



MELODY MINERS
PRESENTA
EL PROYECTO FINAL

EQUIPO



PILAR PEREZ



ELISA GUTIÉRREZ



ALEXANDER GRAJALES



ALEJANDRO CATALAN



PABLO DE ALVA



ALANNA TORRES

INTRODUCCIÓN

Hemos desarrollado un sistema de Machine Learning que clasifica automáticamente el estado de ánimo de las canciones en función de sus características sonoras. Utilizando datos de Spotify, aplicamos técnicas de clustering y modelos supervisados para habilitar recomendaciones musicales basadas en emociones a través de un chatbot, con aplicaciones prácticas en los campos de la tecnología y el entretenimiento.

Este proyecto combina el impacto emocional de la música con la inteligencia artificial, permitiendo recomendar canciones de acuerdo con el estado de ánimo del usuario. Basado en modelos psicológicos validados, como el de valencia y arousal, ofrece soluciones innovadoras en plataformas de streaming, salud mental y tecnología personalizada.

PLAN GENERAL

PROCESAMIENTO DE DATOS

Limpieza y normalización.

GENERACIÓN DE ETIQUETAS - KMEANS

Agrupar canciones en clusters (moods) sin etiquetas iniciales.

CLASIFICACIÓN SUPERVISADA

Entrenar modelos de Machine Learning.

EVALUACIÓN Y SELECCIÓN DE MODELO

Comparar modelos con métricas.

RECOMENDACIÓN DE CANCIONES CON CHATBOT

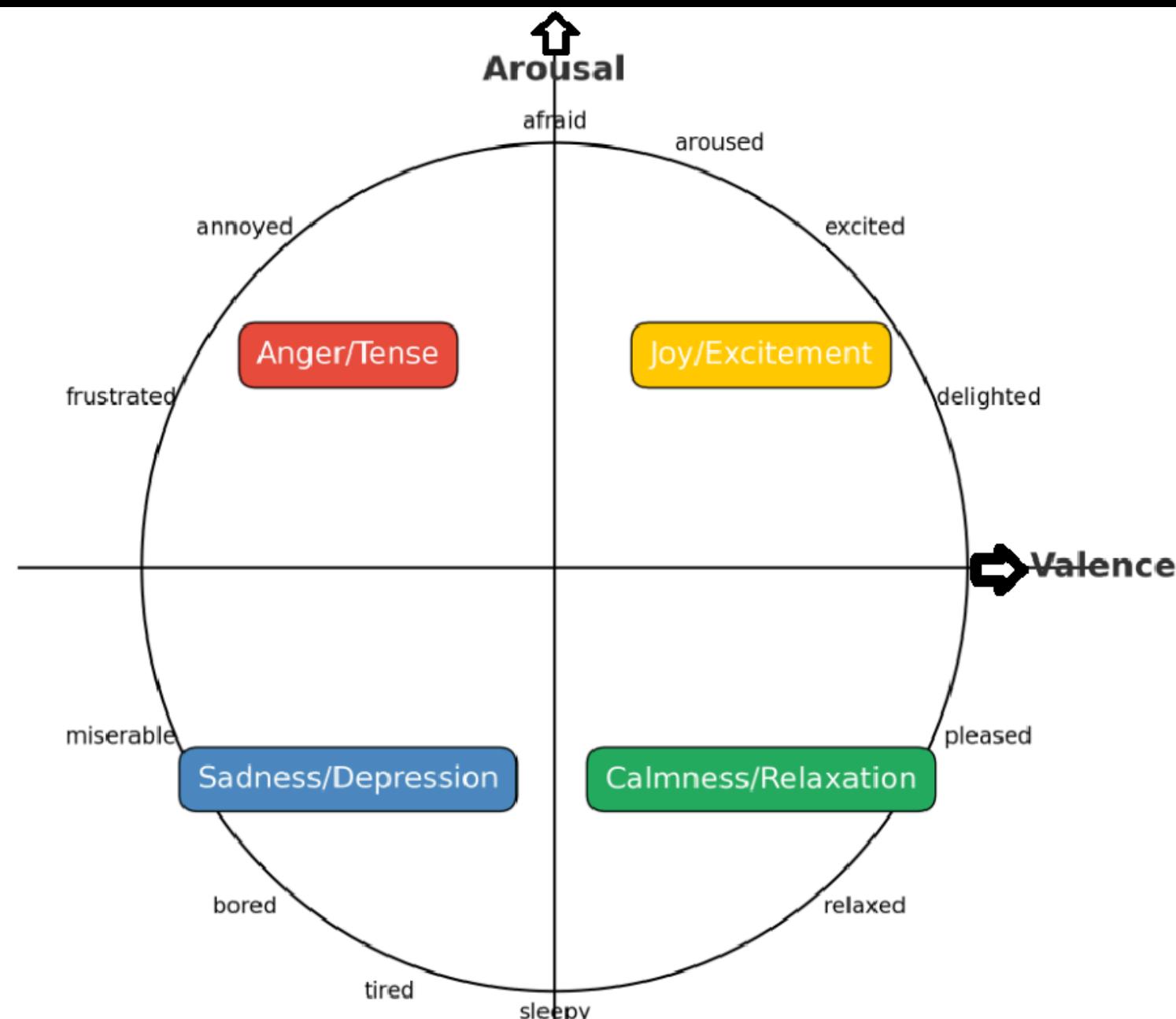
Clusterizar y predecir un nuevo dataset con el modelo elegido para recomendar canciones según el mood del usuario con un chatbot.

