
 Universidad de los Andes Colombia	Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Primer Proyecto Semestre: 2025-10	
---	--	---

Contenido

I. Descripción general sobre noticias falsas en política 1

II. Entregas 1

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 12

I. Descripción general sobre noticias falsas en política

I. La divulgación de noticias falsas sobre política a nivel nacional e internacional es una amenaza que puede impactar la gobernabilidad de países y la seguridad global. A nivel interno de un país puede generar polarización y manipulación de la opinión pública debilitando la confianza en instituciones democráticas. Esto puede llevar a influir en elecciones, justificar medidas impopulares entre otras consecuencias. A nivel internacional, estas noticias pueden generar tensiones diplomáticas, influir en decisiones de política exterior y ser usadas en campañas de desestabilización por parte de gobiernos o grupos específicos. Esta reflexión motivó a un grupo de académicos a recolectar y generar noticias con el fin de utilizar aprendizaje automático para facilitar la detección de noticias falsas. Este conjunto de datos es el que se va a utilizar en el contexto de este proyecto y está disponible en la sección de Bloque Neón de la sección unificada del curso. Los datos originales contienen noticias de periódicos en línea como: 'Público,' 'La Marea,' y 'El Común.' y tienen licencia **ATTRIBUTION 4.0 INTERNATIONAL**, la cual pueden consultar en este enlace: [Deed - Attribution 4.0 International - Creative Commons](#).

II. Entregas

El desarrollo de este proyecto se realiza en dos etapas:

 Universidad de los Andes Colombia	Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Primer Proyecto Semestre: 2025-10	 ABET Engineering Accreditation Commission
---	--	---

Etapas 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 22 de febrero 20:00.**

Etapas 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 29 de marzo 20:00.**

Si la entrega o la fecha de los entregables está después de las 20:00 y hasta las 2:00 am del siguiente día, tendrá una penalización del 30% y su entrega se califica sobre 3.5. Cualquier entrega posterior a las 2:00 a.m. por otros canales tendrá una nota de 0.

III. Actividades



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

- a. **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.
- b. **Registrar el grupo.** Deben registrar su grupo en BloqueNeón en la categoría “Grupos proy1 Etapa 1” para habilitar el enlace de entrega del proyecto.
- c. **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.

 Universidad de los Andes Colombia	Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Primer Proyecto Semestre: 2025-10	
---	---	---

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 noticias recolectadas de varios periódicos en línea y enriquecidas con noticias falsas. El objetivo es detectar noticias falsas e identificar palabras presentes en dichas noticias que guíen a un usuario final en la propuesta de acciones para mitigar esa situación.






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. (20%) Documentación del proceso de aprendizaje automático.

Complete el siguiente Canvas de aprendizaje automático. Adaptación de *OWNML MACHINE LEARNING CANVAS*. [CRISP-ML(Q). The ML Lifecycle Process]



 <div>Universidad de los Andes Colombia</div> <div><div>Acreditación institucional de alta calidad</div><div>10 años</div><div>Mineducación</div><div>Resolución 161 del 9 de enero de 2015</div></div>	Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Primer Proyecto Semestre: 2025-10	
--	--	---

<div>TAREA DE APRENDIZAJE</div> <div></div> <div>¿Cuál es el tipo de aprendizaje? ¿Si es aprendizaje supervisado, indicar qué se predice? ¿Cuáles son los posibles resultados de la tarea de aprendizaje? ¿Cuándo se observan los resultados de esta tarea? Por ejemplo, si es un modelo predictivo indicar si el resultado se obtiene unas horas, días, semanas, o meses antes.</div>	<div>DECISIONES</div> <div></div> <div>¿Cómo se convierten los resultados del modelo en recomendaciones o decisiones procesables para el usuario final?</div>	<div>PROPUESTA DE VALOR</div> <div></div> <div>¿Quién es el beneficiario final? ¿De qué empresa es? ¿Qué problemas específicos se abordan? ¿Qué riesgo puede tener para ese beneficiario el uso de este modelo?</div>	<div>RECOLECCIÓN DE DATOS – NO SE DEBE DILIGENCIAR</div> <div></div> <div>¿Cómo se obtiene el conjunto inicial de entidades y resultados (por ejemplo, extractos de bases de datos, extracciones de API, etiquetado manual)? ¿Qué estrategias se aplican para actualizar los datos continuamente, controlando los costos y manteniendo la vigencia?</div>	<div>FUENTES DE DATOS</div> <div></div> <div>¿Qué fuentes de datos se utilizan? (Mencione tablas de bases de datos internas y externas o métodos API). ¿De dónde se toman los datos? ¿Se pueden utilizar para realizar el objetivo del análisis?</div>
---	--	--	--	---

<div><div><div>Universidad de los Andes Colombia</div></div><div><div>Acreditación institucional de alta calidad</div><div>10 años</div><div>Mineducación Permisos 101 del 9 de enero de 2015</div></div></div>	Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Primer Proyecto Semestre: 2025-10	<div><div><div>ABET</div></div><div>Engineering Accreditation Commission</div></div>
--	--	---

