Procesamiento de Lenguaje Natural (NLP)

Especialización y Maestría en Analítica Estratégica de Datos

2020-II

Plan de Proyecto Final

1. Estudiante(s) — Nicaulis Chica Brand
2. Título provisional del Proyecto: De que hablan las canciones
3. ¿De qué se trata el proyecto? Una aplicación de NLP para analizar texto: partiendo de un conjunto de canciones.
4. ¿Cuál es la motivación? Aplicar aprendizaje automático en un ejemplo práctico y conocido como es la música
5. ¿Cómo se relaciona con NLP? El procesamiento del lenguaje natural va de la mano de la analítica de texto, la cual cuenta, agrupa y categoriza palabras para extraer estructura y significado. La analítica de texto se utiliza para explorar contenido textual y derivar nuevas variables de texto crudo que se pueden visualizar, filtrar o utilizar como entradas para modelos predictivos u otros métodos estadísticos.
6. ¿Dónde va a conseguir los datos? ¿Son etiquetados? En la Web
7. De las cosas que hemos visto (o vamos a ver) en clase, ¿qué piensa usar?

**Primero Extraer y preparar el texto.**

Extraer el texto Utilizando la técnica llamada web scraping dado que la información se extraer de la paginas de internet o blogs de música.

Luego pro-prosesar el texto limpiar el texto de ruido y estandarizarlo para que sea fácil de comparar

Estos pasos incluir quitar puntaciones, transformar el texto en minúscula, eliminar palabras frecuentemente usadas, preposiciones Artículos Etc.

**Segundo Convertir el texto en una representación numérica**

una vez limpiado el texto se procede a aplicar el algoritmo de aprendizaje automático para clasificar en texto

**Tercero Clasificar**

A partir de esta representación ya es posible clasificar con la técnica NO supervisada se llama supervisada dado que se conocen los temas

¿Cómo piensa presentar los datos y sus resultados?: un document donde se evidence el desarrollo del Proyecto y el codigo

1. Otros comentarios: