Grupo 2:

Pablo Ortega

Juana Mejía

Daniela Espinosa 202022615

**PROYECTO 1 – ETAPA 2**

**ITELIGENCIA DE NEGOCIOS**

**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. Ingeniería de Sistemas y Computación Pregrado ISIS-3301 – Inteligencia de Negocios Primer Proyecto Semestre: 2024-20 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 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 integrantes del grupo e identifique puntos a mejorar para la siguiente entrega del proyecto.

**Otros entregables:**

(12%) Aporte individual. Con base en lo descrito en el documento se calcula este aporte.

Entregar el conjunto de datos resultado de la fase de entendimiento y preparación de los datos. Requerido para realizar la sustentación.

Utilizar el github de la entrega anterior para dejar el código y dese ser referenciado en los entregables.

**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/exposici-n-proyecto-anal-tica-de-texto-de-bi-202420-2bbr4z5cpypb3bis>

Deben entregar todo el software relacionado con el proyecto para poder ejecutarlo durante la sustentación o revisión del proyecto.