	PRUEBA DE CONOCIMIENTOS DATOS NO ESTRUCTURADOS II		
	VICEPRESIDENCIA BANCA PERSONAL Y MERCADEO ANALÍTICA - DEPTO. DATOS NO ESTRUCTURADO		
	Depto. Datos No Estructurados	Elaborado 11/01/2022	Versión:1

Prueba de Conocimiento Datos No Estructurados

Entrega: juan.pena@davivienda.com

En el Departamento de Datos No Estructurados buscamos personas con excelentes capacidades técnicas a las que les guste explorar los últimos avances en IA para asumir retos de especial dificultad. Aunque la siguiente prueba busca, en primera instancia, corroborar su idoneidad técnica, para nosotros es muy importante que esta destreza venga acompañada de dos elementos adicionales:

1. Creatividad para encarar desafíos técnicos en procesamiento de datos no estructurados.
2. Habilidad para comunicar el trabajo hecho y sus resultados a un público general o experto.


Por lo tanto, el objetivo de la prueba es tener una visión completa de cómo es su compromiso, capacidad analítica, creatividad técnica y comunicación en la solución de un problema puntual de datos no estructurados, lo cual, en la práctica, se verá reflejado en una presentación a los jurados.

Para el desarrollo de esta prueba cuentan con una semana, tiempo en el que deben desarrollar la solución del problema mediante código debidamente documentado, depositado en un cuaderno de Python, así como una presentación gerencial de su trabajo. De parte de los jurados se evaluarán estos dos elementos y se comprometen a respetar la propiedad intelectual de los desarrollos hechos por ustedes donde haya lugar (modelos personalizados, APIs propias, técnicas especiales, etc). Si lo desea, puede hacerlo explícito mediante licencia Creative Commons.

Puede hacer los supuestos que considere necesarios. Esto depende de la forma en que usted aborde el problema. Aunque los cuadernos de Python suministrados sugieren caminos, no hay una solución única. Fundamentalmente queremos ver soluciones ordenadas, creativas (si consideran que esta cualidad es pertinente en la resolución del problema) y bien comunicadas.

Introducción

Según Gartner, en este momento, los datos no estructurados conforman el 80-90% del total de los datos que manejan las compañías. Dentro de ellos, la información que está contenida en texto desestructurado (correos, decisiones judiciales, evaluaciones de clientes, comentarios en redes sociales, etc) representa un enorme potencial para desarrollar analítica dentro de las empresas.


	PRUEBA DE CONOCIMIENTOS DATOS NO ESTRUCTURADOS II		
	VICEPRESIDENCIA BANCA PERSONAL Y MERCADEO ANALÍTICA - DEPTO. DATOS NO ESTRUCTURADO		
	Depto. Datos No Estructurados	Elaborado 11/01/2022	Versión:1

Un objetivo evidente de procesamiento de texto no estructurado se encuentra en la analítica de opiniones. Mediante ella, podemos obtener información de primera mano de lo que piensan nuestros clientes, sin que dichas opiniones nos lleguen a través de encuestas u otros elementos logísticos que conformaban en un pasado reciente la única forma de tener información estructurada acerca de lo que opinan.

El objetivo de esta prueba es desarrollar una analítica preliminar sobre un conjunto de tweets en los que se menciona de una u otra forma a nuestra organización. Mediante herramientas de procesamiento de lenguaje natural (NLP) buscamos que desarrollen el procesamiento de los textos contenidos en los tweets, que encuentren insights que puedan ser de interés para áreas como marketing, servicio al cliente, etc; y que visualicen estos datos de una forma clara para tomadores de decisiones. Además, en la medida de lo posible, buscamos una agrupación de estos tweets, en las categorías que ustedes consideren las mejores para hacer una buena analítica descriptiva de lo que opinan nuestros clientes en Twitter.

Instrucciones Importantes

- Se suministrará el conjunto de tweets a analizar.
- Los resultados y visualizaciones producidas por su desarrollo deben estar sustentados en la presentación.
- **Tiene que mostrar conocimiento de uso de modelos LLM o Trasformer**
- El desarrollo es libre, siempre y cuando quede contenido en el cuaderno suministrado. Siéntase autónomo en usar la mejor técnica que considere para resolver el reto. Puede utilizar cualquier herramienta, propia o que encuentre en Internet.
- El cuaderno debe tener las líneas de comando que instalen las librerías y versiones utilizadas para correr el desarrollo.
- El desarrollo hecho debe estar documentado, tanto en el código como en comentarios que expliquen los pasos hechos en el cuaderno suministrado.
- Puede utilizar otros conjuntos de datos en su desarrollo, siempre que los resultados del análisis se limiten al conjunto de tweets proporcionados.
- En la presentación debe justificar las razones por las que escogió el enfoque desarrollado.
- Si ejecuta algún proceso previo o auxiliar en la solución del reto, por favor justifíquelo en la presentación.
- Si por algún motivo su desarrollo no tiene los resultados esperados por usted, recuerde que esto también es un resultado. Elabore la presentación y justifique su proceso. ¡Es posible que su trabajo sea **muy interesante y tenga mucho potencial!**

