



Spike Challenge - Detector de Reggaetones (usando Spotify)

¡Gracias por participar en el proceso de selección de Spike! Como parte del proceso, este desafío nos ayudará a entender la manera en que te enfrentas a problemas nuevos y, además, podremos evaluar tus conocimientos actuales.

Algunos puntos importantes,

1. Este desafío no te debiera tomar más de 2-3 horas de tu tiempo. Por lo mismo, no esperamos respuestas muy pulidas ni perfectas.
2. Las preguntas irán aumentando en dificultad, por lo que responde hasta donde puedas. Si por algún motivo hay alguna parte que no lograste completar, no hay problema.
3. Tendrás 5 días desde la fecha de recepción para enviar tus respuestas al desafío.
4. No hay restricciones de herramientas, lenguajes ni algoritmos para resolver el desafío.
5. Lee bien las instrucciones!

Instrucciones

Como dice el título, vamos a hacer un detector de reggaetones. Para eso armamos un dataset usando la API de Spotify, que nos da información detallada de cada canción ([documentación](#)). Para cada canción se tienen los siguientes parámetros: *acousticness*, *danceability*, *duration_ms*, *energy*, *instrumentalness*, *key*, *liveness*, *loudness*, *mode*, *speechiness*, *tempo* y *time_signature*.

Encontrarás tres archivos,

- *data_todotipo.csv* tiene canciones NO reggaeton.
- *data_reggaeton.csv* tiene SOLO canciones reggaeton.
- *data_test.csv* será el dataset para testear la solución.

Desafío

1. Analiza los dataset *data_todotipo.csv* y *data_reggaeton.csv*. ¿Qué puedes decir de los datos, distribuciones, missing, etc? ¿En qué se diferencian? Entregable: texto/imágenes.
2. Consolida los dos datasets en uno solo y genera una marca para las canciones que sean reggaeton en base a los datasets. Entregable: csv con dataset y código (si es que usaste).
3. Entrena uno o varios modelos (usando el/los algoritmo(s) que prefieras) para detectar qué canciones son reggaeton. Entregable: modelo en cualquier formato y código (si es que usaste).
4. Evalúa tu modelo. ¿Qué performance tiene? ¿Qué métricas usas para evaluar esa performance? ¿Por qué elegiste ese algoritmo en particular? ¿Cómo podrías mejorar la performance? Entregable: texto/imágenes.
5. Aplica tu modelo sobre el dataset "*data_test.csv*", agregándole a cada registro dos nuevos campos: *marca_reggaeton* (un 1 cuando es reggaetón, un 0 cuando no lo es) y *probabilidad_reggaeton* (probabilidad de que la canción sea reggaeton). ¿Cómo elegiste cuáles marcar? ¿De qué depende que predigas la existencia de mayor o menor cantidad de reggaetones? Entregable: texto/imágenes.

Envía un .zip conteniendo todas las respuestas, con el formato "[Apellido-Nombre].zip" a jobs@spikelab.xyz usando el asunto: "Spike Challenge".

Hemos creado además un slack para que todos puedan colaborar entre sí. Ingresa al slack en https://join.slack.com/t/spike-challenge/shared_invite/enQtMzE4NzkzOTkyNzM3LTEyMGM5NmExMDAzNDc3MDRhOTI4NzE3YzQ5MDJjOWE4NTNkMjA1OGZlMmMxMjZhN2EwMDc3YWw1ZTJjNDc0Nzg.