



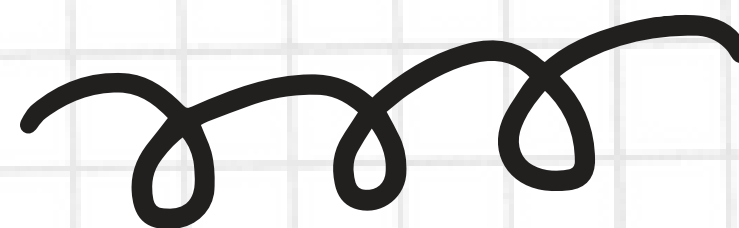
cc5205 - MINERÍA DE
DATOS

GRUPO 1

HITO 2:

¿QUÉ HACE QUE LA MÚSICA SEA MÚSICA?

EXPLORANDO LOS GÉNEROS MUSICALES



PATRICIO ESPINOZA

JAVIERA ROMERO

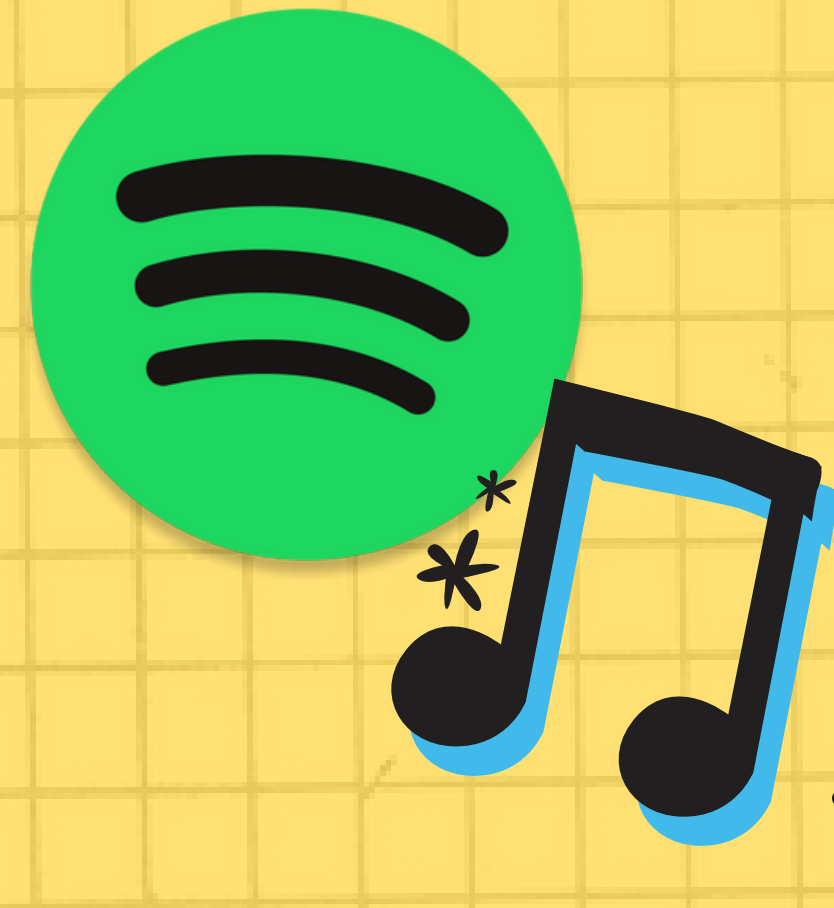
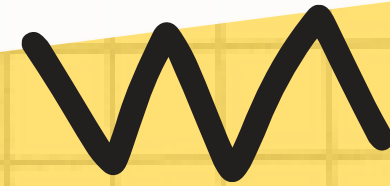
RODRIGO DIAZ

SCARLETT PLAZA

VICENTE THIELE



CONTEXTO



ESTUDIO DE UN DATASET DE ENFOCADO EN CANCIONES DE LA PLATAFORMA SPOTIFY

SE QUIERE INVESTIGAR LA RELACIÓN DE LOS ATRIBUTOS MUSICALES CON EL GÉNERO MUSICAL, PARA ESTO SE BUSCA UTILIZAR TÉCNICAS DE CLASIFICACIÓN Y CLUSTERING CON TAL DE ENCONTRAR RELACIONES INTERESANTES Y DARLE RESPUESTA A LAS PREGUNTAS PLANTEADAS EN EL HITO ANTERIOR

PREGUNTAS HITO 1



- P1. ¿ES POSIBLE PREDECIR EL GÉNERO AL QUE PERTENECE UNA CANCIÓN DE ACUERDO A LAS VARIABLES MUSICALES TALES COMO EL TEMPO, LA VALENCIA MUSICAL, CANTIDAD DE BEATS, ENTRE OTROS?
- P2. ¿EXISTEN GÉNEROS QUE PESE A SER DISTINTOS, SUS CANCIONES TIENEN ATRIBUTOS SIMILARES? ¿Y VICEVERSA?
- P3. ¿LOS GÉNEROS CON MÁS CANCIONES CON LETRAS EXPLÍCITAS COMPARTEN CARACTERÍSTICAS ENTRE ELLOS?

The header features a light purple background with a white grid pattern. Several stylized musical notes in black, green, and blue are scattered across the top. In the center, a white banner with a torn edge contains the text 'PROPUESTAS EXPERIMENTALES' in a bold, black, sans-serif font. A black wavy line is positioned below the banner.

PROPUESTAS EXPERIMENTALES

1.

