**Universidad de Santiago de Chile**
**Facultad de Ingeniería**
**Departamento de Ingeniería Informática**

**Asignatura** : Taller de minería de datos avanzada
**Programa** : Magister en Ingeniería Informática
**Profesor** **:** Dr. Max Chacón Pacheco
**Ayudante** **:** Felipe-Andrés Bello Robles

**Fecha Entrega Oral** **:** 2 de Mayo de 2018
**Fecha Entrega Escrito** **:** 9 de Mayo de 2018

# TALLER 3: Máxima Entropía

**Objetivos:**

- Comprender y presentar el problema de clasificación de texto (categorización)
- Realizar pre-procesamiento de texto para problemas de clasificación.
- Selección de muestra para entrenamiento y test
- Realizar proceso de calibración de parámetros del algoritmo de máxima entropía.
- Comprender de forma práctica el funcionamiento del método de máxima entropía mediante la configuración de sus parámetros
- Evaluar el rendimiento del algoritmo mediante sus índices "precisión", "recall" y "F1", definiendo un subconjunto de categorías "relevantes" para dicho propósito.

**Aspectos importantes a considerar:** Para obtener los resultados y cumplir los objetivos del laboratorio, se debe tener en cuenta los siguientes puntos:

- Utilizar "R" http://www.r-project.org/ y sus librerías "maxent", "tm" y "SnowballC".
- Realizar una comparación la literatura, de manera de establecer la efectividad del método a la resolución del problema, incluyendo ventajas y desventajas de éste.

Archivos de datos: generalmente archivo *.data, archivo *.names

**Escrito:** Se debe elaborar un *Artículo* de máximo 6 páginas, según el formato:
https://www.springer.com/gp/computer-science/lncs/conference-proceedings-guidelines

| | Puntos a evaluar | Porcentaje |
|---|---|---|
| **Estructura del Artículo** | Presentación, ortografía y redacción | 5% |
| | Abstract e Introducción | 10% |
| | Métodos (explicación del funcionamiento) y Datos (pre-procesamiento usado) | 20% |
| | Resultados | 20% |
| | Discusión | 25% |
| | Conclusiones | 20% |

**Observaciones:**

Consultas al mail Felipe.bello@gmail.com, Felipe.bello@usach.cl

El trabajo debe ser presentado de forma oral (50%) y escrita (informe 50%) en horario de clases el día 2 de Mayo y 9 de Mayo de 2018. Disponen de 15-20 minutos de exposición y 10 para contestar preguntas de la comisión.

La información de las bases de datos se encuentra en la página:

http://archive.ics.uci.edu/ml/

https://cran.r-project.org/web/packages/maxent/maxent.pdf

https://cran.r-project.org/web/packages/tm/tm.pdf

ftp://cran.r-project.org/pub/R/web/packages/SnowballC/SnowballC.pdf

Nota Final: Promedio simple de las experiencias.