

# Proyecto Final de Minería de Textos

Andrés Auquilla

28 de enero de 2021

# Índice

<b>1. Introducción</b>	<b>3</b>
<b>2. Alance del Proyecto</b>	<b>3</b>
2.1. Tema del Proyecto . . . . .	3
2.2. Conformación de los Grupos . . . . .	3
2.3. Datasets a utilizar . . . . .	3
<b>3. Partes del Proyecto</b>	<b>3</b>
3.1. Propuesta . . . . .	3
3.2. Reporte Final . . . . .	4
<b>4. Evaluación</b>	<b>4</b>

# 1. Introducción

El proyecto final de la materia Minería de Textos tiene por objetivo que el estudiante aplique las técnicas aprendidas a lo largo del semestre. Adicionalmente, que el estudiante pueda tomar decisiones sobre la metodología a realizar en base a un fuerte fundamento teórico. Además, que pueda generar conclusiones basadas en evidencia; para de esta manera, interpretar correctamente los resultados obtenidos.

## 2. Alcance del Proyecto

### 2.1. Tema del Proyecto

El tema del proyecto es abierto; es decir, no existe un tema definido para todos los grupos. Esto implica que el estudiante deberá proponer un tema y su respectivo alcance. Más detalles al respecto en las siguientes secciones.

Todo tema propuesto deberá utilizar alguna o algunas de las técnicas analizadas a lo largo del semestre, i.e. clasificación de textos, clustering de textos, modelamiento de textos, etc. A continuación, los elementos mínimos que se deberán cubrir en cada proyecto:

- a. Análisis descriptivo de los datos: gráficas, tendencias, correlaciones, etc.
- b. Selección de los modelos a probar
- c. Determinación de la metodología a utilizar
- d. Análisis y comparación de los resultados obtenidos
- e. Conclusiones (lecciones aprendidas)

### 2.2. Conformación de los Grupos

Los grupos serán conformados por 2 estudiantes.

### 2.3. Datasets a utilizar

Si bien existen una variedad de sitios donde descargar datasets para problemas de Aprendizaje Automático en general, se recomienda que todos los grupos escojan un dataset del siguiente repositorio digital: <http://archive.ics.uci.edu/ml/datasets>. Es posible, también, que los grupos seleccionen un dataset de otros repositorios. Sin embargo, no se permitirán datasets de sitios como *Kaggle*, ya que normalmente tienen asociado código a los mismos.

El grupo deberá seleccionar un dataset en un área que le interese. Si dos o más grupos seleccionan el mismo dataset, estos deberán ponerse de acuerdo, de tal forma que ningún grupo trabaje sobre el mismo dataset.

## 3. Partes del Proyecto

### 3.1. Propuesta

Una vez seleccionado un dataset, el estudiante deberá escribir una propuesta para la aprobación por parte del profesor. Esta propuesta deberá indicar claramente:

- El dataset escogido: No. instancias, No. atributos, características generales, etc. Además, el hipervínculo hacia la descripción del dataset.
- El objetivo del proyecto (que problema pretende solucionar), e.g. se desea realizar un análisis de sentimientos en base a críticas de productos.
- Metodología a aplicar: esquemáticamente, que pasos se van a realizar
- Selección de modelos: que modelos piensa entrenar y comparar. ¿Por que escogió esos modelos?

Esta propuesta no deberá exceder una plana y deberá ser presentada el día 05 de Febrero de 2021 durante la última sesión de ejercicios. Adicionalmente, los grupos deberán registrarse a través del siguiente documento compartido <https://bit.ly/36mpi4u>

### 3.2. Reporte Final

El reporte final deberá ser de no mas de 6 hojas, sin contar las referencias. El formato del reporte final deberá ser el siguiente:

- Abstract (1 párrafo)
- Introducción (0.5 páginas): contexto, necesidad (por que el problema es relevante), tarea (que es lo que se hizo para solucionar el problema e indicar claramente las entradas y salidas del problema), objetivo (objetivo general del proyecto).
- Trabajos relacionados (0.5 páginas): una revisión breve de trabajos similares.
- Descripción del dataset (0.5 - 1 páginas): No. de instancias de entrenamiento/prueba, ¿Se realizó algún tipo de pre-procesamiento?, ¿Se aplicó algún tipo de transformación?, etc.
- Metodología (1 - 1.5 páginas): Descripción de los pasos realizados para el proyecto.
- Resultados y Discusión (1 - 3 páginas): Resultados más relevantes de los experimentos realizados y su discusión. Por asuntos de espacio, solo incluir los resultados que considere más relevantes para elaborar sus conclusiones.
- Conclusión (1 - 3 párrafos): Conclusiones generales del proyecto. Cuales fueron las ideas principales, mejores modelos, como se obtuvieron mejores resultados, y trabajo futuro.
- Contribuciones (1 párrafo): Quién contribuyó con qué.
- Referencias (sin límite)

El código del proyecto deberá ser entregado en formato .zip o .rar. Para ello, se habilitará a tiempo un enlace en el evritual para que los grupos puedan cargar sus respectivos archivos.

Se sugiere que los grupos escojan un template en word o Latex de alguna conferencia/journal. Cuando suba su archivo, deberá hacerlo en formato pdf con la nomenclatura típica de nombre de archivo.

## 4. Evaluación

El reporte será juzgado en base a ciertos aspectos como: claridad, relevancia del tema, correcto uso de las técnicas, y lecciones aprendidas.

El puntaje máximo que se obtendrá por el proyecto son 15 puntos (50% de la nota del examen final). Adicionalmente, el trabajo será evaluado individualmente de forma oral.

La fecha de sustentación del proyecto será el miércoles 12 de Febrero de 2021 en horas de clase.