
 <p>Universidad de los Andes Colombia</p>	<p>Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Semestre: 2022-20</p>	
--	---	---

ENUNCIADO PRIMER PROYECTO

La Analítica de Textos (AT) es un campo interdisciplinario que conjuga el *machine learning* (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 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 nos van a permitir resolver por ejemplo, problemas relacionados con:

- Categorización de documentos.
- Análisis de sentimientos/opiniones.
- Análisis de historias médicas, transacciones bancarias, documentos legales, aplicaciones laborales, entre otros.
- Clasificación automática de incidencias en centros de atención al cliente.
- Filtrado de documentos.
- Creación automática de resúmenes.
- Extracción de información especializada (en artículos científicos, reportes financieros, entre otros)

Es en el contexto descrito previamente donde se desarrolla este proyecto, el cual tiene como objetivo general fortalecer las competencias necesarias para una aplicación efectiva del proceso de descubrimiento de conocimiento a partir de textos, guiado por una metodología propia del mundo de la analítica.

A nivel de contexto y problemática a abordar en el proyecto se proponen las siguientes opciones, de las cuales, cada grupo debe seleccionar una:

- a) **Contexto de salud mental.** Apoyo a la detección de intentos de suicidio a partir de información recolectada de Reddit a nivel de comunidades que sufren de depresión o han intentado suicidarse. Los datos originales los pueden encontrar en este enlace

 <p>Universidad de los Andes Colombia</p>	<p>Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Semestre: 2022-20</p>	
--	---	---

(<https://www.kaggle.com/datasets/nikhileswarkomati/suicide-watch?resource=download>). Sin embargo, los datos a utilizar en el proyecto no son los mismo y los encuentra disponibles en la sección proyecto de BloqueNeón curso Unificado de BI.

- b) **Contexto de negocios.** Determinar el tipo o la calificación dada a un mensaje compartido por usuarios de Yelp. Los datos originales los pueden encontrar en este enlace (<https://www.yelp.com/dataset/documentation/main>). Sin embargo, los datos a utilizar en el proyecto no son los mismos y los encuentra disponibles en la sección proyecto de BloqueNeón curso Unificado de BI.

El desarrollo de este proyecto se realiza en dos etapas:

1. **Etap 1.** Construcción de modelos analíticos, donde el rol principal es el de un científico de datos.
2. **Etap 2.** Despliegue de modelos analíticos y desarrollo de una aplicación, donde el rol principal es el de un ingeniero de datos.

A continuación, se describe cada una de las etapas a desarrollar.

Etapa 1. Analítica de textos

Objetivo



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

Pasos previos

Cada grupo debe realizar los siguientes pasos, antes de iniciar el desarrollo del proyecto:

1. Determinar la temática a trabajar en el proyecto a partir de las opciones dadas en los contextos: a) Salud mental, b) Negocios.
2. Registrar la temática a trabajar por cada grupo en este archivo excel: <https://docs.google.com/spreadsheets/d/1aISBeRl1S6SiwzCftVvquEO25cGLCDamSOygVrfSUGs/edit?usp=sharing>. Este archivo es el mismo donde registro su grupo de trabajo para los laboratorios.

Descripción de la entrega

	Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Semestre: 2022-20	
---	---	---



A continuación, se describen cada uno de los entregables solicitados en esta etapa.

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 del enfoque analítico para alcanzar los objetivos del negocio. Descripción de cómo el requerimiento de negocio es resuelto por el enfoque analítico propuesto, para lo cual debe diligenciar la tabla que se presenta a continuación:

Oportunidad/problema Negocio	
Enfoque analítico (Descripción del requerimiento desde el punto de vista de aprendizaje de máquina)	
Organización y rol dentro de ella que se beneficia con la oportunidad definida	
Técnicas y algoritmos a utilizar	

- (25%) **Entendimiento y preparación de los datos.** Perfilamiento y análisis de la calidad de los datos. Tratamiento de los datos (preparación o transformaciones requeridas), de acuerdo con el dominio, las técnicas y los algoritmos seleccionados para resolver las tareas. Recuerde que es importante definir criterios para determinar los tokens que van a utilizar a nivel del modelo ya que esto afecta la

 <p>Universidad de los Andes Colombia</p>	<p>Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Semestre: 2022-20</p>	
--	---	---

calidad y el tiempo de construcción del modelo. Por ejemplo, incluir tokens que tienen alta frecuencia no genera información útil para el modelo.



