
	<b>Ingeniería de Sistemas y Computación</b> <b>Pregrado</b> <b>ISIS-3301 – Inteligencia de Negocios</b> <b>Semestre: 2021-10</b>	
---	---	---

## ENUNCIADO PRIMER PROYECTO

### **Objetivo.**

- Aplicar la metodología de minería de textos para la construcción de soluciones alineadas con los objetivos del negocio en un contexto de aplicación.

### **Descripción.**



La Minería de textos (MT), también conocida como Analítica de Textos, es un campo interdisciplinario que conjuga la Minería de Datos y el Procesamiento del Lenguaje Natural, y que tiene como objetivo procesar de manera automática grandes cantidades de textos para la extracción de conocimiento que apoye la toma de decisiones.

Las organizaciones, empresas e instituciones disponen de bases de datos documentales relacionadas con sus procesos de negocio y sería conveniente para ellas disponer de mecanismos que les faciliten un análisis automático de estos documentos. También, en muchos casos, sería de gran utilidad el análisis de información textual proveniente de diferentes fuentes como la Web, redes sociales, librerías digitales, entre otros. La MT proporciona soluciones en este sentido, al suministrar un conjunto de técnicas que permiten resolver problemas relacionados con la categorización automatizada de textos, la identificación de los temas ocultos presentes en una colección de documentos (modelado de tópicos), la síntesis de textos o creación automática de resúmenes, entre otras.

Una de las tareas más comunes de la minería de textos es la clasificación o categorización de documentos. Esta consiste en la asignación de una o más categorías predefinidas a un texto con base en su contenido. Su ámbito de aplicación es muy amplio como, por ejemplo: la clasificación de noticias, el análisis de sentimientos/opiniones en redes sociales, filtrado de correos spam, automatización de líneas de atención a clientes, entre otras.

Es en el contexto descrito previamente donde se desarrolla este proyecto, el cual tiene como objetivo general promover las competencias necesarias para una aplicación efectiva del proceso de descubrimiento de conocimiento a partir de textos. Se sugiere trabajar en contextos tales como:

- a) Seguridad informática, por ejemplo, analizando correos phishing con base en su contenido.
- b) Análisis de sentimientos/opiniones a partir de revisiones de los usuarios.

 <b>Universidad de los Andes</b> Colombia	<b>Ingeniería de Sistemas y Computación</b> <b>Pregrado</b> <b>ISIS-3301 – Inteligencia de Negocios</b> <b>Semestre: 2021-10</b>	
---	---	---

- c) Analítica en salud, por ejemplo, rastreando e identificando efectos imprevistos de medicamentos (como vacunas) a partir de comentarios de los consumidores.

La definición de las temáticas específicas del proyecto se trabajará en la sesión del 18 de febrero. Es indispensable que, previo a la sesión, propongan temáticas en alguno de los contextos sugeridos, o en otros similares, con el fin de seleccionar tres con datos puntuales a trabajar. Los grupos del proyecto podrán seleccionar una de esas temáticas, de acuerdo con las reglas establecidas en la sesión del 18 de febrero. Una temática consiste de un contexto, la definición clara del conjunto de datos de textos a analizar y de una organización que pueda beneficiarse y se concreta en una propuesta de enunciado corto.

Algunos repositorios públicos de datos de textos que pueden consultar son:

<https://www.kaggle.com/>

<https://research.google/tools/datasets/>

<https://analyticsindiamag.com/10-open-source-datasets-for-text-classification/>



<https://dasavisha.github.io/IWSPA-sharedtask/>

## **Descripción de la entrega.**

I. Documento que describa cómo fue aplicada la metodología de minería de textos y los resultados obtenidos en cada una de las fases:

- (15%) Comprensión del negocio y enfoque analítico. Definición de los objetivos y criterios de éxito desde el punto de vista del negocio. Determinación de las tareas de minería de textos que se considera son las adecuadas para alcanzar los objetivos del negocio. Descripción de cómo el requerimiento de negocio es resuelto con el o los requerimientos de minería propuestos, para lo cual debe diligenciar la tabla que se presenta a continuación:

Oportunidad/problema Negocio	



	<b>Ingeniería de Sistemas y Computación</b> <b>Pregrado</b> <b>ISIS-3301 – Inteligencia de Negocios</b> <b>Semestre: 2021-10</b>	
---	---	---

Descripción del requerimiento desde el punto de vista de minería de datos		
Detalles de la actividad de minería de datos		
Tarea	Técnica	Algoritmo y parámetros utilizados (con la justificación respectiva)

- (25%) Comprensión de los datos y preparación de los datos. Perfilamiento de los datos. Análisis de la calidad de los datos. Tratamiento de los datos (preparación o transformaciones requeridas), sobre la base del conocimiento del dominio y de los algoritmos seleccionados para resolver las tareas.
- (30%) Modelado y evaluación. Aplicación de las técnicas de minería de datos para la construcción de los modelos y presentación de los resultados de la evaluación cuantitativa. En este punto deben justificar la selección del modelo, mostrando las diferentes opciones probadas.
- (15%) Análisis de resultados. Descripción de los resultados obtenidos que permita a la organización comprenderlos, haciendo énfasis en el análisis de las medidas arrojadas por los modelos utilizados y cómo aportan en la consecución de los objetivos del negocio. Incluir posibles estrategias que la organización debe plantear relacionadas con los resultados obtenidos en los modelos y una justificación de por qué esa información es útil para ellos. Este resultado debe estar en el documento y deben generar un video y una presentación de máximo 7 minutos explicando su proyecto y resultados.

Importante: justifique cada decisión tomada en cada una de las fases.

- (15%) Trabajo en equipo

	<b>Ingeniería de Sistemas y Computación</b> <b>Pregrado</b> <b>ISIS-3301 – Inteligencia de Negocios</b> <b>Semestre: 2021-10</b>	
---	---	---

- (5%) Describir los retos enfrentados en el proyecto y las formas planteadas para resolverlos. De la misma manera deben describir la forma como se organizaron los integrantes del grupo para desarrollar este proyecto, lo cual incluye, definir las tareas realizadas por cada integrante y los tiempos, en número de horas, utilizados para realizarlas. Adicionalmente, de acuerdo con la planeación establecida para realizar el proyecto y la realidad en el desarrollo del mismo, reparta 100 puntos entre los integrantes del grupo y justifique su decisión.
- (10%) Sustentación y evaluación del aporte individual

II. Conjunto de datos resultado de la fase de comprensión y preparación de los datos. Requerido para realizar la sustentación.

### **Evaluación**

- El proyecto se realiza en grupos de 3 estudiantes.
- El documento a entregar tiene máximo 10 páginas (sin incluir portada, tabla de contenido, ni referencias), a una columna y con letra arial, tamaño 12.
- Deben entregar los modelos analíticos desarrollados (Sw) y todo lo relacionado con el proyecto para poder ejecutarlos en la sustentación.
- Se sustentará el proyecto y se realizará una evaluación del aporte de cada integrante en la solución propuesta para el proyecto.
- La fecha máxima de entrega es el **Domingo 14 de marzo 22:00.**