
	<b>Ingeniería de Sistemas y Computación</b> <b>Pregrado</b> <b>ISIS-3301 – Inteligencia de Negocios</b> <b>Primer Proyecto</b> <b>Semestre: 2024-20</b>	
---	---	---

## Contenido

I. Descripción general Fondo de Poblaciones de las Naciones Unidas .....	1
II. Entregas .....	2
III. Actividades .....	2
IV. Etapa 1. Construcción de modelos de analítica de textos .....	2
V. Etapa 2. Automatización y uso de modelos de analítica de textos .....	8
VI. Condiciones de las entregas.....	11



## I. Descripción general Fondo de Poblaciones de las Naciones Unidas

El Fondo de Poblaciones de las Naciones Unidas (UNFPA<sup>1</sup>) junto con entidades públicas y haciendo uso de diferentes herramientas de participación ciudadana, busca identificar problemas y evaluar soluciones actuales, relacionando la información dada por los ciudadanos con los diferentes Objetivos de Desarrollo Sostenible (ODS). Los ODS fueron adoptados por las Naciones Unidas en 2015 como un llamamiento universal para poner fin a la pobreza, proteger el planeta y garantizar que para el 2030 todas las personas disfruten de paz y prosperidad<sup>2</sup>.

En este contexto, uno de los procesos que requiere mayor esfuerzo es el análisis de la información textual recopilada, ya que consume muchos recursos, que incluyen la participación de un experto. Es así como el UNFPA quiere desarrollar un proyecto con ustedes, donde el objetivo principal es relacionar de forma automática opiniones de los ciudadanos con los ODS 3, 4 y 5. A nivel de la solución a plantear deben aplicar la metodología de desarrollo de aplicaciones analíticas para crear un modelo analítico que sea utilizado y reentrenado por medio de una aplicación web o móvil a partir de un conjunto de opiniones que contienen texto en lenguaje natural.

<sup>1</sup> <https://colombia.unfpa.org/es/unfpa-en-colombia>

<sup>2</sup> <https://www.undp.org/es/sustainable-development-goals>

	<p>Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Primer Proyecto Semestre: 2024-20</p>	
---	---	---

## II. Entregas

El desarrollo de este proyecto se realiza en dos etapas:

**Etapla 1. Construcción de modelos de analítica de textos**, donde el rol principal es el de un científico de datos. El trabajo se centra en la construcción de modelos analíticos sin dejar de lado el entendimiento del negocio, los datos y por supuesto la evaluación y utilidad de los resultados obtenidos con el modelo seleccionado. **Fecha máxima de entrega 7 de septiembre 20:00.**

**Etapla 2. Automatización y uso de modelos de analítica de textos**, donde el rol principal es el de un ingeniero de datos. Esta etapa se enfoca en el desarrollo de la aplicación para un usuario final y en garantizar que la solución analítica que incluye la construcción del modelo analítico funciona en diferentes momentos del tiempo con nuevos conjuntos de datos. **Fecha máxima de entrega 12 de octubre 20:00.**

**Si la entrega la hacen después de las 20:00 y hasta las 2:00 am del siguiente día, tendrá una penalización del 30%. Cualquier entrega posterior a las 2:00 a.m. por otros canales tendrá una nota de 0/5.**

## III. Actividades



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

- Conformar su grupo de mínimo 2 y máximo 3 estudiantes.** Recuerde que los grupos de proyecto pueden involucrar estudiantes de distintas secciones y pueden ser diferentes a los grupos de los laboratorios.
- Registrar el grupo** en BloqueNeón. Deben registrar su grupo en BloqueNeón para habilitar el enlace de entrega del proyecto.
- Leer todo este documento**

## IV. Etapa 1. Construcción de modelos de analítica de textos

### Objetivos

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

 <p>Universidad de los Andes Colombia</p>	<p>Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Primer Proyecto Semestre: 2024-20</p>	
--	---	---

### **Descripción de la etapa 1**

En esta etapa será científico de datos. 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. En este caso, cada grupo debe trabajar con opiniones que representan la voz de los habitantes locales para identificar las problemáticas de su entorno en términos de los ODS 3,4 y 5. En esa tarea se requiere construir un modelo analítico que permita relacionar de forma automática nuevas opiniones, al igual que analizar las palabras que llevan a relacionar una opinión con un ODS.



