**ANEXO 1. FOMULARIO DE POSTULACIÓN DE RETOS**

Fecha: 20/05/2020

Ciudad: \_\_\_CALI\_\_\_\_\_\_\_\_\_\_

Señores

Ministerio MinTIC

Atendiendo a la convocatoria realizada por el Ministerio de Tecnologías de la Información y las Comunicaciones, a continuación, nos permitimos proponer un reto para el proceso de formación que se realizará en Ciencia de Datos durante el 2020, el cual atiende a una necesidad particular de la entidad.

Hemos leído los términos y condiciones de la convocatoria y disponemos de la información necearía para que el equipo pueda desarrollar el proyecto.

|  |  |
| --- | --- |
| **DATOS BÁSICOS** | |
| * + - 1. Nombre de la entidad o empresa: | **DATA360** |
| 1. Tipo de empresa o entidad: | Empresa privada: \_\_\_X\_\_  Sector: \_\_\_\_Tecnología y Analítica Avanzada  Entidad pública: \_\_\_\_\_\_  Nacional \_\_\_\_\_\_  Territorial \_\_\_\_\_\_  Otra - Cual? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| 1. Área/ departamento o Dirección de la empresa o entidad a la que está asociada el reto. | Valle del cauca/ Zona América vía Cali - Jamundí |
| 1. Punto de contacto operativo para el Proyecto **(Líder)** | |
| Nombres y apellidos | Natalia Raffo |
| Cargo | Chief Data Officer |
| Correo electrónico | natalia.raffo@data360.com.co |
| 1. Un punto de **contacto técnico** para el Proyecto que puede responder preguntas técnicas sobre el conjunto de datos y el problema | |
| Nombres y apellidos | Laura Orozco |
| Cargo | Científica de datos Senior |
| Correo electrónico | laura.orozco@data360.com.co |
| **DESCRIPCIÓN DEL PROYECTO** | |
| **Descripción general de la empresa o entidad pública.** | **DATA360º** es una empresa focalizada en servicios de analítica avanzada e inteligencia artificial, utilizando las tecnologías y herramientas más robustas de Inteligencia Artificial y analítica avanzada otorgamos a nuestros clientes la capacidad de tomar decisiones más asertivas, oportunas y relevantes.  Generamos valor a través de 3 pilares que nos diferencian de nuestros competidores y crean nuestra cultura empresarial. |
| **Descripción del problema que debe resolverse** | Herramienta automática de extracción de propiedades de un documento pre- impreso y manuscrito: Esta herramienta debe incluir la segmentación de documentos según el tipo, la extracción de propiedades y minería de texto. |
| **¿Por qué es importante este problema?** | Actualmente hemos evidenciado en nuestras empresas Colombianos necesidades del negocio enfocadas a utilizar herramientas más robustas para extraer texto de altos volúmenes de documentos en tiempos cortos para tomar decisiones efectivas generando conocimiento y aporte en la región. |
| **¿Cuáles son los conjuntos de datos que los participantes pueden acceder para resolver este problema?** | Documentos en formato pdf, Word, imágenes en formato :tiff, png, jpg, jpeg, facturas. |
| **Enumere y describa el conjunto de datos específicos** | * + - 1. DATA SET DE DOCUMENTOS EN PDFS, 2. DATA SET DE IMÁGENES DE FACTURAS Y DOCUMENTOS. Descripción de variables: en el primer data set: nombre del docuemento, fecha, resumen, contenido, descripción, metodología, implementación, tesis. Para el data set 2: nombre del proveedor, marca, fecha, nombre del asesor, productos, precios, valor total, observaciones. |
| **¿Qué tipo de conjuntos de datos externos pueden ser de interés para el problema?** | Datos de las fuentes MNIST, NIST, IAM |
| **Resultados esperados del proyecto.** | Herramienta automática de extracción de propiedades de un documento pre- impreso y manuscrito: Esta herramienta debe incluir la segmentación de documentos según el tipo, la extracción de propiedades y minería de texto, permita extraer el texto sin importar la estructura del documento y procesar altos volúmenes de manera automática. |

**Nota: Con la presentación de esta propuesta de proyecto nos comprometemos a dar respeto a todos y cada uno de los términos previstos en la Ley de protección de datos descritas en la Ley 1581 de 2012 así como las normas que la complementen, adicionen o sustituyan.**