

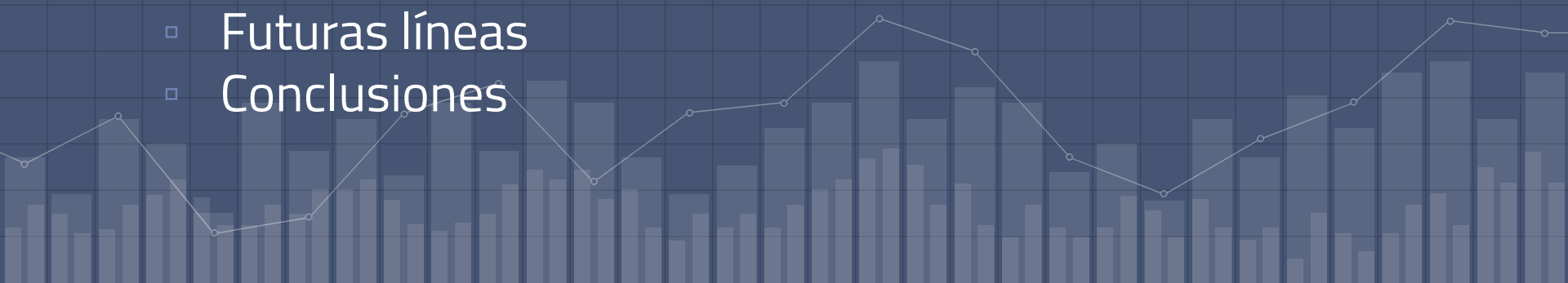
Presentación Ejecutiva

Por Christian Bastias

3 de sept. de 2022

CODER HOUSE

Contenidos

- ▣ Descripción de la problemática
 - ▣ Objetivo de estudio
 - ▣ Descripción de los datos
 - ▣ Análisis exploratorio y hallazgos
 - ▣ Algoritmos escogidos
 - ▣ Métricas
 - ▣ Futuras líneas
 - ▣ Conclusiones
- 
- A decorative background graphic at the bottom of the slide. It features a white line chart with circular markers and a bar chart with vertical bars of varying heights, all in a light blue color against the dark blue grid background.

Descripción de la problemática

Este proyecto aborda, esencialmente, la estimación e identificación de aquellos factores que permiten determinar la popularidad de una canción.

A lo largo de la historia moderna de la música, se han revelado al mundo artistas cuyas canciones siguen siendo escuchadas hasta la actualidad. Provengan de músicos conocidos por una sola canción (como *Gotye*) o por una larga y exitosa trayectoria (*Eminem* y *Rihanna*, por ej.), estos deben conservar su relevancia en el medio para asegurar su puesto como el mejor de su género.

Para esto, tanto artistas como productoras pueden apoyarse en métricas para obtener una ventaja ante este mercado altamente competitivo.