### **Descripción de los entregables**

A continuación, se presentan las secciones que debe incluir el documento y los diferentes entregables a realizar y espacios de evaluación.

#### **Sección 1. (10%) Entendimiento del negocio y enfoque analítico.**

Complete la siguiente tabla:

Oportunidad/problema Negocio	
Objetivos y criterios de éxito desde el punto de vista del negocio.	
Organización y rol dentro de ella que se beneficia con la oportunidad definida	
Impacto que puede tener en Colombia este proyecto.	

	<b>Ingeniería de Sistemas y Computación</b> <b>Pregrado</b> <b>ISIS-3301 – Inteligencia de Negocios</b> <b>Primer Proyecto</b> <b>Semestre: 2024-20</b>	
---	---	---

Enfoque analítico. Descripción de la categoría de análisis (descriptivo, predictivo, etc.) , tipo y tarea de aprendizaje e incluya las técnicas y algoritmos que propone utilizar.



**Sección 2. (20%) Entendimiento y preparación de los datos.** Perfilamiento y análisis de la calidad de los datos. Además, deben incluir el tratamiento de los datos (preparación o transformaciones requeridas), de acuerdo con el dominio, las técnicas y los algoritmos seleccionados para resolver el problema planteado.

**Sección 3. (20%) Modelado y evaluación.** Aplicación de **mínimo tres algoritmos** diferentes para la tarea de aprendizaje automático seleccionada. En esta parte deben describir los algoritmos utilizados para la construcción de los modelos y presentar los resultados de las métricas de evaluación para justificar la selección del modelo. Además, deben incluir los nombres de los estudiantes que trabajaron en cada modelo analítico.

Al igual que en los laboratorios, en la nota del proyecto hay un porcentaje específico para el aporte de cada miembro del grupo.

**Sección 4. (20%) Resultados.**

- Descripción de los resultados obtenidos, que permita a la organización comprenderlos, haciendo énfasis en el análisis de las métricas de calidad arrojadas por los modelos utilizados y cómo aportan en la consecución de los objetivos del negocio.
- Incluir el análisis de las palabras identificadas para relacionar las opiniones con los ODS y posibles estrategias que la organización debe plantear utilizando los resultados obtenidos en los modelos analíticos y una justificación de por qué esa información es útil para ellos.
- Entregar los datos de prueba compartidos, en formato excel, con una columna adicional que contiene la etiqueta asignada por el modelo que seleccionaron. Este archivo se utiliza para determinar el f1-score de su modelo analítico y compararlo con los de los diferentes grupos para asignar parte de la nota del grupo.

 <div> <b>Universidad de los Andes</b> Colombia </div> <div> <small>           Acreditación Institución de alta calidad <b>10 años</b> Ministerio de Educación Presidencia de la Rectoría 9 de enero de 2024         </small> </div>	<b>Ingeniería de Sistemas y Computación</b> <b>Pregrado</b> <b>ISIS-3301 – Inteligencia de Negocios</b> <b>Primer Proyecto</b> <b>Semestre: 2024-20</b>	
---	---	---



- d. Generar un video de **máximo 5 minutos** explicando su proyecto y resultados. El video debe ser publicado en el padlet respectivo (ver condiciones de entrega).

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

### Sección 5. (10%) Mapa de actores relacionado con el producto de datos creado.

Descripción de una organización que puede beneficiarse del resultado del modelo analítico planteado. Dentro de la organización identifique posibles tipos de actores relacionados con el producto tales como clientes/financiadores/proveedores/beneficiados/etc. y plantee un beneficio y un riesgo que puede tener dicho tipo de actor. Por ejemplo, en el contexto de un modelo de deserción de estudiantes en una universidad, este entregable podría ser el siguiente:

