PROCESAMIENTO Y CLASIFICACIÓN DE DATOS: ROADMAP DEL PROYECTO

ARNOLDO OLIVA

PROYECTO

Tema

 Clasificación de Audio (sentimientos)

Problemática

 Dicho proceso podría servir para Predicción del estado emocional de un cliente

Datos

- Archivos .wav, .mp3, etc.
- Fuente: Kaggle

PROYECTO: PASOS

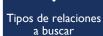
Extracción y lectura de datos: parseo de los files en formato .wav y .mp3 en librerías de python Feature Engineering: limpieza de datos, eliminación de tiempos vacíos, remoción de outliers, etc. Exploración de datos: espectrogramas, decibeles, etc.

Modelado y evaluación: splitting de los datos, modelo en sí, evaluación del mismo. Predicción de ciertos audios al azar.

PROYECTO: PREGUNTAS EXTENDIDAS

¿Qué está buscando? • Predicción de audios (clasificación de sentimiento)

¿Refleja el problema? • No en sí; pero se puede aterrizar para caso de uso de una empresa interesada en predecir el estado de ánimo de clientes recién atendidos.



• Posibles patrones entre características de las ondas de sonido, sus decibeles, etc.

PROYECTO: PREGUNTAS EXTENDIDAS

Resultado o atributo a predeci

• Una etiqueta, que muestre si la manera en la que la persona dijo una oración sea feliz o triste.

Realizar predicciones o buscar patrones Realizar predicciones

Tipo de datos que tiene Audios en formato .mp3 y/o .wav

PROYECTO: PREGUNTAS EXTENDIDAS

Hay varias tabla

• No hay varias tablas, cada archivo es independiente entre sí, pero están contenidos en folders diferentes.



• Se tiene que explorar los datos para determinar la limpieza, pero a grandes rasgos sería limpiar espacio vacío, eliminar outliers de decibeles, entre otros.



• Las muestras son independientes entre sí, son la misma frase pero dicha por diferentes personas en diferentes tonos; cada archivo no dura mas de 15 seg.



• No. No son datos de los procesos de ninguna empresa.