
 <p>Universidad de los Andes Colombia</p>	<p>Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Semestre: 2021-20</p>	
--	---	---

ENUNCIADO PRIMER PROYECTO

Objetivo.

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

Descripción.



La Analítica de Textos (AT) es un campo interdisciplinario que conjuga el aprendizaje de máquina y el Procesamiento de 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 AT 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 analítica 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) Analítica en salud, por ejemplo, rastreando e identificando efectos imprevistos de medicamentos (como vacunas) a partir de comentarios de los consumidores.
- b) Análisis de opiniones/sentimientos a partir de noticias, comentarios o revisiones de usuarios a productos o servicios.

 <p>Universidad de los Andes Colombia</p>	<p>Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Semestre: 2021-20</p>	
--	---	---

c) Deportes, música, turismo.

La definición de las temáticas específicas del proyecto se trabajará en la sesión del 2 de septiembre. 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 2 de septiembre. 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. Recuerden que, para trabajar tareas de clasificación, requieren datos que representen las distintas categorías en las cuales se concentra el análisis.

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

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

<https://datasets.simula.no/index.html>

<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 analítica de textos y los resultados obtenidos en cada una de las fases:



- (10%) **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 analítica 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 aprendizaje de máquina propuestos, para lo cual debe diligenciar la tabla que se presenta a continuación:

Oportunidad/problema Negocio	

	Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Semestre: 2021-20	
---	---	---

Descripción del requerimiento desde el punto de vista de aprendizaje de máquina	
Detalles de la actividad de minería de datos	
Tarea	Técnica Algoritmo e hiper-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 **mínimo tres algoritmos** diferentes para la técnica de aprendizaje de máquina seleccionada. En esta parte deben describir los algoritmos utilizados para la construcción de los modelos y presentación de los resultados de la evaluación cuantitativa. Adicionalmente, deben justificar la selección del modelo, mostrando los diferentes modelos probados e incluir los nombres de los estudiantes que trabajaron en cada uno. Al igual que en los laboratorios, en la nota habrá un porcentaje específico para el aporte individual.
- (15%) **Resultados.** Descripción y visualización en un tablero de control, 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

 <p>Universidad de los Andes Colombia</p>	<p>Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Semestre: 2021-20</p>	
--	---	---

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.

- (10%) Trabajo en equipo.
 - Para el trabajo en equipo se sugiere los siguientes roles para cada integrante.
 - **Líder de proyecto:** Está a cargo de la gestión del proyecto. Define las fechas de reuniones, pre-entregables del grupo y verifica las asignaciones de tareas para que la carga sea equitativa. Se encarga de subir la entrega del grupo. Si no hay consenso sobre algunas decisiones, tiene la última palabra.
 - **Líder de negocio:** Es responsable de velar por resolver el problema o la oportunidad identificada y estar alineado con la estrategia del negocio para el cual se plantea el proyecto. Debe garantizar que el producto se puede comunicar de forma apropiada.
 - **Líder de datos:** Se encarga de gestionar los datos que se van a usar en el proyecto y de las asignaciones de tareas sobre datos. Debe dejarlos disponibles para todo el grupo.
 - **Líder de analítica:** Se encarga de gestionar las tareas de analítica del grupo. Se encarga de verificar que los entregables cumplen con los estándares de análisis y que se tiene el “mejor modelo” según las restricciones existentes.
 - Se recomiendan al menos las siguientes reuniones de grupo:
 - **Reunión de lanzamiento y planeación:** Para definir roles y forma de trabajo del grupo. Se genera lluvia de ideas sobre el proyecto.
 - **Reunión de ideación:** Una vez se han definido los datos que proponen trabajar, la reunión de ideación busca definir el problema y los dolores.
 - **Reuniones de seguimiento:** Se recomienda mínimo una reunión de seguimiento semanal corta. También pueden ser correos de avance según lo defina el grupo. Pueden tener un tablero de control de las tareas, utilizando herramientas como Trello.

