



ANEXO 1. FOMULARIO DE POSTULACIÓN DE RETOS

Fecha: 11/05/2020

Ciudad: Bucaramanga

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 necesaria para que el equipo pueda desarrollar el proyecto.

DATOS BÁSICOS	
1. Nombre de la entidad o empresa:	K-2 INGENIERIA S.A.S.
2. Tipo de empresa o entidad:	Empresa privada: <input checked="" type="checkbox"/> X Sector: <u>Actividades de arquitectura e ingeniería y otras actividades conexas de consultoría técnica</u> Entidad pública: _____ Nacional _____ Territorial _____ Otra - Cual? _____
3. Área/ departamento o Dirección de la empresa o entidad a la que está asociada el reto.	División de Ingeniería
4. Punto de contacto operativo para el Proyecto (Líder)	
Nombres y apellidos	Raúl Fernando Castro Suárez
Cargo	Gerente División Ingeniería
Correo electrónico	raul.castro@k2ingenieria.com
5. Un punto de contacto técnico para el Proyecto que puede responder preguntas técnicas sobre el conjunto de datos y el problema	
Nombres y apellidos	Jerson Fabian Bueno Guarín
Cargo	Coordinador Redes
Correo electrónico	j.bueno@k2ingenieria.com



DESCRIPCIÓN DEL PROYECTO	
Descripción general de la empresa o entidad pública¹.	<p>K2 Ingeniería, es una empresa colombiana, constituida en 1999, que integra la experiencia en diferentes áreas de ingeniería ambiental, civil, mecánica, electrónica y sistemas para ofrecer un amplio portafolio de servicios de consultoría, monitoreo e integración de tecnologías.</p> <p>K2 Ingeniería, se ha especializado en el área ambiental, sector en el que desarrolla proyectos sobresalientes a nivel nacional.</p> <p>En el área de consultoría se ejecutan proyectos en un marco de desarrollo sostenible donde se busca principalmente la solución de problemas que afectan al ambiente y el entorno. K2 Ingeniería año tras año ha tenido un crecimiento exponencial en conocimiento especializado, que le da un valor agregado al poder asesorar a los clientes y guiarlos al diagnóstico y solución de su problemática ambiental y social. Los ejes principales en el desarrollo de proyectos son Aire, Olores, Ruido y vibraciones, Agua, Suelo, Meteorología, Recurso fauna y flora.</p>
Descripción del problema que debe resolverse²	<p>Con el proyecto se busca aplicar analítica de datos a la información histórica de monitoreos medioambientales para generar indicadores que apoyen la gestión ambiental e inteligencia artificial para realizar predicciones futuras de las variables de interés que soporten la toma de decisiones de los actores involucrados en la problemática y guíen las acciones que benefician a las comunidades.</p> <p>Para lo anterior, se debe tener presente que el monitoreo de variables medioambientales en tiempo real (calidad del aire, meteorología, ruido, aguas, etc.) provenientes de estaciones automáticas³ o incluso las mediciones manuales⁴ suele generar un alto volumen de datos que normalmente es de difícil consulta y rara vez es utilizado más allá de lo estrictamente requerido (p.ej. generar reportes con gráficas de los datos o realizar comparaciones con las normas aplicables).</p>

¹ <https://www.k2ingenieria.com/>

² Referencias:

https://en.wikipedia.org/wiki/Data_analysis

https://en.wikipedia.org/wiki/Data_mining

https://en.wikipedia.org/wiki/Artificial_intelligence

https://en.wikipedia.org/wiki/Machine_learning

³ Con diversidad de frecuencias: hay variables captadas cada minuto, 10 minutos o cada hora.

⁴ Con resultados cada 24 horas (diarios) o muestreos puntuales.



