



Universidad Simón Bolívar

Departamento de Cómputo Científico y Estadística

CO3321 Estadística para Ingenieros

Enero-Marzo 2016

## **Proyecto 1**

Alexis Ibarra, carné 06-39741

9 de mayo de 2016

# Índice

1. Antecedentes	3
2. Objetivos	4
3. Descripción del experimento	5
4. Procedimiento	6

# 1. Antecedentes

El analisis automatico de las redes sociales es uno de los topicos de mayor movimiento en la estadistica y en las areas de Data Science en la ultima decada. El volumen y velocidad con que se genera la informacion hace extremadamente relevante poder identificar y clasificar el comportamiento de las interacciones de los usuarios de redes sociales.

Uno de los topicos en estudio es la deteccion de Topicos, o elementos que pudieran ser de interes masivo en la red. Cada topico (por ejemplo un hashtag en twitter) involucra una serie de interacciones en diferentes discusiones (twits) entre diferentes autores que pueden o no involucrar varios otros topicos o discusiones. Cada topico, ademas, tiene un tiempo de vida durante el cual el topico muestra mucha actividad.

Los datos del presente proyecto forman parte de una recoleccion de datos usada para crear modelos que puedan predecir la aparicion de Buzz's o topicos de alta relevancia en la red social. La base de datos solo considera varios topicos que recibieron contribuciones por varios dias, y se tomaron varias medidas sobre el topico durante siete dias consecutivos. En general, un topico se considera un Buzz si presenta a lo menos 500 contribuciones por dia.

## 2. Objetivos

El objetivo del proyecto es explorar la informacion para entender el comportamiento de los topicos en las medidas tomadas. Especificamente deseamos entender:

- Cual es la probabilidad de ocurrencia de un Buzz
- En promedio, durante la semana, cuanto dura el interes en un topico y como varia este cuando se trata de un Buzz.
- Como se diferencian los comportamientos alrededor de los topicos comparando el dia en que inicia versus al final de la semana? Y a los cuatro dias?
- ¿Que se puede decir de la cantidad de contribuciones que tiene un topico?
- ¿Como cae la tasa de contibuciones en funcion del interes que se tiene en el topico? Y respecto al numero de contibuyentes? ¿Como se diferencia este comportamiento respecto a un Buzz?

### 3. Descripcion del experimento

Los datos presentados contienen indicadores de informacion de Twitter tomados en 2013. Cada fila de esta data representa siete dias de observacion de un topico especifico ( por ejemplo, Coldplay ) considerando un par de dias luego de su primera observacion. Se considera que un topico es un Buzz! si hay a lo menos 500 discusiones activas por dia (en promedio, con respecto a la observacion inicial) Todas las observaciones son independientes En total hay 140.707 filas. Cada fila contiene 11 atributos medidos a lo largo de siete dias:

- Columnas 1-7: NCD1 ... NCD7 numero de discusiones creadas respecto al topico por dia
- Columnas 8-14: AI1 .. AI7 Incremento de nuevos autores interactuando (Popularidad)
- Columnas 15-21: AL1 .. AL7 Nivel de atencion, es decir, cantidad de autores por dia medido en escala intervalo (0,1)
- Columnas 22-28: BL1 .. BL7 Nivel de impacto (burst) medido como el numero de discusiones creadas entre el la cantidad de discusiones (NCD / NAD)
- Columnas 29-35: AC1 .. AC7 Numero de Contenedores Atomicos en las redes sociales
- Columnas 36-42: AL1 .. AL7 Una medida de la atencion prestada en las redes sociales
- Columnas 43-49: CS1 .. CS7 Dispersion de la contribucion al topico
- Columnas 50-56: AT1 .. AT7 Interaccion de los autores, el promedio de autores contribuyendo en el topico
- Columnas 57-63: NA1 .. NA7 Numero total de autores contribuyendo en el topico por dia
- Columnas 64-70: DL1 .. DL7 Longitud de la discusion.
- Columnas 71-77: NAD1 .. NAD7 Cantidad de discusiones en el topico

## 4. Procedimiento