PROCESAMIENTO DE DATOS

DATASET

Feature	Útil para el mood	Razón concreta
<code>id</code>	X	Identificador único, no impacta el mood.
<code>name</code>	X	Nombre no influye directamente.
<code>popularity</code>	X	Popularidad no directamente relacionada.
<code>duration_ms</code>	X	Duración no influye directamente.
<code>explicit</code>	X	Contenido explícito sin relación directa.
<code>artists</code>	X	Nombre artista, percepción subjetiva.
<code>id_artists</code>	X	ID artistas no influye directamente.
<code>release_date</code>	X	Fecha lanzamiento sin impacto directo.
<code>danceability</code>	✓	Directamente relacionado con arousal.
<code>energy</code>	✓	Impacta directamente el arousal.
<code>key</code>	X	Tono musical sin relación directa.
<code>loudness</code>	✓	Volumen influye en intensidad emocional.
<code>mode</code>	X	Modalidad efecto emocional sutil.
<code>speechiness</code>	X	Contenido hablado sin impacto directo.
<code>acousticness</code>	X	Cantidad acústica sin influencia clara.
<code>instrumentalness</code>	X	Instrumentalidad sin efecto directo.
<code>liveness</code>	X	Factor 'en vivo' sin relación directa.
<code>valence</code>	✓	Valencia afecta directamente mood.
<code>tempo</code>	✓	Ritmo influye en excitación/arousal.
<code>time_signature</code>	X	Pulsos por compás sin impacto claro.

DATASET



El Modelo Circumplejo de Emoción de James Russell, organiza las emociones humanas en un espacio bidimensional con dos ejes principales:

- Valencia (de negativa a positiva, indicando si la emoción es agradable o desagradable).
- Arousal (de baja a alta activación, reflejando el nivel de energía o intensidad, como relajación frente a excitación o enojo).