LA PRIMERA PARTE SE ENFOCA EN EL PREPROCESAMIENTO DE LOS DATOS, PARA ESTE LO PRIMERO QUE HICIMOS FUE ELIMINAR LAS COLUMNAS QUE NO ENTREGABAN INFORMACIÓN DE LOS ATRIBUTOS MUSICALES, COMO LO PUEDE SER LA POPULARIDAD, EL NOMBRE DEL ALBUM, ETC. PARA LUEGO HACER UN REDUCCIÓN DE LOS GÉNEROS CON LOS QUE SE VA A TRABAJAR, ESTO NOS AYUDA A TENER UN MEJOR CLASIFICADOR.

LUEGO SE HACE LA PARTICIÓN DEL DATASET EN TRAIN Y TEST, CON 0.8 PARA TRAIN Y 0.2 PARA TEST, POSTERIOR A ESTO VIENE UN RESAMPLING DE LOS DATOS.

FINALMENTE SE IMPLEMENTARON 3 MODELOS CLASIFICADORES: BAYES, SVM Y GRADIENT BOOSTING MACHINE, ADEMÁS DE UTILIZAR GRIDSEARCHCV PARA ENCONTRAR LOS MEJORES HIPERPARÁMETROS DE CADA MODELO.



PROPUESTAS EXPERIMENTALES



2.

PARA EL PREPROCESAMIENTO SE TOMAN 10 GÉNEROS A MANO PARA REALIZAR EL ESTUDIO, PARA ESTO SE BUSCARON GENEROS QUE SE PAREZCAN Y GENEROS QUE NO TENGAN NADA EN COMÚN

PARA REALIZAR EL CLUSTERING SE UTILIZARA K-MEANS Y DBSCAN, LA EVALUACIÓN DE ESTAS TÉCNICAS SE CENTRARON EN MEDIR LA COHESIÓN Y SEPARACIÓN, PARA ESTO SE UTILIZA EL COEFICIENTE DE SILHOUTTE, TAMBIÉN SE APLICARÁN MATRICES DE SIMILITUD

DEBIDO A LA ALTA DIMENSIONALIDAD DE LOS DATOS SE UTILIZAN TÉCNICAS DE PCA PARA OBTENER UNA REDUCCIÓN QUE FACILITE LA VISUALIZACIÓN DE LOS DATOS

A decorative header featuring stylized musical notes in green, yellow, and blue, some with asterisks. A central white banner with a torn edge contains the text 'PROPUESTAS EXPERIMENTALES'.

PROPUESTAS EXPERIMENTALES

3.

PARA ESTE PREPROCESAMIENTO SE SELECCIONARON LOS GÉNEROS MUSICALES QUE TIENEN MÁS CANCIONES CON CONTENIDO EXPLICITO

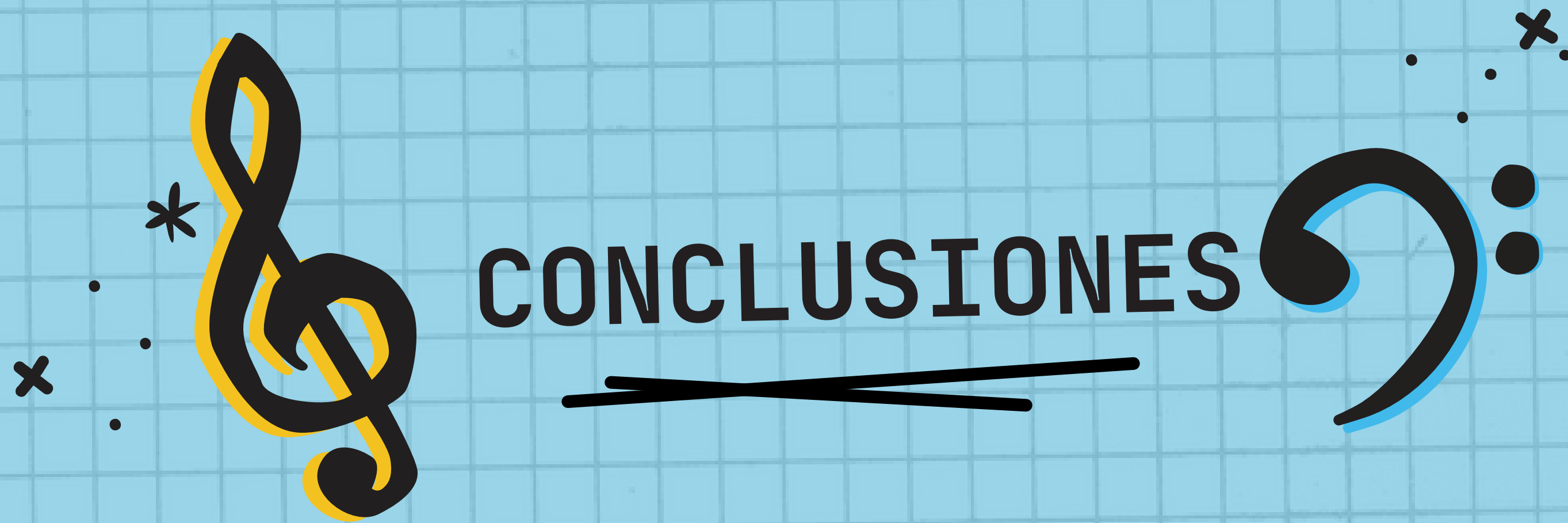
AL IGUAL QUE EN EL DESARROLLO DE LA PROPUESTA ANTERIOR SE UTILIZARA K-MEANS Y DBSCAN, JUNTO CON LAS MÉTRICAS DE EVALUACIÓN MENCIONADAS ANTERIORMENTE.

RESULTADO DE EXPERIMENTO 1

MATRIZ DE CONFUSIÓN CON LOS
RESULTADOS DEL CLASIFICADOR
SVM:

EL CLASIFICADOR QUE DIO MEJORES
RESULTADOS FUE SVM QUE ENTREGÓ
UNA PRECISIÓN PROMEDIO DE 0.62





CONCLUSIONES