o solo contiene el vídeo o solo contiene el documento de diseño o cualquier forma incompleta del trabajo.
- El código entregado en el archivo .zip en Google Classroom, debe ser el mismo que se ilustra en el vídeo y se explica en el documento de diseño. En caso de no cumplirse, se considera que el código no fue entregado y, por tanto, no se realizó la práctica por parte del grupo.

NOTA1: Si en la fecha de la sustentación no se presenta el video, el trabajo de la práctica se considerará no entregado.

NOTA 2: La sustentación del trabajo se realizará en la fecha indicada y se realizará mediante el vídeo entregado y las preguntas orales que se realizarán a los integrantes de cada equipo después de la presentación del pitch.

NOTA 3: La redacción de este documento de practica puede presentar algunos errores o la ausencia de algunos requisitos específicos, los cuales deberán ser informados por parte de los estudiantes al profesor, para ser aclarados de manera colectiva durante la clase. Estas aclaraciones harán parte de los requisitos de la práctica.