	<p>Por lo anterior, las siguientes preguntas ante esta realidad serían válidas:</p> <ol style="list-style-type: none">1. ¿Qué indicadores, relaciones o conclusiones podemos obtener de los datos históricos y las diferentes variables recopiladas aplicando <i>data analysis / mining</i> que ayuden a los tomadores de decisiones a realizar una mejor gestión ambiental y reducir el impacto de las actividades monitoreadas?2. ¿Cómo predecimos o pronosticamos el comportamiento futuro (con una probabilidad aceptable) de las diferentes variables de interés medioambiental (p.ej. calidad del aire y meteorología) usando <i>artificial intelligence (AI) / machine learning</i> con base en los datos de monitoreo obtenidos históricamente por la empresa y la información de libre acceso en el país?3. ¿Cómo automatizamos el procesamiento y la validación de los datos históricos y la generación de reportes y gráficas con software libre como <i>Python</i> o <i>R</i> para prestar un servicio más ágil a los clientes y volver más competitiva a la empresa?
<p>¿Por qué es importante este problema?⁵</p>	<p>En 2015, la Organización de las Naciones Unidas (ONU) aprobó la Agenda 2030 sobre el Desarrollo Sostenible, una oportunidad para que los países y sus sociedades emprendan un nuevo camino con el que mejorar la vida de todos. La Agenda cuenta con 17 Objetivos de Desarrollo Sostenible (ODS), que incluyen entre otros la defensa del medio ambiente y el combate al cambio climático. La meta 3.9 del Objetivo 3, contempla para 2030, reducir sustancialmente el número de muertes y enfermedades producidas por productos químicos peligrosos y la contaminación del aire, el agua y el suelo. De la misma manera, la meta 11.6 del Objetivo 11, plantea de aquí a 2030, reducir el impacto ambiental negativo per cápita de las ciudades, incluso prestando especial</p>

⁵ Referencias:

<https://www.un.org/sustainabledevelopment/es/>

<https://www.un.org/sustainabledevelopment/es/health/>

<https://www.un.org/sustainabledevelopment/es/cities/>

<https://www.minambiente.gov.co/index.php/asuntos-ambientales-sectorial-y-urbana/gestion-del-aire/contaminacion-atmosferica>

<https://www.minambiente.gov.co/index.php/noticias/4404-gobierno-nacional-presenta-el-estado-de-la-calidad-del-aire-en-colombia-y-la-primera-estrategia-nacional-de-calidad-del-aire-del-pais#>

<https://www.minambiente.gov.co/index.php/estrategia-nacional-de-calidad-del-aire-enca>



	<p>atención a la calidad del aire y la gestión de los desechos municipales y de otro tipo.</p> <p>El Estado colombiano no ha sido ajeno a dicha realidad, y de acuerdo con los análisis realizados por el Ministerio de Ambiente y Desarrollo Sostenible (MADS) la contaminación atmosférica en Colombia es uno de los problemas ambientales de mayor preocupación para los colombianos por los impactos generados tanto en la salud como en el ambiente, además, es el tercer factor generador de costos sociales después de la contaminación del agua y de los desastres naturales.</p> <p>Por todo lo anterior, Gobierno Nacional ha generado la primera Estrategia Nacional de Calidad del Aire (ENCA) del país, dicha estrategia busca mejorar la calidad del aire con énfasis en la reducción de material particulado en áreas urbanas, garantizando la protección del ambiente y la salud de los colombianos. Se debe tener presente que la estrategia le apunta a la meta del Plan Nacional de Desarrollo 2018-2022, que busca que más estaciones de monitoreo de calidad del aire cumplan con el estándar más estricto de la Organización Mundial de la Salud (OMS).</p> <p>En el numeral 2.1 de la ENCA perteneciente al capítulo de diagnóstico y problemática, el MADS identificó que uno de los grandes desafíos de esta estrategia se centra en el fortalecimiento del conocimiento científico y técnico, y que es uno de los temas de interés es el fortalecimiento del análisis de la información y bases de datos. El objetivo específico II de la Estrategia se alinea a lo anterior y esto se ve plasmado en la línea de acción 5.2. OE-II que busca fortalecer el conocimiento técnico científico para avanzar en la gestión de la calidad del aire en el país y afirma que con el fin de fundamentar y orientar las acciones para mejorar la calidad del aire en el país es necesario fortalecer la investigación científica y tecnológica, y <u>contar con datos fiables, estandarizados, íntegros, y comparables</u> en todo el territorio.</p>
¿Cuáles son los conjuntos de datos que los participantes pueden acceder para resolver este problema?	<p>La empresa cuenta con más de 270 millones de registros de aproximadamente 650 estaciones ubicadas en cerca de 20 proyectos ubicados a largo de todo el país en zonas mineras o industriales o de infraestructura clave para el desarrollo de Colombia. Los datos son obtenidos en tiempo real con sensores</p>



