

Practica 1 (PREPROYECTO) de Metodología de DataScience

Serás asistido en todo momento por el Dr Miguel Félix y todo el equipo (Roberto, Jacobo, Memo, Jorge, Carlos, Sergio)

Puedes preguntar y pedir apoyo en todo el proceso y fases

NOTAS IMPORTANTES:

- Se realiza en Equipo de 2 a 4 personas (puede ser el mismo que el de su proyecto final)
- Se entrega el próximo Lunes 16 de Julio de 2018 en un documento que incluye las fases de la metodología descrita a continuación

DESCRIPCIÓN: Se desarrollara un experimento de DataScience para pronosticar si un meme político genera una reacción positiva, negativo, tristeza, enojo, risa, etc. en la red social Facebook

Objetivo General: Aplicar la metodología de DataScience para pronosticar el impacto de un meme con mensaje político en la red social Facebook (tipo de reacción:like, triste, share, etc)

Fases de la Metodología:

1. Recuperar 30 memes (por equipo) de la red social facebook relacionadas con las elecciones 2018 y generar un dataset (incluyendo descargar la imagen y el texto del meme o explicación)
2. Compartir los memes descargados y su estadística con los otros equipos del grupo, para que así se tengan de 100 a 300 memes en un dataset global)
3. Aplicar algoritmo de Machine learning (agrupación- clustering) para identificar a cual preferencia política pertenece una persona en la red social Facebook, a partir de la reacción del meme (puede incluso compartir el meme para obtener datos de la reacción) . **Este proceso implica descubrir la caracterización del meme** (si hay dudas pregunta)
4. Aplicar algoritmo de regresión para pronosticar el impacto de un meme específico dentro de la red social Facebook (una vez que se tienen los grupos identificados)
5. Evaluar el nivel de desempeño del algoritmo de regresión y clustering con las métricas de Machine Learning y validando con el resultado real obtenido en FB
6. Generar el reporte final explicando los resultados

Fase 1: En la red social Facebook, desde cada uno de los perfiles del alumno o de grupos o fan pages, descubrir memes de indole politica, guardar la imagen y también en un archivo excel incluir la estadística de la reacción del meme. (ESTE ES UN PROCESO SEMIAUTOMÁTICO) *Deberá generar un carpeta en Drive, con las imágenes y en otra el archivo excel*

Ejemplo:



04072018. Meme AMLO-PEÑA

Id de Imagen	Descripción o texto del meme	# de reacciones (likes, enojo, etc)	Lugar o ubicación de las personas que reaccionaron	Nombre del perfil de donde se identificó el meme (público o privado)	Clase o Categoría	Comentarios en el meme
04072018	EPN hace una pregunta grotesca AMLO, le responde reflejando que la pregunta es tonta	10 Likes 5 hahaha	-Aide Ruíz Monterrey J Carlos Ramirez N/A Kike Vazquez N/A Yolanda Rodriguez -Saltillo Fabián Avila Jose Juan Hernandez -Texas	Qué hay de Nuevo León (público)	-Risa -Like	N/A

La columna Descripción o texto del meme , hay que dividirla en otras tres columnas como sigue:

texto del meme	texto dentro de la imagen	Texto escrito asociado a la imagen
EPN: ¿y como es vivir en plaza sesamo? AMLO: ¡ese es ELMO yo soy AMLO!	¿y como es vivir en plaza sesamo? -ese es ELMO yo soy AMLO	El mejor meme del encuentro AMLO/PEÑA

La meta es que el algoritmo de Clustering ayude a:

- Identificar memes de política que generen humor-risa ,
- Identificar memes de política que generan odio,
- Identificar memes de política neutros (no generan reacción)

Esto a partir de las estadísticas de la tabla 1 y de los comentarios.

- Etiquetar los datos de los memes de acuerdo a las clases identificadas:

- a. odio
- b. humor
- c. neutro

Fase 2: Comparta los memes descargados y su estadística con los otros equipos del grupo e integrelos en un unico dataset

Fase 3.- Aplicar algoritmo de Machine learning (agrupación- clustering) para identificar a cuál preferencia política pertenece una persona en la red social Facebook

Fase 4: Aplicar algoritmo de regresión para el pronosticar usando Python

Fase 5. Evaluar el nivel de desempeño del algoritmo de regresión y clustering con las métricas de Machine Learning y validando con el resultado real obtenido en FB

Fase 6: Generar el reporte final explicando los resultados