- (30%) **Modelado y evaluación.** Aplicación de **mínimo tres algoritmos** diferentes para la tarea 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 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.

- (8%) Trabajo en equipo.
 - Para el trabajo en equipo se sugiere los siguientes roles para cada integrante. Un integrante puede desempeñar más de un rol.
 - **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.

 <p>Universidad de los Andes Colombia</p>	<p>Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Semestre: 2022-20</p>	
--	---	---

- **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 la forma de resolver el proyecto.
 - **Reunión de ideación:** Una vez se han explorado los datos del proyecto, la reunión de ideación busca definir el problema y los dolores que van a resolver.
 - **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.
 - **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 y las tareas realizadas por cada integrante del grupo, los tiempos, en número de horas, dedicados para realizar lo que le fue asignado, al igual que el algoritmo trabajado, los retos enfrentados en el proyecto y las formas planteadas para resolverlos. Adicionalmente, de acuerdo con la planeación establecida para realizar el proyecto y la realidad en el desarrollo del mismo, reflexionen sobre la forma como repartiría 100 puntos entre los integrantes del grupo e identifique puntos a mejorar para la siguiente entrega del proyecto.
- (12%) 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.

	Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Semestre: 2022-20	
---	---	---

III. Se debe crear un repositorio para el proyecto del curso, en el cual debe crear una wiki para incluir los diferentes entregables. Debe almacenar el código de la solución creada y referenciarlo en los entregables. Puede acceder a beneficios de estudiante en <https://education.github.com/pack>. El entregable en BloqueNeón es la dirección de la wiki (dentro del repositorio), donde registró todo lo solicitado en esta etapa del proyecto. El github de su proyecto puede ser un repositorio público o compartido a través de githubfront para garantizar que los tutores tengan acceso.

Condiciones de la entrega

- El proyecto se realiza en grupos de 3 estudiantes. Los estudiantes pueden estar en distintas secciones.
- 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/cttzj7kw3elic09a> con clave: BI222
- Deben entregar los modelos analíticos desarrollados (Software) y todo lo relacionado con el proyecto para poder ejecutarlos el día de 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 **miércoles 19 de octubre, 22:00 pm.**



Bono

Los grupos que obtengan los mejores dos resultados de acuerdo con la métrica de desempeño f1 de los modelos propuestos (0.5/5.0)

Etapas 2. Automatización de analítica de textos

Objetivos

- Automatizar un proceso replicable para aplicar la metodología de analítica de textos en la construcción de modelos analíticos

	Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Semestre: 2022-20	
---	---	---

- Desarrollar una aplicación que utilice un modelo analítico basado en aprendizaje automático y sea de interés para un rol dentro de una organización.

Contexto



El desarrollo de soluciones analíticas basadas en modelos analíticos para una organización incluye la participación de diferentes roles. En particular, en ese proceso de desarrollo, haremos énfasis en los roles del científico de datos, ingeniero de datos e ingeniero de software.

El científico de datos tiene la responsabilidad de seguir un proceso replicable y de calidad, para entender los datos, prepararlos, construir el modelo analítico y evaluarlo, tanto a nivel cuantitativo como cualitativo. Este proceso fue realizado en gran medida en la etapa 1 de este proyecto. Por su parte, el ingeniero de datos tiene la responsabilidad de tomar el resultado entregado por el científico de datos, completarlo incluyendo elementos como pruebas, organización de código, validación de atributos de calidad (e.g., eficiencia), además de persistir el modelo generado en un repositorio específico para este fin. Adicionalmente, deben extender el código para incluir el registro o log, que permita en caso de falla, revisar las entradas al modelo, los datos utilizados para el entrenamiento y prueba del modelo y el resultado de la preparación. Finalmente, el ingeniero de software toma ese resultado como entrada de la aplicación web o móvil que debe diseñar y construir, la cual es utilizada por el usuario final para ingresar su solicitud al modelo, interactuar con el resultado y apoyar sus acciones o decisiones. Esta etapa 2 del proyecto se centra en este proceso de desarrollo de la solución analítica.