<div><div>SIMULACIÓN DE IMPACTO</div><div></div><div>¿Cuáles son los valores de costo/beneficio de las decisiones (in)correctas? ¿Cuáles son los criterios de éxito del modelo para su posterior despliegue? ¿Existen restricciones de equidad?</div></div>	<div><div>APRENDIZAJE (USO DEL MODELO)</div><div></div><div>¿El uso del modelo es por lotes o en tiempo real? ¿Con qué frecuencia se usa?</div></div>	<div><div>CONSTRUCCIÓN DE MODELOS</div><div></div><div>¿Cuántos modelos se necesitan? ¿Cuándo deben actualizarse? ¿De cuánto tiempo se dispone para generar el modelo (incluido el proceso de ingeniería de características y el análisis o evaluación del mismo)?</div></div>	<div><div>INGENIERÍA DE CARACTERÍSTICAS</div><div></div><div>¿Qué variables/características se utilizan en el modelo? ¿Qué agregaciones o transformaciones se aplican a las fuentes de datos originales – incluir las más importantes--?</div></div>
	<div><div>MONITOREO NO SE DEBE DILIGENCIAR</div><div></div><div>¿Qué métricas y KPI se utilizan para hacer un seguimiento del impacto de la solución de ML una vez desplegada, tanto para los usuarios finales como para la empresa? ¿Con qué frecuencia deben revisarse?</div></div>		



 Universidad de los Andes Colombia	Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Primer Proyecto Semestre: 2025-10	
---	--	---

Sección 2. (20%) Entendimiento y preparación de los datos (a nivel de código y análisis). 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. Es posible utilizar algoritmos no vistos en clase. 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.

Cada estudiante del grupo debe realizar al menos uno de esos algoritmos. La nota obtenida en esta sección se utiliza para asignar parte de la nota individual de la entrega.



Sección 4. (20%) Resultados.

- a. 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.
- b. Incluir el análisis de las palabras seleccionadas para identificar las noticias falsas y una justificación de por qué esa información es útil para la organización.
- c. Entregar los datos de prueba compartidos, en el mismo formato, 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 grupal.
- d. Generar un video de **máximo 5 minutos** explicando su proyecto y resultados. El video debe estar en la wiki respectiva.

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

Sección 5. (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, el algoritmo trabajado, los retos enfrentados en el proyecto, las formas planteadas para resolverlos y el uso que le dieron a ChatGPT en este proceso.

 Universidad de los Andes Colombia	Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Primer Proyecto Semestre: 2025-10	
---	--	---

- Reportar la forma como distribuyen 100 puntos entre los integrantes del grupo e identificar puntos a mejorar para la siguiente entrega del proyecto. Tenga en cuenta el compromiso en el logro de resultados, la calidad de la contribución, el cumplimiento de compromisos adquiridos con el grupo y la participación equitativa en el trabajo del grupo.
- 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 y se cubra la totalidad de los entregables. Se encarga de subir la entrega del grupo. Si no hay consenso sobre algunas decisiones, tiene la última palabra.
 - **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:
 - **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

- I. **(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.
- II. Conjunto de datos resultado de la fase de entendimiento y preparación de los datos. Requerido para realizar la sustentación.

 Universidad de los Andes Colombia	Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Primer Proyecto Semestre: 2025-10	
---	--	---

- III. **Repositorio:** Se debe crear un repositorio para el proyecto del curso, en el cual debe crear una wiki **para incluir todos los 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 URL 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 (0.3/5.0)

Si incluye nuevas fuentes o extiende la fuente actual para enriquecer el modelo analítico que va a construir, esto será bonificado, hasta (0.3/5.0)



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.

 Universidad de los Andes Colombia	Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Primer Proyecto Semestre: 2025-10	
---	--	---



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. Recuerden registrar el grupo en BloqueNeón en la categoría “Grupos proy1 Etapa 2” para tener acceso al enlace de entrega.

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. 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 y la probabilidad de esa predicción.
- 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”. En este punto debe investigar y plantear tres definiciones distintas de re-entrenamiento. En cada una de ellas debe incluir una descripción del significado, acompañada de una ventaja o desventaja de esa propuesta. Al final debe concluir indicando la que implementó en esta etapa del proyecto.

 Universidad de los Andes Colombia	Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Primer Proyecto Semestre: 2025-10	
---	---	---

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. En este punto en el momento de solicitar la información al usuario de la aplicación piense en la forma más sencilla para que él introduzca la información e internamente si lo requiere haga los ajustes para que la API funcione.



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.
- Desarrollo de una aplicación web o móvil para interactuar con el resultado del modelo analítico a partir de un texto o textos dados por el usuario. Recuerde que la interacción incluye tanto la predicción como la probabilidad asociada a la misma. A nivel de usuario experto en modelos analíticos, debe brindar la opción de re-entrenar el modelo, según la semántica que seleccionó.
- Resuelva las siguientes preguntas con respecto a la aplicación a desplegar: ¿Qué recursos informáticos requiere para entrenar, ejecutar, persistir el modelo analítico y desplegar la aplicación? ¿Cómo se integrará la aplicación construida a la organización, estará conectada con algún proceso del negocio o cómo se pondrá a disposición del usuario final? ¿Qué riesgos tiene para el usuario final usar la aplicación construida?

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 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 y reflexionando sobre el posible impacto a nivel Colombia el uso de la aplicación desarrollada.

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.

	<p>Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Primer Proyecto Semestre: 2025-10</p>	
---	---	---

- 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.
- Reportar la forma como distribuyen 100 puntos entre los integrantes del grupo e identificar puntos a mejorar para la siguiente entrega del proyecto. Tenga en cuenta el compromiso en el logro de resultados, la calidad de la contribución, el cumplimiento de compromisos adquiridos con el grupo y la participación equitativa en el trabajo del grupo.
- 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.
 - **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.

Otros entregables y elementos a evaluar

- 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.
- III. Utilizar el github de la entrega anterior para dejar el nuevo código y los diferentes los entregables. Recuerden que pueden acceder a beneficios de estudiante en <https://education.github.com/pack> u que lo que entrega en Bloque Neón es la dirección de la wiki donde están todos los entregables.

 Universidad de los Andes Colombia	Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Primer Proyecto Semestre: 2025-10	
---	---	---

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.
- El repositorio debe tener todo el código fuente de la aplicación y del modelo analítico para poder ejecutarlo durante la revisión del proyecto.