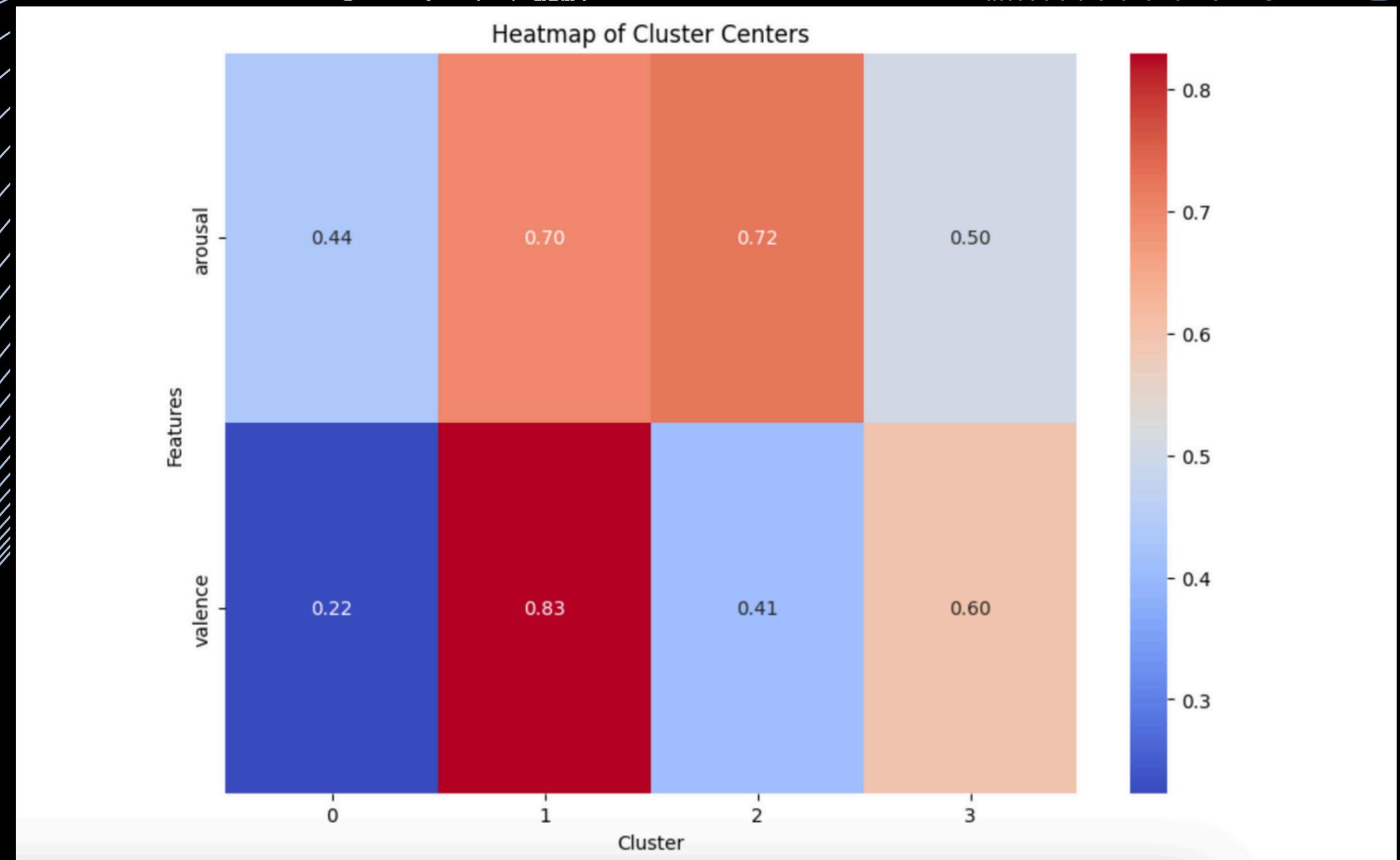
Este modelo permite visualizar y comprender las relaciones entre las emociones en un gráfico.

GENERACIÓN DE ETIQUETAS

K - MEANS

¿Por qué se eligió K=4?

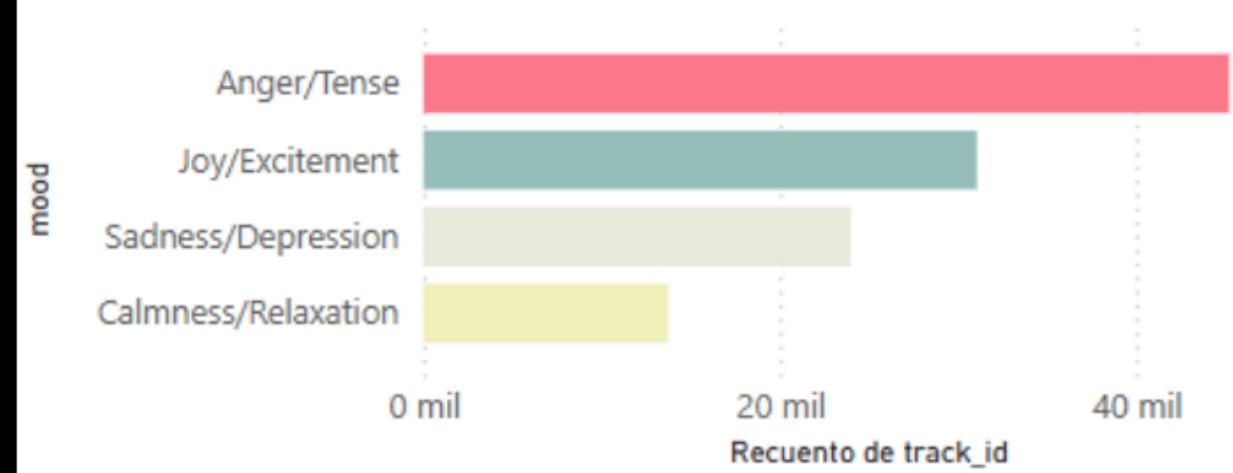
Ofrece la mejor combinación de separación, clara relación con "mood" y simplicidad interpretativa para clasificar canciones de Spotify, según sus características emocionales. Utilizando el modelo de Circumplejo de Emoción