	PRUEBA DE CONOCIMIENTOS DATOS NO ESTRUCTURADOS II		
	VICEPRESIDENCIA BANCA PERSONAL Y MERCADEO ANALÍTICA - DEPTO. DATOS NO ESTRUCTURADO		
	Depto. Datos No Estructurados	Elaborado 11/01/2022	Versión:1

Bono: Adicionalmente, si tiene conocimientos de despliegue, “apificación” o puesta en producción de su desarrollo, puede hacerlos explícitos en la presentación mediante un bosquejo sustentado de lo que usted haría para que su desarrollo sea utilizado por un usuario final. Esto reemplazará el peor puntaje parcial que haya obtenido en la calificación de su presentación. El despliegue propuesto no se tiene que desarrollar. (Aunque si lo hace será tenido en cuenta)

Objetivo

El propósito de la prueba es medir su capacidad de análisis, programación, elección de técnicas y comunicación de procesos y resultados. Dentro del Departamento de Datos No Estructurados creemos que la unión de estos elementos determina la buena calidad de las soluciones que le aportamos a la organización.

Nos interesa conocer de su propia voz el proceso que hizo para resolver el reto, ver la forma como abordó el problema, el orden en su código y la manera como sintetiza las partes interesantes de su trabajo en una presentación.

Insumos


Para el desarrollo se compartirá una carpeta **‘Prueba_de_conocimientos_davivienda’** con la subcarpeta de este ejercicio **‘Ejercicio_2_chats’** con el siguiente contenido:

1. Carpeta de datos ‘davivienda_tweets.csv’ con los tweets a analizar.
2. Cuaderno de Google Colab ‘Prueba_de_conocimientos_II_DNE_Davivienda’.
3. Documento Explicativo del Ejercicio.

Entregables

Se debe entregar en archivo zip, marcado con su nombre, una carpeta que contenga los siguientes elementos:

- **Presentación Gerencial:** Una presentación (en formato PDF) donde se encuentre:
 - Exploración de datos.
 - Técnicas usadas o exploradas.
 - Dificultades encontradas.
 - Aspectos importantes.
 - Resultados.
 - Costos de la ejecución de su desarrollo, si corresponde (Ej: si utilizó una API paga, una VM con hardware especial, etc).

	PRUEBA DE CONOCIMIENTOS DATOS NO ESTRUCTURADOS II		
	VICEPRESIDENCIA BANCA PERSONAL Y MERCADEO ANALÍTICA - DEPTO. DATOS NO ESTRUCTURADO		
	Depto. Datos No Estructurados	Elaborado 11/01/2022	Versión:1

- Detalles adicionales que considere relevantes para la presentación de su proceso.
- (Opcional) - Bosquejo de puesta en producción del desarrollo hecho.
- **Códigos:** [Repositorio de Github](#) con el cuaderno de Python suministrado con el desarrollo hecho, debidamente documentado. Los resultados de la ejecución de cada celda deben ser visibles sin necesidad de correr las celdas. Tenga en cuenta que no vamos a modificar rutas de ubicación de input/output, archivos auxiliares, etc. por lo que todo el contenido debe ser completo y consistente dentro del cuaderno entregado.

NOTA: Verificar que el repositorio sea público para su posterior consulta.

- **(Opcional)** - Si su desarrollo necesita de archivos auxiliares para funcionar, por favor inclúyalos dentro del repositorio de manera que sea consistente con la ejecución de los códigos hechos.

Nota: Los resultados deben entregarse por medio de correo electrónico a juan.pena@davivienda.com

- Con el siguiente asunto: Prueba de Conocimiento Datos no Estructurados [nombre participante]
- Con la siguiente nomenclatura: datos_no_estructurados_02_NombreParticipante.zip

Evaluación

Se evaluarán sus capacidades de análisis, manejo de datos, programación e interpretación. Esto se realizará bajo la siguiente metodología:

- Desarrollo de los códigos: 30%
 - Orden (5%)
 - Documentación (5%)
 - Técnica (20%)
- Presentación con los resultados: 70%
 - Orden en el relato del proceso (20%)
 - Conocimientos técnicos (20%)
 - Completitud en los datos relevantes de su desarrollo (10%)
 - Claridad en la presentación (10%)
 - Bono: Si demuestra suficiencia de conocimientos de despliegue de la analítica presentada, obtendrá un buen puntaje en el bono y éste reemplazará la peor calificación que tenga en los cuatro criterios anteriores.