

### Parcial, Primera Oportunidad. Primer Cuatrimestre 2017: Criterio

1- De la forma en que está planteado el ejercicio es conveniente primero quedarse unicamente con el registro de mayor cantidad de asistencias en partidos de playoffs porque si ese no supera al promedio entonces ninguno lo hace, eso evita tener que hacer un join con mayor cantidad de registros. Si no usan esta estrategia descuento de 3 puntos.

Extraños usos de groupby o cosas por el estilo descuento MINIMO de 5 puntos.

2- Para el punto a) Si no tiene en cuenta un mínimo de ratings por libro antes de considerar el promedio -3. Si agrupa mal o calcula mal el promedio -5, si la lógica está mal vale cero. Errores mínimos de sintaxis no serán tenidos en cuenta, pero si utilizan funciones no existentes se descuentan puntos en función de que hagan. Para el punto b) Al igual que antes, si calculan el promedio sin evaluar la cantidad de ratings para un book\_id dado -3, si hace mal el join/merge -5, si hacen mal la validación de la condición -3.

3- V/F. Poner Verdadero o Falso de por si no otorga puntaje, debe estar justificado. Si no está justificado vale cero.

- a) Falso, ya que BM25 agrega a su vez un factor de normalización de las longitudes de documentos, por lo que aun tomando  $k \rightarrow \infty$  el resultado de BM25 podria ser distinto al de TF-IDF. Si dicen que es Verdadero y aclaran que esto es para documentos con longitudes similares puede estar bien también.
- b) Falso, es posible modificar la búsqueda binaria para que funcione con la estructura de Front Coding Parcial. Si dicen que es Verdadero aclarando que se puede modificar la búsqueda binaria, pero aun asi hay que recorrer unos pocos registros secuencialmente está bien también.
- c) Falso,  $\log(N+1/ft_i)$  es un factor que otorga a cada termino un peso inversamente proporcional a su frecuencia de aparición en los documentos. Como parte de la justificación deben aclarar que esto no tiene que ver con la longitud de los documentos y para qué sirve realmente.
- d) Verdadero, Deben justificar idealmente con un ejemplo, o con una explicación clara de por que es asi.

4- Hay que calcular el signo del PI entre los 3 vectores y los 6 hiperplanos aleatorios, el resultado es una matriz de  $6 \times 3$  (o  $3 \times 6$ ) con +1 y -1 o con 0 y 1. Luego hay que comaprar el AND de 3 y el OR de los 2 resultados con el OR de a 2 y el AND de los 3 resultados. En uno queda una colisión y en el otro quedan 2 colisiones (tres pero una se repite). Si no hay una breve explicación sobre por qué ocurre esto descuento de 2 puntitos.

5- Se espera el desarrollo completo del PPMC. Si actualizan todo sobre la misma tabla y no se puede corregir la evolución de cada modelo se tendrá en cuenta exclusivamente lo que indique que emite y la actualización de los intervalos pero no se podrá considerar si arrastran algún error. Si no aplican exclusión (siempre que se pueda) -5, si no indica la salida en binario del archivo comprimido -7, si solo consideran A, B y EOF en el modelo -1 son -8puntos ya que se aclaro en clase que siendo dinamico el modelo -1 debia contener todos los caracteres posibles.

6- Para el primer punto simplemente hay que graficar tomando las 2 primeras columnas de U como coordenadas. Para el segundo punto hay que hacer el producto entre la primera columna de U y el primer vector de V resultando en una matriz de  $5 \times 5$ . Para el tercer punto se puede simplemente observar que tenemos la matriz original completa al tener la SVD y calculando  $U \cdot S \cdot V$  se observa que la matriz es positiva.

7- Hay 3 colisiones en el primer nivel. Por lo tanto en el segundo nivel hay que crear una tabla de 9 posiciones ( $n^2$ ). Si no la crean de 9 posiciones -5 puntos.

Para esta tabla hay que buscar una función que sea perfecta para las 3 claves que colisionaron y las 9 posiciones posibles.