	<p>automáticos o de manera manual con monitoreos realizados en campo.</p> <p>Debido al alto volumen de datos de estaciones automáticas la empresa tiene almacenada en 39 bases de datos no relacionales (MongoDB) los datos provenientes de los equipos de medición y cuenta con la misma cantidad de bases de datos relacionales (MySQL) para variables de interés de los proyectos. La información se podría consultar ya sea por copias de las bases de datos (p.ej. <i>dumps</i> .json / .bson) o exportar a archivos planos CSV o incluso generar los servicios web necesarios para tener acceso a los datos a trabajar.</p> <p>Adicional a lo anterior, se cuentan con información no estructurada como reportes y formatos de clientes o autoridades ambientales en los cuales se debe reportar la información captada en los diferentes proyectos. En general se pueden clasificar en hojas de cálculo e informes por temática o recurso ambiental (calidad del aire, ruido, aguas, etc.)</p>																																																		
Enumere y describa el conjunto de datos específicos	<p>Una muestra de las cantidades de datos para diferentes proyectos se muestra en el siguiente cuadro:</p> <table><tr><th>ID Proyecto</th><th>Cantidad de Datos</th></tr><tr><td>1</td><td>27,168,590</td></tr><tr><td>2</td><td>4,757,326</td></tr><tr><td>3</td><td>9,601,088</td></tr><tr><td>4</td><td>21,013,033</td></tr><tr><td>5</td><td>11,489,024</td></tr><tr><td>6</td><td>9,422,333</td></tr><tr><td>7</td><td>15,573,219</td></tr><tr><td>9</td><td>281,372</td></tr><tr><td>10</td><td>186,145</td></tr><tr><td>11</td><td>982,165</td></tr><tr><td>12</td><td>258,285</td></tr><tr><td>13</td><td>110,968</td></tr><tr><td>14</td><td>13,518</td></tr><tr><td>16</td><td>3,787,658</td></tr><tr><td>17</td><td>2,557,161</td></tr><tr><td>18</td><td>80,256,753</td></tr><tr><td>19</td><td>80,343,977</td></tr><tr><td>20</td><td>460,456</td></tr><tr><td>21</td><td>47,109</td></tr><tr><td>32</td><td>152,239</td></tr><tr><td>33</td><td>1,713,744</td></tr><tr><td>34</td><td>420,226</td></tr><tr><td>35</td><td>116,899</td></tr><tr><td>39</td><td>250,864</td></tr></table>	ID Proyecto	Cantidad de Datos	1	27,168,590	2	4,757,326	3	9,601,088	4	21,013,033	5	11,489,024	6	9,422,333	7	15,573,219	9	281,372	10	186,145	11	982,165	12	258,285	13	110,968	14	13,518	16	3,787,658	17	2,557,161	18	80,256,753	19	80,343,977	20	460,456	21	47,109	32	152,239	33	1,713,744	34	420,226	35	116,899	39	250,864
ID Proyecto	Cantidad de Datos																																																		
1	27,168,590																																																		
2	4,757,326																																																		
3	9,601,088																																																		
4	21,013,033																																																		
5	11,489,024																																																		
6	9,422,333																																																		
7	15,573,219																																																		
9	281,372																																																		
10	186,145																																																		
11	982,165																																																		
12	258,285																																																		
13	110,968																																																		
14	13,518																																																		
16	3,787,658																																																		
17	2,557,161																																																		
18	80,256,753																																																		
19	80,343,977																																																		
20	460,456																																																		
21	47,109																																																		
32	152,239																																																		
33	1,713,744																																																		
34	420,226																																																		
35	116,899																																																		
39	250,864																																																		