Rol dentro de la empresa	Tipo de actor	Beneficio	Riesgo
Coordinación académica	Usuario-cliente	Apoya la labor de consejería preventiva de los estudiantes	Si el modelo no tiene un buen desempeño, puede estar alertando a estudiantes que no van a desertar o dejando de lado los que sí.
Dirección de ingeniería de sistemas y computación	Financiador	Mecanismo de toma de decisiones basado en información, menos sensible a la intuición del personal	En caso de que el modelo no funcione es dinero mal invertido y pudo dejarse de hacer un proyecto con mayor impacto y viabilidad.
Dirección de servicios de tecnologías de información	Proveedor	Garantiza el cumplimiento de estándares de calidad de los productos desarrollados, que incluye elementos como la seguridad y privacidad de los datos utilizados.	Manejo incorrecto de los datos que lleve a la violación de la privacidad de los datos.
Estudiante	Beneficiado	Recibe una consejería a tiempo que le permite buscar las ayudas que pueden impedir su deserción del programa.	Afectación psicológica que lo lleva a bajar el nivel académico dado que fue alertado o



	<b>Ingeniería de Sistemas y Computación</b> <b>Pregrado</b> <b>ISIS-3301 – Inteligencia de Negocios</b> <b>Primer Proyecto</b> <b>Semestre: 2024-20</b>	
---	---	---

			acompañado de forma incorrecta.
--	--	--	---------------------------------

Ten presente esta frase de Albert Einstein: "La preocupación por el hombre y su destino debe ser siempre el principal interés de todos los esfuerzos técnicos. Nunca olvides esto en medio de tus diagramas y ecuaciones".

### Sección 6. (8%) Trabajo en equipo

- Describir los roles y las tareas realizadas por cada integrante del grupo (ver sugerencias en la siguiente viñeta), 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. Además, según la planeación establecida para realizar el proyecto y la realidad en el desarrollo de este, reflexionen sobre cómo repartir 100 puntos entre los integrantes del grupo e identifique puntos a mejorar para la siguiente entrega del proyecto.
- A nivel del trabajo en equipo se sugieren 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.
  - **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 y garantizar la entrega en el repositorio de git.
  - **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:

 <p>Universidad de los Andes Colombia</p>	<p>Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Primer Proyecto Semestre: 2024-20</p>	
--	---	---



- **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 la organización/empresa/institución y el rol dentro de ella, que se beneficia de la solución analítica que van a desarrollar.
- **Reuniones de seguimiento:** Se recomienda mínimo una reunión de seguimiento semanal corta. También pueden ser correos de avance según lo defina el grupo.
- **Reunión de finalización:** Para consolidar el trabajo final, verificar el trabajo del grupo y analizar los puntos a mejorar para la siguiente etapa del proyecto.

#### Otros entregables y espacios de evaluación

- (12%) Sustentación y evaluación del aporte individual. Se revisa lo entregado y se valida el aporte individual y la concordancia entre el documento y la sustentación.
- Conjunto de datos resultado de la fase de entendimiento y preparación de los datos. Requerido para realizar la sustentación.
- Repositorio:** 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 monitores tengan acceso.

#### **Bono**

**Los grupos que obtengan los mejores cinco resultados de acuerdo con las métricas de desempeño de los modelos propuestos (0.5/5.0)**

 <p>Universidad de los Andes Colombia</p>	<p>Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Primer Proyecto Semestre: 2024-20</p>	
--	---	---

## V. Etapa 2. Automatización y uso de modelos de analítica de textos

### **Objetivos**

- Automatizar un proceso replicable para aplicar la metodología de analítica de textos en la construcción de modelos analíticos.
- Desarrollar una aplicación que utilice un modelo analítico basado en aprendizaje automático y sea de interés para una organización, empresa o institución y en particular para un rol existente en alguna de ellas.

### **Descripción de la etapa 2**

Esta etapa se centra en el rol de ingeniero de datos y hay actividades del ingeniero de software. 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. 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.



En esta etapa 2 deben automatizar el desarrollo del modelo de analítica de textos creado en la etapa 1 del proyecto y desarrollar una aplicación orientada a un usuario con sentido y apoyo a cualquier acción o decisión de este.

### **Descripción de la entrega**

A continuación, se presentan las secciones que debe incluir el documento y los diferentes entregables a realizar y espacios de evaluación.

**Sección 1. (20%) Proceso de automatización del proceso de preparación de datos, construcción del modelo, persistencia del modelo y acceso por medio de API:** Descripción del proceso e implementación realizados por el ingeniero de datos, acompañado del código.