VISUALIZACIÓN DE DATOS

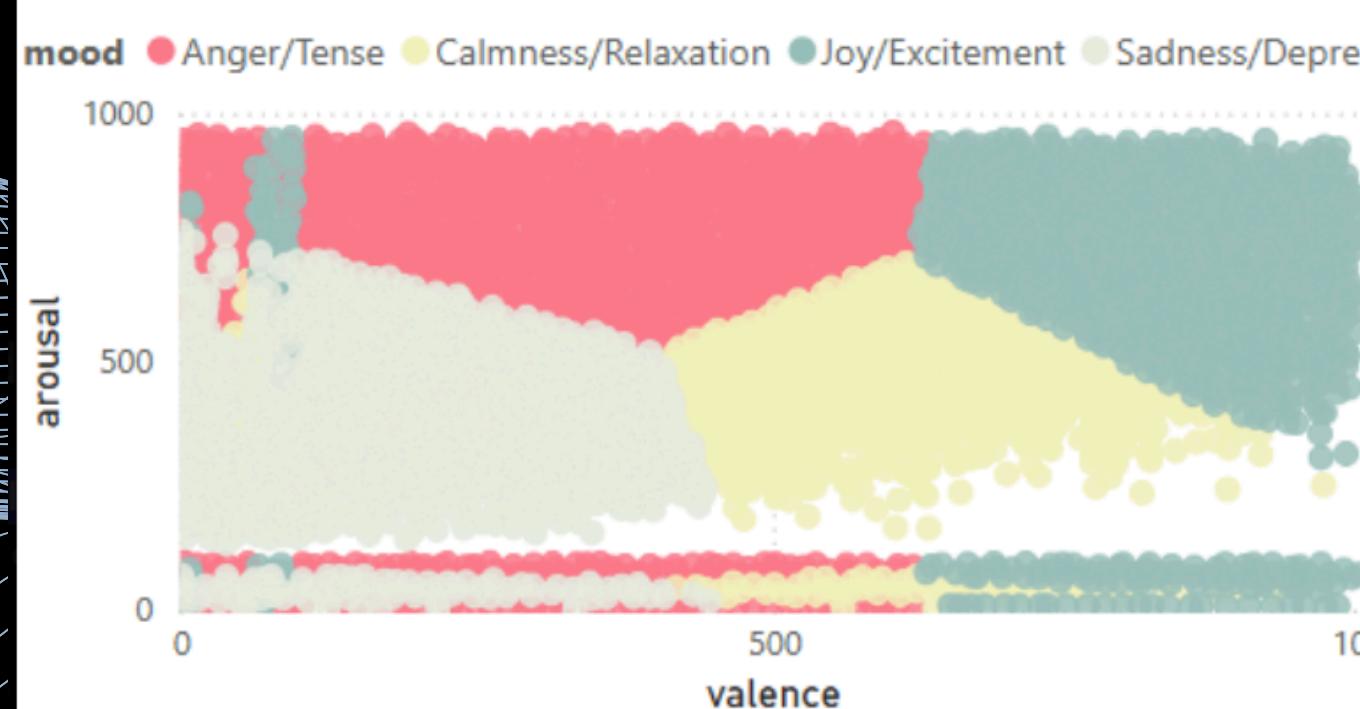
Recuento de Canciones por Mood



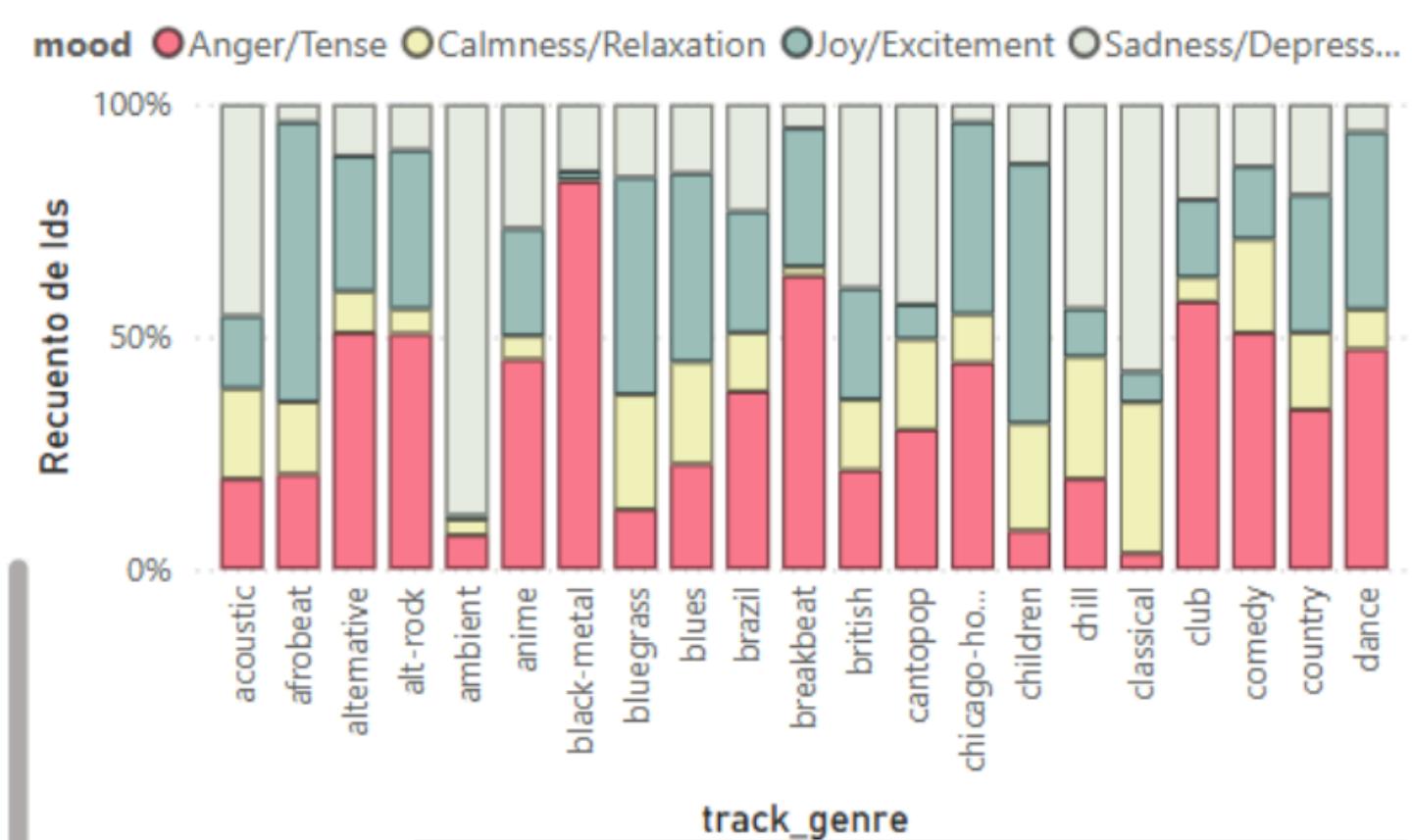
Valor Medio de las Features por Mood

mood	Promedio de energy	Promedio de tempo	Promedio de loudness	Promedio de danceability	Promedio de valence
Anger/Tense	714,01	546,43	817,04	499,48	333,68
Calmness/Relaxation	392,07	445,17	753,17	565,51	527,06
Joy/Excitement	685,93	528,81	807,90	601,50	728,64
Sadness/Depression	309,29	446,90	702,93	415,93	171,15
Total	582,48	508,50	782,86	517,66	430,42