Objetivo de estudio

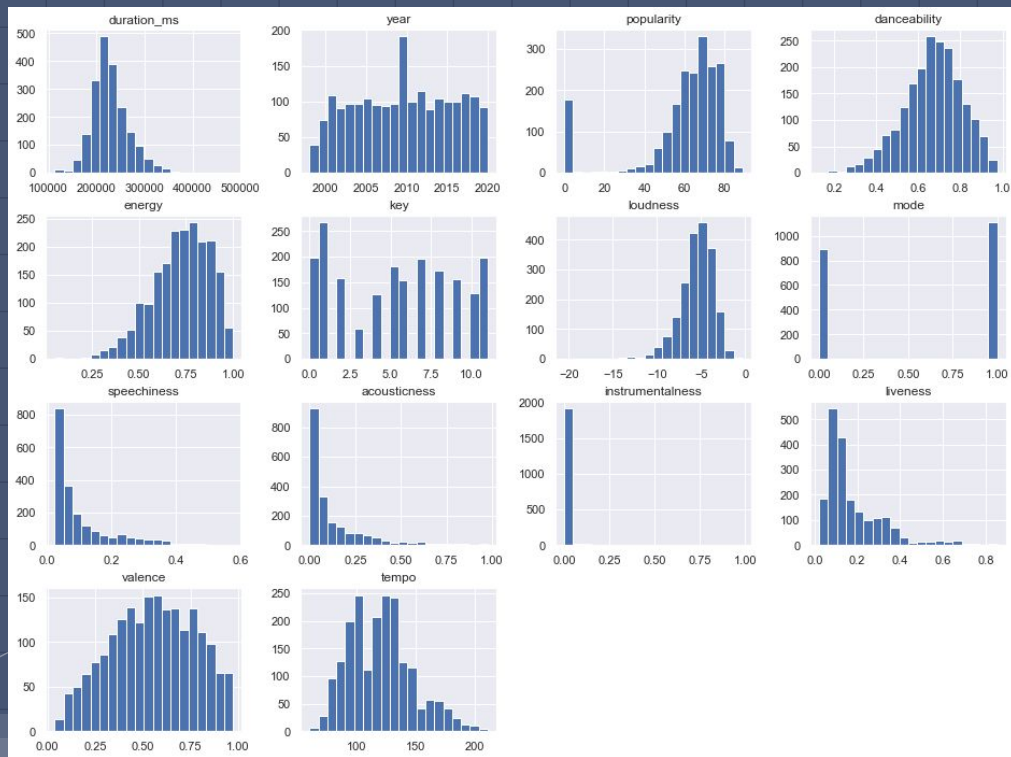
Estudio de las dimensiones de una canción, análisis de sus factores y elaboración de un modelo predictivo que permita estimar la popularidad de una canción.

Descripción de los datos

Se utilizan 18 dimensiones sobre la cual se puede describir una canción, siendo todos estos relevantes con el propósito de extraer información relevante de las mismas.

- Artista
- Canción
- Duración
- Explicitud
- Año de publicación
- Popularidad
- Bailabilidad
- Energía
- Tono
- Volumen
- Escala o modo
- Discurso
- Acustica
- Instrumentalidad
- Viveza
- Valencia o positividad
- Tempo
- Género

