

IIC 2433 – Minería de datos
Tarea 5

1. Enunciado

La meta de esta tarea es entrenar un clasificador de texto que pueda reconocer atributos binarios sobre los legisladores chilenos, basado en las intervenciones que estos hacen en el parlamento.

Esta tarea es individual, se entrega como un notebook con todas las celdas corridas.

Fechas. La fecha de entrega de la tarea es el **Viernes 22 de Noviembre**, a las 20:00 hrs.

2. Tareas a resolver

Tienes acceso a dos archivos csv en

<https://drive.google.com/file/d/1WbW-RRyx5S0oaIKedvL7EZg0SuHsKveN/view?usp=sharing>.

El primero, `parlamentarios_info_general.csv` contiene información general sobre los parlamentarios vigentes hoy. Para esta tarea son de especial atención el identificador, el género, y su partido político. El segundo, `participacion_descripcion.csv`, contiene información sobre lo que han hablado cada uno de estos parlamentarios en los últimos años. Este documento cruza con la info de los parlamentarios usando el identificador.

Parte 1. Clasificador de género Debes implementar un clasificador que reciba un string y determine si ese string corresponde a un discurso hecho por un hombre o una mujer. Para esto puedes usar cualquiera de los métodos y algoritmos vistos en clases, pero no puedes usar algo que no hayamos visto. Tu clasificador debe ser implementado mediante una clase **ClasText** con dos métodos:

- Un método `fit`, que recibe una lista o un array de strings X , y una lista, array o series y que solo contiene números 0 o 1. Este método entrena un clasificador que puede predecir si un string dado debe ser etiquetado como 0 o 1, en base a lo visto en el entrenamiento.
- Un método `predict`, que recibe una lista o un array de strings X , y entrega una lista o un array de números 0 o 1 del mismo tamaño que X .

Parte 2. Análisis Contesta usando las herramientas vistas en clases: ¿cómo es el rendimiento de tu clasificador? Prueba ahora otros dos clasificadores distintos, usando otros algoritmos u otras formas de procesar texto. ¿Mejora el rendimiento?

Parte 3. Extension: clasificando ideología política Considera que los partidos de derecha son: Partido Unión Demócrata Independiente, Partido Republicano de Chile, Partido Renovación Nacional, Partido Social Cristiano, Partido Evolución Política (Evópoli), Partido Demócratas Chile y Movimiento Amarillos por Chile.

Considera que los partidos de izquierda son: Partido Comunista de Chile, Partido Por la Democracia, Partido Frente Amplio, Partido Socialista de Chile, Partido Radical Socialdemócrata, Partido Demócrata Cristiano, Partido Liberal de Chile, Federación Regionalista Verde Social, Partido Ecologista Verde, Partido Humanista y Partido Acción Humanista.

Muestra como puedes usar **ClasText** para clasificar discursos de alguien de izquierda o de derecha.

Parte 4. ¡Entregale nuevo texto a tus clasificadores! Ahora que tienes clasificadores, aprovecha de jugar un rato. Entregale textos que tu creas que pertenecen claramente a una clase u otra, y ve que predice tu clasificador. ¿Te hace sentido?