Para una estación con información de calidad del aire y meteorología, el conjunto de datos está estructurado de la siguiente manera:

```
1 Estación;Serial;Variable;Unidad;Componente;Fecha;Valor
2 "Estación Aire";SN011;"Presión Atmosférica";mmHg;Meteorología;"2020-03-01 00:00:39";627.7150268547
3 "Estación Aire";SN011;"Temperatura Ambiente";Celsius;Meteorología;"2020-03-01 00:00:39";23.666700363159
4 "Estación Aire";SN011;"Humedad Relativa";%;Meteorología;"2020-03-01 00:00:40";74
5 "Estación Aire";SN011;"Precipitación";mm;Meteorología;"2020-03-01 00:00:41";0
6 "Estación Aire";SN011;"Radiación Solar";W/m²;Meteorología;"2020-03-01 00:00:42";0
7 "Estación Aire";SN011;"Velocidad del Viento";m/s;Meteorología;"2020-03-01 00:00:52";0.4470399916172
8 "Estación Aire";SN011;"Dirección del viento";°;Meteorología;"2020-03-01 00:00:53";303
9 "Estación Aire";SN016;PM10;µg/m³;Aire;"2020-03-01 00:01:00";47.430000305176
10 "Estación Aire";SN016;"INDICE CALIDAD AIRE PM10";µg/m³;Aire;"2020-03-01 00:01:00";47.430000305176
11 "Estación Aire";SN011;"Velocidad del Viento";m/s;Meteorología;"2020-03-01 00:01:52";1.3411200046539
12 "Estación Aire";SN011;"Dirección del viento";°;Meteorología;"2020-03-01 00:01:53";301
13 "Estación Aire";SN011;"Velocidad del Viento";m/s;Meteorología;"2020-03-01 00:02:52";1.7881599664688
14 "Estación Aire";SN011;"Dirección del viento";°;Meteorología;"2020-03-01 00:02:53";268
15 "Estación Aire";SN011;"Velocidad del Viento";m/s;Meteorología;"2020-03-01 00:03:52";1.7881599664688
```

Ejemplo estructura archivo plano CSV

Estación	Serial	Variable	Unidad	Componente	Fecha	Valor
"Estación Aire"	SN011	Presión Atmosférica	mmHg	Meteorología	2020-03-01 00:0	627.72
"Estación Aire"	SN011	Temperatura Ambiente	Celsius	Meteorología	2020-03-01 00:0	23.67
"Estación Aire"	SN011	Humedad Relativa	%	Meteorología	2020-03-01 00:0	74
"Estación Aire"	SN011	Precipitación	mm	Meteorología	2020-03-01 00:0	0
"Estación Aire"	SN011	Radiación Solar	W/m²	Meteorología	2020-03-01 00:0	0
"Estación Aire"	SN011	Velocidad del Viento	m/s	Meteorología	2020-03-01 00:0	0.45
"Estación Aire"	SN011	Dirección del viento	°	Meteorología	2020-03-01 00:0	303
"Estación Aire"	SN016	PM10	µg/m³	Aire	2020-03-01 00:0	47.43
"Estación Aire"	SN016	INDICE CALIDAD AIRE PM10	µg/m³	Aire	2020-03-01 00:0	47.43
"Estación Aire"	SN011	Velocidad del Viento	m/s	Meteorología	2020-03-01 00:0	1.34
"Estación Aire"	SN011	Dirección del viento	°	Meteorología	2020-03-01 00:0	301
"Estación Aire"	SN011	Velocidad del Viento	m/s	Meteorología	2020-03-01 00:0	1.79
"Estación Aire"	SN011	Dirección del viento	°	Meteorología	2020-03-01 00:0	268
"Estación Aire"	SN011	Velocidad del Viento	m/s	Meteorología	2020-03-01 00:0	1.79

Ejemplo datos tabulados

Abc 2020-03_rev.csv	Abc 2020-03_rev.csv	Abc 2020-03_rev.csv	Abc 2020-03_rev.csv	Abc 2020-03_rev.csv	Abc 2020-03_rev.csv	# 2020-03_rev...
Estación	Serial	Variable	Unidad	Componente	Fecha	Valor
"Estación Aire"	SN011	"Dirección del viento"	°	Meteorología	3/13/2020 6:07:11 PM	253.000
"Estación Aire"	SN011	"Velocidad del Viento"	m/s	Meteorología	3/13/2020 6:07:13 PM	3.576
"Estación Aire"	SN011	"Dirección del viento"	°	Meteorología	3/13/2020 6:08:11 PM	311.000
"Estación Aire"	SN011	"Velocidad del Viento"	m/s	Meteorología	3/13/2020 6:08:13 PM	4.470
"Estación Aire"	SN011	"Intensidad de señal"	RSSI	Sistema	3/13/2020 6:08:32 PM	24.000
"Estación Aire"	SN011	"Dirección del viento"	°	Meteorología	3/13/2020 6:09:11 PM	275.000
"Estación Aire"	SN011	"Velocidad del Viento"	m/s	Meteorología	3/13/2020 6:09:13 PM	4.470
"Estación Aire"	SN016	"INDICE CALIDAD AIR...	µg/m³	Aire	3/13/2020 6:10:00 PM	6.090
"Estación Aire"	SN011	"Dirección del viento"	°	Meteorología	3/13/2020 6:10:11 PM	319.000
"Estación Aire"	SN011	"Velocidad del Viento"	m/s	Meteorología	3/13/2020 6:10:13 PM	4.023
"Estación Aire"	SN016	PM10	µg/m³	Aire	3/13/2020 6:10:43 PM	6.090
"Estación Aire"	SN011	Precipitación	mm	Meteorología	3/13/2020 6:10:47 PM	0.000
"Estación Aire"	SN011	"Presión Atmosférica"	mmHg	Meteorología	3/13/2020 6:10:48 PM	624.814
"Estación Aire"	SN011	"Temperatura Ambie...	Celsius	Meteorología	3/13/2020 6:10:48 PM	26.444
"Estación Aire"	SN011	"Humedad Relativa"	%	Meteorología	3/13/2020 6:10:49 PM	60.800