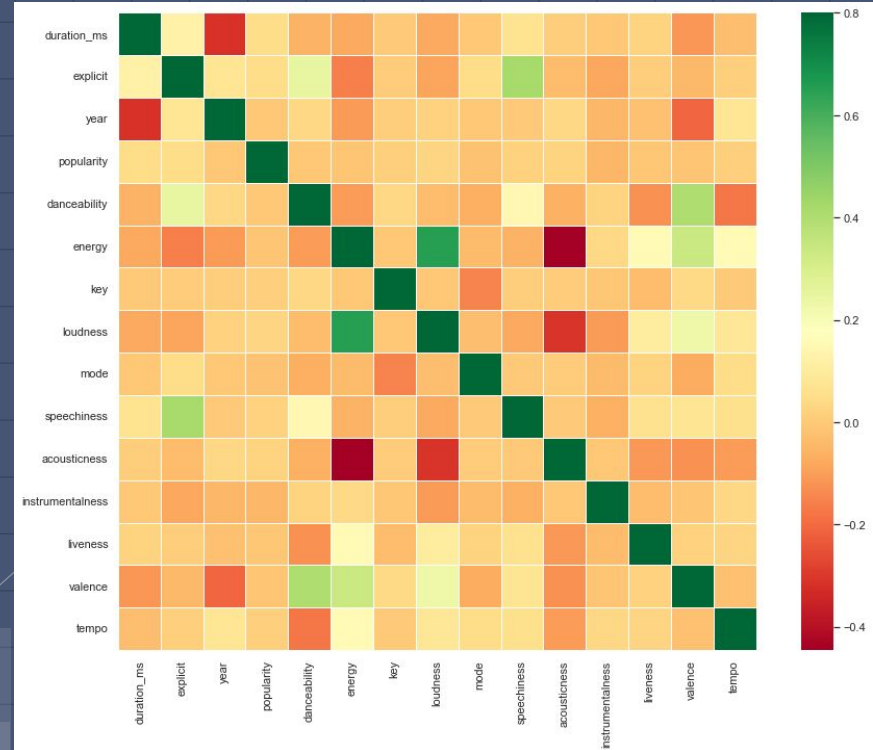
Análisis Exploratorio



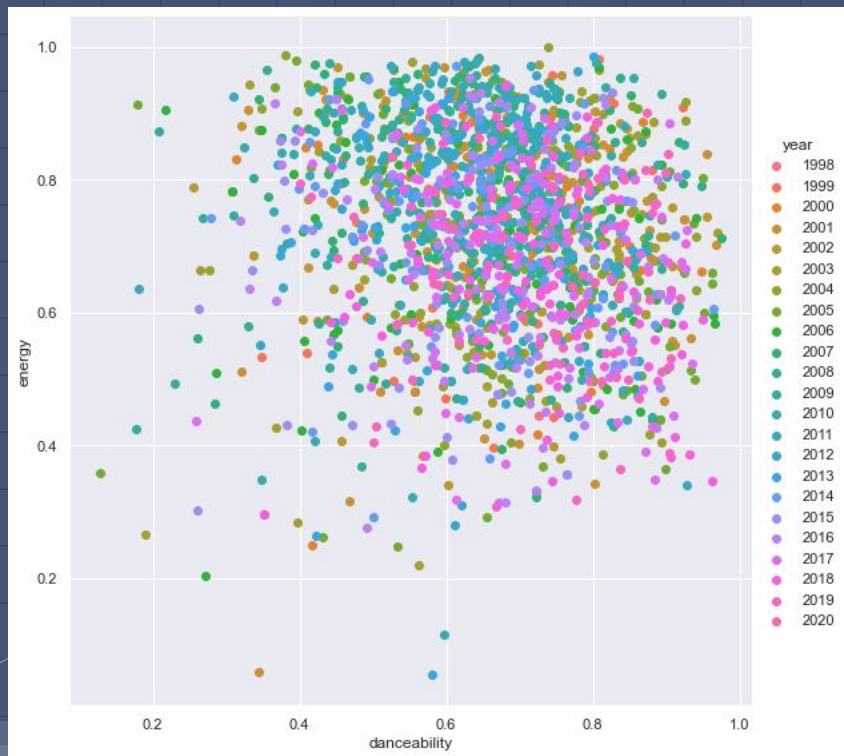
Se observa una tendencia en canciones en canciones de alta bailabilidad, volumen y energía como factores altos en canciones populares. Por otro lado, la gran mayoría de las canciones presentan baja acústica y discurso.

Por otro lado, se observa una baja correlación entre los datos en función de su popularidad.

En particular, destaca una correlación entre energía y volumen, y disponibilidad y volumen, siendo consistente con la información anterior.



Análisis Exploratorio



En función de su año de publicación, los factores de energía y bailabilidad son congruentes con su popularidad, es decir, estas dimensiones son relevantes a lo largo de los últimos 20 años.

Algoritmos utilizados

*Logistic
Regression*

*Random
Forest*

*Decision
Tree*

Algoritmos utilizados

(métricas)

*Logistic
Regression*

	precision	recall	f1-score	support
0	0.70	0.89	0.78	266
1	0.53	0.25	0.34	134
accuracy			0.68	400
macro avg	0.62	0.57	0.56	400
weighted avg	0.64	0.68	0.63	400

Cross val. scores

Media: 0.7038

Std: 0.0424

*Random
Forest*

	precision	recall	f1-score	support
0	0.71	0.86	0.78	266
1	0.53	0.31	0.39	134
accuracy			0.68	400
macro avg	0.62	0.59	0.59	400
weighted avg	0.65	0.68	0.65	400

