

## Proyecto

# Aplicativo: Análisis de Información

Nombres: Daniela Lucía Ramos Betin - Diego Nicolás Avila Moreno.

**Tema:** Análisis descriptivo del contexto musical de la banda Gojira usando NLP, data mining en R.

## Índice

1. Introducción
2. Objetivo de Investigación
3. Descripción de las Fuentes de Información
4. Propuesta de la metodología.
5. Resultados Esperados.

## Introducción

El procesamiento de datos y las herramientas de software que permiten tratar datos han abierto una gran posibilidad de oportunidades a diferentes sectores de la sociedad en las que la estadística y la ciencia de datos son foco central y punto de partida. Ya sea en medicina, biología, finanzas y demás áreas del conocimiento, en todas se ha visto la posibilidad de aplicar dicha teoría así como una gran variedad de presentaciones de resultados por medio de aplicativos cada vez más utilizados y con mayor acceso en aprendizaje y uso.

Es por esto que en áreas como la música es posible aplicar la ciencia de datos, aún más cuando existe la posibilidad de encontrar tendencias que permitan a los artistas de próximas generaciones la manera adecuada de escribir letras que llamen más la atención del público y que les permita ser más reconocidos según el género en el que se desarrollan.

En nuestro caso, como miembros de este proyecto, estamos interesados en explorar los **contextos líricos de géneros musicales como el Death Metal.**

*Se elegirá en este caso la banda Gojira.*



Se trata de una banda de Trash Metal francesa, la cual desde 1996 ha producido varios éxitos en el género que, para los fanáticos del metal como lo somos los miembros del equipo, son un símbolo en la escena del rock.



Algo particular de ellos son los mensajes de sus canciones, pues estos están centrados en temas como el medio ambiente, la ecología y lo trascendental.

**Las letras son el recurso que queremos explorar en este proyecto y confirmar mediante las herramientas de análisis en R lo mencionado.**

## Objetivo de la Investigación

**Objetivo:** Desarrollar un aplicativo (*Shiny*) que permita hacer un análisis de las letras de Gojira por medio de minería de texto y análisis descriptivo que permitan entender el éxito detrás de la música de la banda.

## Fuentes de información

Hasta la fecha de entrega de este documento seguimos en la tarea de encontrar una fuente que contenga dicha información, generando un reto adicional en caso que deba obtenerse de manera propia. En medio de dicha investigación también han surgido diversas ideas que permitan recopilar los datos por medio de varias fuentes:

1. **Uso de Api's:** existen diversas API's que nos permitirían encontrar la información de las letras y demás datos antes mencionados. Algunas de ellas son Shazam (Core), Spotify (Core) y Genius.com las cuales nos permitirían sacar la información de manera práctica. Estas se encuentran dispuestas en el sitio ***rapidapi.com***.



2. **Scraping / Mining:** Minería de páginas web, enfocado más que todo en las letras de las canciones. Existen diversas páginas que contienen la información de las letras de las canciones, por lo que hacer scraping resultaría un método alternativo para sacar la data, en R se implementa con la librería **rvest**.



3. **Investigación formal:** la información sobre álbumes, ganancias y demás datos diferentes a las letras pueden ser recopilados por medio de investigación sin necesidad de usar herramientas de programación. *Gojira* tiene 7 álbumes en total (82 canciones), por lo la búsqueda de datos la discografía no sería tan demorada de buscar a comparación de otros artistas con un mayor número de trabajos en el mismo tiempo de existencia de la banda.

Canción	Álbum	Letra	Año	No. Reproducciones	Ranking Top	Unidades Vendidas
String	String	Doc	Int (1996-2022)	Int (0-Inf)	String (Top 10, Top 100, Unranked)	Int (0-Inf)

Total de registros: 82, representado en el total de canciones de la banda.

## Propuesta de la metodología

La aplicación iniciaría con una breve introducción sobre Gojira, de tal manera que sea posible para el usuario entender quienes son y en qué se caracterizan. Después de esto vendría el despliegue del aplicativo.

**Análisis Cualitativo:** Seguido, aplicaríamos minería de texto con la cual tomaremos las letras de las canciones y aplicaremos frecuencias para hacer diversos análisis. El objetivo es identificar estas frecuencias y cómo actúan a lo largo del tiempo de existencia de la banda. Saber si existen palabras que puedan estar relacionada con *rankings* o incluso saber si existen términos que puedan correlacionarse con las ventas de unidades.

Figure 1 consists of three bar charts, each showing the distribution of GDR (Global Developmental Risk) across seven pathways (Pathway1 to Pathway7). The y-axis represents the count, and the x-axis represents the pathway. The legend indicates the GDR categories: Definitive (red), Limited (green), Moderate (blue), Heavily (yellow), and Strong (purple).

- Left Chart (GDR):** Shows the distribution of GDR across seven pathways. The y-axis ranges from 0 to 400. The x-axis labels are Pathway1, Pathway2, Pathway3, Pathway4, Pathway5, Pathway6, and Pathway7. The legend indicates the GDR categories: Definitive (red), Limited (green), Moderate (blue), Heavily (yellow), and Strong (purple).
- Middle Chart (GDR):** Shows the distribution of GDR across seven pathways. The y-axis ranges from 0 to 1000. The x-axis labels are Pathway1, Pathway2, Pathway3, Pathway4, Pathway5, Pathway6, and Pathway7. The legend indicates the GDR categories: Definitive (red), Limited (green), Moderate (blue), Heavily (yellow), and Strong (purple).
- Right Chart (GDR):** Shows the distribution of GDR across seven pathways. The y-axis ranges from 0 to 1000. The x-axis labels are Pathway1, Pathway2, Pathway3, Pathway4, Pathway5, Pathway6, and Pathway7. The legend indicates the GDR categories: Definitive (red), Limited (green), Moderate (blue), Heavily (yellow), and Strong (purple).





## Resultados Esperados

En cuanto a los resultados cualitativos, esperamos obtener la confirmación de que la letra de las canciones de la banda tiene un contexto vegano ambientalista.

En cuanto a la aplicación, esperamos poder desarrollarla en base a la estructura modularizada presentada durante el curso, permitiendo obtener ***insights*** alrededor de las letras de la banda donde sea posible identificar si las letras de sus canciones puedan estar relacionadas a su mensaje ambientalista, su éxito comercial, en ranqueo o incluso en número de unidades vendidas.

Igualmente se espera que con este aplicativo se refuercen los conocimientos que se vean en clase y que harán parte del diseño final del mismo.