Distribución de Moods según Arousal y Valence



Recuento de Canciones por Género y Mood



Cálculo Campo Arousal

Track Name	Energy (50%)	Tempo (25%)	Loudness (20%)	Danceability (5%)	Arousal
A Niente	0	303	543	183	194
Asmr: Sleepy Box Fan	0	0	773	0	155
Baby White Noise Series - Pure White Noise	0	0	437	0	87
Beach Rain	0	307	618	61	203
Beach Waves and Soothing Brown Noise	0	0	704	0	141
Box Fan - Loopable With No Fade	0	0	283	0	57
Box Fan Long Loop For Sleep	0	0	574	0	115
Box Fan Noise	0	0	507	0	101
Box Fan on High	0	484	714	149	271
Broad Spectrum White Noise: Constant Wind	0	275	687	17	215

Lista Canciones, Artistas y Mood

Name	Artists	mood
pagadoff	!Invite	Anger/Tense
strolling	!Invite	Sadness/Depression
Going on a Mission	"Puppy Dog Pals" Cast	Joy/Excitement
Puppy Dog Pals Main Title Theme	"Puppy Dog Pals" Cast	Joy/Excitement
King of Suede	"Weird Al" Yankovic	Anger/Tense
The Night Santa Went Crazy	"Weird Al" Yankovic	Anger/Tense
Amish Paradise (Parody of "Gangsta's Paradise" by Coolio)	"Weird Al" Yankovic	Calmness/Relaxation
Grapefruit Diet (Parody of "Zoot Suit Riot" by Cherry Poppin' Daddies)	"Weird Al" Yankovic	Calmness/Relaxation
Spatula City	"Weird Al" Yankovic	Calmness/Relaxation
Total		

CLASIFICACIÓN SUPERVISADA

MODELOS

EVALUACIÓN DE LOS MODELOS

Model	Best Grid Accuracy	Test Accuracy
Random Forest	0.99903	0.99932
KNN	0.99912	0.99944
Logistic Regression	0.99876	0.99902
SVC	0.99944	0.99951
Hist Gradient Boost	0.99752	0.99756
XGBoost	0.99747	0.99761

1 RANDOM FOREST (Combina múltiples árboles de decisión para mejorar la precisión y reducir el sobreajuste)

2 KNN (Clasifica datos basándose en la proximidad a sus vecinos más cercanos en el espacio)

3 LOGISTIC REGRESSION (Un modelo lineal que predice probabilidades para clasificación binaria)

4 SVC (Encuentra el hiperplano óptimo que separa las clases con el mayor margen posible)

5 HITS GRADIENT BOOST (técnica eficiente basada en histogramas que mejora iterativamente modelos débiles)

6 XGBOOST (versión rápida y optimizada del gradient boosting que maneja bien datos grandes y complejos)

CONCLUSIONES

K=4 ofrece una clasificación clara y útil para sistemas de recomendación.

SVC es el modelo más efectivo para clasificar 'mood' de canciones.

Creación de Chatbot para recomendación de canciones según el mood del usuario

¿Quieres verlo?

CHATBOT

Sad Mood - Joy Music



clideo.com

INGESTA DE DATOS (SIMULACIÓN)

The image shows a Windows desktop environment with three separate PowerShell windows open, each displaying command-line output. The windows are arranged vertically.

- Top Window:** Displays the command `python .\simu_send_data.py` being run in the directory `C:\Users\duverg\Documents\workspace-bootcamp_github\ProyectoFinal_BDML14\Code`. The output shows the script is executing successfully.
- Middle Window:** Displays the command `python .\simu_endpoint_receiver.py` being run in the same directory. The output indicates that a Flask app named 'simu_endpoint_receiver' is running on port 8888 in debug mode. It also includes a warning from Werkzeug about using a development server in production and instructions to press `CTRL+C` to quit.
- Bottom Window:** Displays the command `python .\simu_spark_processing.py` being run in the same directory. The output is currently blank, suggesting the process has just started or is still running.

¡GRACIAS!