Descripción de la entrega



La entrega de esta etapa del proyecto contiene un documento que describe el proceso de despliegue de la solución analítica en el ambiente de producción de una organización. Esto incluye la automatización del proceso desde la preparación de los datos hasta la persistencia del modelo analítico construido, al igual que el dejar disponible una aplicación que toma el resultado del modelo y lo pone a disposición de un usuario final. Adicionalmente, como parte de la entrega se debe incluir un video y todo el software desarrollado para el proyecto. En particular debe entregar:

- (20%) **Proceso de automatización del proceso de preparación de datos, construcción del modelo, persistencia del modelo y acceso por medio de API:**

	Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Semestre: 2022-20	
---	---	---

Descripción del proceso e implementación realizados por el ingeniero de datos, acompañado del código. En este punto es importante que revise si es posible enriquecer el modelo analítico construido previamente con integración de nuevas columnas del mismo archivo original o de otras fuentes para obtener mejores resultados tanto desde el punto de vista cuantitativo, como cualitativo. Se sugiere utilizar *pipelines* para este proceso y un Framework para desarrollo de la API.

- (40%) **Desarrollo de la aplicación y justificación.** Descripción del usuario/rol de la organización que va a utilizar la aplicación, la conexión entre esa aplicación y el proceso de negocio que va a apoyar (si aplica), y la importancia que tiene para ese rol la existencia de esta aplicación. Adicionalmente, el desarrollo de una aplicación web o móvil para interactuar con el resultado del modelo (y su calidad) a partir de un texto o textos datos por el usuario.
- (18%) **Resultados.** Video de máximo ocho (7) minutos con la descripción y visualización en la aplicación de los resultados del modelo, que permita a un rol dentro de la organización comprenderlos y usarlos. El video debe simular la interacción del usuario final con la aplicación y describir dos acciones que puede realizar como resultado de dicha interacción, haciendo énfasis en la forma como el resultado del modelo aporta en esas acciones.
- (10%) **Trabajo en equipo.**
 - Para el trabajo en equipo se sugiere los siguientes roles para cada integrante. Un integrante puede desempeñar más de un rol.
 - **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.
 - **Ingeniero de datos:** Es responsable de velar por la calidad del proceso de automatización relacionado con la construcción del modelo analítico.
 - **Ingeniero de software responsable del diseño de la aplicación y resultados:** Se encarga de liderar el diseño de la aplicación y de la generación del video con los resultados obtenidos.
 - **Ingeniero de software responsable de desarrollar la aplicación final:** Se encarga de gestionar el proceso de construcción de la aplicación.
 - Se recomiendan al menos las siguientes reuniones de grupo:

	Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Semestre: 2022-20	
---	---	---



- **Reunión de lanzamiento y planeación:** Para definir roles y forma de trabajo del grupo. Se genera lluvia de ideas sobre la forma de resolver el proyecto.
 - **Reuniones de seguimiento:** Se recomienda mínimo dos reuniones de seguimiento semanal cortas. También pueden ser correos de avance según lo defina el grupo. Pueden tener un tablero de control de las tareas, utilizando herramientas como Trello.
 - **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 y las tareas realizadas por cada integrante del grupo, los tiempos, en número de horas, dedicados para realizar lo que le fue asignado, al igual que los retos enfrentados en el proyecto y las formas planteadas para resolverlos. Adicionalmente, de acuerdo con la planeación establecida para realizar el proyecto y la realidad en el desarrollo del mismo, reflexionen sobre la forma como repartiría 100 puntos entre los integrantes del grupo e identifique puntos a mejorar para la siguiente entrega del proyecto.
 - (12%) Sustentación y evaluación del aporte individual

Adicionalmente deben:

1. Entregar el conjunto de datos resultado de la fase de entendimiento y preparación de los datos. Requerido para realizar la sustentación.
2. Especificar en el enlace de la wiki en el github de su proyecto (puede ser un repositorio público o compartirlo a través de githubfront). Todo el código debe quedar en el repositorio del proyecto del curso y referenciado en los entregables. Recuerden que pueden acceder a beneficios de estudiante en <https://education.github.com/pack>.

Evaluación y entregables

- El proyecto se realiza en grupos de 3 estudiantes. Los estudiantes pueden estar en distintas secciones.

 <p>Universidad de los Andes Colombia</p> <p> <small> Acreditación Institución de alta calidad 10 años Mineducación Resolución 1853 9 de enero de 2015 </small> </p>	<p>Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Semestre: 2022-20</p>	
---	--	---

- 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/cttzj7kw3elic09a> con clave: BI222
- Deben entregar todo lo relacionado con el proyecto para poder ejecutarlo el día de la sustentación.
- **Se debe incluir 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 **domingo 13 de noviembre, 22:00 pm.**