	<b>Ingeniería de Sistemas y Computación</b> <b>Pregrado</b> <b>ISIS-3301 – Inteligencia de Negocios</b> <b>Primer Proyecto</b> <b>Semestre: 2024-20</b>	
---	---	---

Debe utilizar *pipelines* para este proceso y un Framework para desarrollo de la API. La API REST debe estar compuesta por dos “endpoints”:



- El primero debe recibir mediante el "body" una o más instancias de datos con la totalidad de las características requeridas para las que se desea realizar la predicción. Este “endpoint” debe devolver una lista con la misma cardinalidad de las instancias recibidas, y en el mismo orden, que contenga las predicciones realizadas por el modelo para cada instancia de datos recibida.
- El segundo “endpoint”, debe estar en capacidad de enviar una cantidad relevante de instancias de datos que servirán para realizar un proceso de re-entrenamiento del modelo, razón por la cual, además de enviar las características, también será requisito enviar la variable objetivo. Como respuesta, el “endpoint” debe devolver algunas métricas de desempeño como pueden ser el Precision, Recall, F1-score. Además, tenga presente que con el proceso de re-entrenamiento del modelo, el archivo binario deberá reemplazarse para ser tenido en cuenta como nueva versión la próxima vez que se solicite realizar una predicción mediante el primer “endpoint”.

Los datos tanto de la solicitud como de la respuesta deben estar en formato JSON y deben respetar el esquema del CSV original proporcionado para este proyecto 1. Recuerde que el pipeline es el encargado de hacer todas las preparaciones requeridas previo a hacer los procesos de entrenamiento o predicción.

**Sección 2. (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, debe desarrollar una aplicación web o móvil para interactuar con el resultado del modelo analítico a partir de un texto o textos datos por el usuario. Recuerde que la interacción incluye tanto la predicción como el *score* asociado a la misma.

**Sección 3. (18%) Resultados.** Video de máximo 5 minutos con la descripción y visualización en la aplicación de los resultados del modelo analítico, que permita a un rol dentro de la organización comprenderlos y usarlos. El video publicado en el padlet respectivo (ver condiciones de entrega), debe simular la interacción del usuario final con la aplicación y describir dos acciones que puede realizar como resultado de dicha interacción, haciendo



	<p>Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Primer Proyecto Semestre: 2024-20</p>	
---	---	---

énfasis en la forma como el resultado del modelo aporta en esas acciones y reflexionando sobre el posible impacto a nivel Colombia que presentó en la entrega 1.

Realice pruebas de facilidad de uso de la aplicación y utilidad para el usuario para el cual se diseñó, puede vincular a otros estudiantes del curso o personas cercanas para esta validación.

#### Sección 4. (10%) Trabajo en equipo.

- Para el trabajo en equipo se sugieren 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:
  - **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.
  - **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. Además, según la planeación establecida para realizar el proyecto y la realidad en el desarrollo de este, reflexionen sobre cómo repartiría 100 puntos entre los

 <p>Universidad de los Andes Colombia</p>	<p>Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Primer Proyecto Semestre: 2024-20</p>	
--	---	---

integrantes del grupo e identifique puntos a mejorar para la siguiente entrega del proyecto.

### **Otros entregables**

- I. (12%) **Aporte individual.** Con base en lo descrito en el documento se calcula este aporte.
- II. Entregar el conjunto de datos resultado de la fase de entendimiento y preparación de los datos. Requerido para realizar la sustentación.
- III. Utilizar el github de la entrega anterior para dejar el código y dese ser referenciado en los entregables. Recuerden que pueden acceder a beneficios de estudiante en <https://education.github.com/pack>.

## VI. Condiciones de las entregas

- El proyecto se realiza en grupos de 3 estudiantes. Los estudiantes pueden estar en distintas secciones.
- El documento 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/exposicion-proyecto-analitica-de-texto-de-bi-202420-2bbr4z5cypyb3bis> con clave: BIProyBI242
- Deben entregar todo el software relacionado con el proyecto para poder ejecutarlo durante la sustentación o revisión del proyecto.