 <p>Universidad de los Andes Colombia</p>	<p>Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Semestre: 2021-20</p>	
--	---	---



- **Reunión de finalización:** Para consolidar el trabajo final, verificar el trabajo del grupo y analizar los puntos a mejorar para el siguiente proyecto.
 - Describir los roles que tuvo cada estudiante, al igual que el algoritmo trabajado y los retos enfrentados en el proyecto y las formas planteadas para resolverlos. De la misma manera deben describir 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.

III. Se debe crear un repositorio para el proyecto del curso, en el que compartirán el código de la solución creada. Puede acceder a beneficios de estudiante en <https://education.github.com/pack>. En el documento debe especificar el enlace de github de su proyecto (puede ser un repositorio público o compartirlo a través de githubfront).

Evaluación y entregables

- 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 publicar los resultados del proyecto usando el video generado, en el padlet <https://uniandes.padlet.org/mavillam/f40gjzxm5cf51mn8>, con clave: BI212
- Deben entregar los modelos analíticos desarrollados (Sw) y todo lo relacionado con el proyecto para poder ejecutarlos en la sustentación.
- **Se debe incluir la presentación y el video**
- 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 **sábado 2 de octubre 22:00.**

	Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Semestre: 2021-20	
---	---	---

Bonos.

En esta ocasión tendremos tres bonos:

- 1. Estudiantes cuyos temáticas y conjuntos de datos son seleccionados para trabajar el proyecto (0.3/5.0)**
- 2. Los grupos que obtengan los mejores dos resultados de acuerdo con las métricas de desempeño de los modelos propuestos (0.5/5.0)**
- 3. Certificación estadística (0.6/0.5)**

A continuación, describimos el bono de certificación estadística.

Objetivo.

- Revisar de forma detallada la parte estadística relacionada con el proyecto 1.



Descripción.

Uno de los elementos clave de éxito en los proyectos de analítica y en particular de minería de datos, es el fundamento estadístico que permite validar el desarrollo del proyecto y los resultados obtenidos.

Es ahí donde queremos darle la oportunidad de mejorar su comprensión en lo relacionado con temas estadísticos y la calidad del proyecto. Para lograr esto, tenemos a disposición del grupo un equipo de expertos en estadística¹, quienes revisarán su proyecto, utilizando inicialmente el video del resultado y el documento entregado. Ellos van a generar una serie de preguntas y recomendaciones relacionadas con la comprensión del proyecto y en particular con la parte estadística del proyecto, las cuales deben ser resueltas por ustedes como grupo. Esas respuestas serán analizadas por el equipo de estadística y si las respuestas son claras y satisfactorias ellos podrán darle la certificación de estadística que los hará merecedores de un bono de 0.6/5.0 en la nota del proyecto 1.

¡Esperamos que aprovechen esta oportunidad y puedan obtener el bono propuesto!!

¹ Grupo de estudiantes guiados por el profesor del departamento de matemáticas de Uniandes, Schweitzer Rocuts Pabón.

	Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Semestre: 2021-20	
---	---	---

El proceso de certificación, es el siguiente:

1. Manifestar el interés en el bono. La forma de participar es muy sencilla. En el padlet (<https://uniandes.padlet.org/mavillam/f40gjxm5cf51mn8> clave: BI212) donde reportan los resultados del proyecto, pueden inscribir a su grupo hasta el sábado 3 de octubre a las 22:00.
2. **Semana 12.** Se realizará una reunión entre los grupos de BI y de estadística, para aclarar dudas del contexto del proyecto
3. **Semana 15.** El grupo de estadística entrega retroalimentación al grupo de BI
 - a. Ver un ejemplo de [documento entregado](#).
4. **Semana 16.** El grupo del curso de BI entrega
 - a. Documento corto de la utilidad de las recomendaciones dadas por el grupo certificador y, las acciones a seguir para responder a esas recomendaciones
 - b. (opc.) video con las acciones a seguir para incluir las recomendaciones