Cross val. scores

0.6963

0.0377

*Decision
Tree*

	precision	recall	f1-score	support
0	0.71	0.71	0.71	266
1	0.43	0.43	0.43	134
accuracy			0.62	400
macro avg	0.57	0.57	0.57	400
weighted avg	0.62	0.62	0.62	400

Cross val. scores

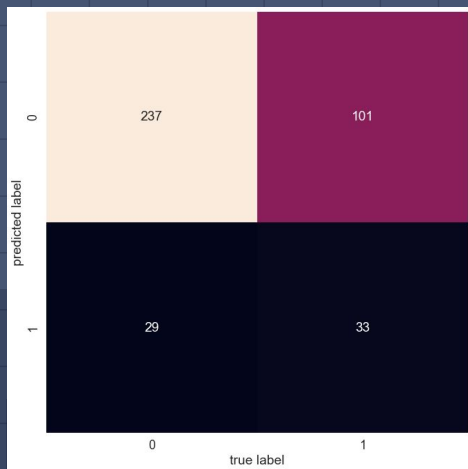
0.6951

0.0395

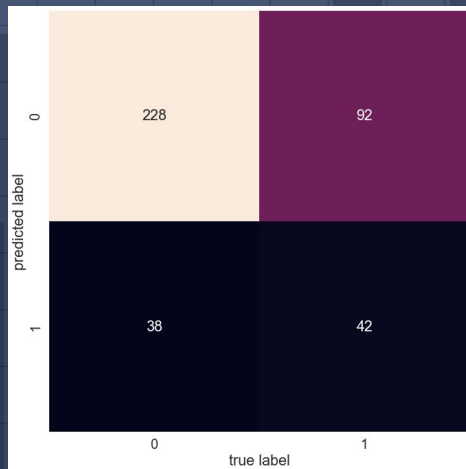
Algoritmos utilizados

(matrices de confusión)

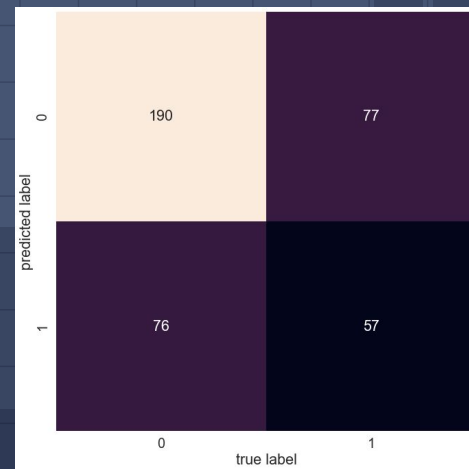
*Logistic
Regression*



*Random
Forest*



*Decision
Tree*



Futuras líneas

En base a los resultados obtenidos, se evalúa la posibilidad de realizar las siguientes líneas para mejorar el modelo.

- Identificar valores de configuración óptimos de los algoritmos para mejorar la precisión de los resultados **sin sacrificar tiempo** de ejecución.
- Aplicar otras técnicas de validación cruzada (*Hyperparameter tuning*) para **controlar mejor el comportamiento** del modelo y mejorar el desempeño.
- Implementar técnicas de procesamiento de lenguaje natural y regularización para **aumentar la fiabilidad** de los datos.

Conclusiones

En base a los resultados obtenidos, se concluye que los resultados emitidos por los algoritmos proveen estimaciones similares en cuanto a la predicción de que una canción sea popular. Sin embargo, se observa que el algoritmo *Random Forest* es el que mejor se ajusta a los datos, con una precisión de 0.71 y 0.53 para una canción no popular y popular, respectivamente.

De igual modo, su f1-score de 0.69 ofrece una mejor estimación de la precisión del modelo en función de su *dataset*, en comparación con los resultados obtenidos por *Logistic regression*.