Ejemplo tipos de datos

Como se puede observar en las imágenes anteriores, los datos en general contienen variables categóricas y numéricas con los siguientes tipos de datos:

- Estación / nombre de la estación: texto (*string*).
- Serial del equipo /monitor: texto (*string*).



	<ul style="list-style-type: none">• Variable monitoreada⁶: texto (<i>string</i>).• Unidad de medición: texto (<i>string</i>).• Componente / temática medición: texto (<i>string</i>).• Fecha y hora de medición: fecha (<i>date / timestamp</i>).• Valor de la medición: número decimal (<i>float</i>)
¿Qué tipo de conjuntos de datos externos pueden ser de interés para el problema? ⁷	<p>En cuanto al acceso a la información en materia del estado de la calidad del aire, Colombia cuenta con 204 estaciones de monitoreo que conforman 26 sistemas de vigilancia de la calidad. A nivel nacional, existe el Subsistema de información sobre calidad del aire (SISAIRE), las ciudades de Bogotá y Medellín cuentan con aplicativos propios.</p> <p>El IDEAM cuenta con más de 8000 estaciones con información meteorológica, hidrológica e hidrometeorológica que puede ser usada para el proyecto.</p> <p>Dependiendo del tipo de estación si está es afectada por el tránsito vehicular se pueden consultar las base de datos de las autoridades del sector como MinTransporte, INVIAS, ANI o el RUNT.</p>
Resultados esperados del proyecto.	<ol style="list-style-type: none">1. Metodología de estandarización y validación de datos de calidad de aire.2. Caracterización de los conjuntos de datos analizados que incluya las relaciones entre variables y los patrones en las series de tiempo.3. Comparación y evaluación del desempeño de modelos o ensamblado de modelos usando <i>machine learning / deep learning / neural networks</i> para la predicción de calidad del aire o contaminación del aire.4. Algoritmo de predicción de datos futuros de calidad del aire (<i>prediction / forecasting / estimation</i>) en <i>Python</i>.

⁶ Depende del proyecto, la estación y el recurso monitoreado.

⁷ Referencias:

<http://aircolombia.sisaire.gov.co/portal/index.col>

<http://www.ideam.gov.co/>

<https://www.datos.gov.co/>

<https://www.mintransporte.gov.co/>

<https://www.invias.gov.co/>

<https://www.ani.gov.co/>

<https://www.runt.com.co/>



	5. <i>Notebooks</i> en <i>Jupyter</i> con automatización de la generación de gráficas de interés del conjunto de datos por temática medioambiental (calidad del aire, meteorología, ruido y calidad y cantidad de aguas).
--	---

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.

CONFIDENTIAL



El presente documento contiene información confidencial, por lo tanto, es de uso exclusivo para sus destinatarios.

Se prohíbe descomponer o desensamblar el documento (incluyendo sus textos e imágenes) y la distribución del mismo a terceros diferentes a MinTIC y/o los evaluadores de la Convocatoria Banco de Retos DS4S 2020-I sin autorización escrita de K2 INGENIERÍA SAS.

www.k2ingenieria.com
info@k2ingenieria.com
+57 (7) 635 2870
+57 313 218 1070

CONFIDENTIAL