**CONVOCATORIA ABIERTA**

**CONVOCATORIA PRUEBAS DE CONCEPTO ACADÉMICAS EN ANALÍTICA Y MACHINE LEARNING DEL CENTRO DE EXCELENCIA Y APROPIACIÓN EN BIG DATA Y DATA ANALYTICS, ALIANZA CAOBA 2022-2**

# PRESENTACIÓN

Alianza CAOBA, es el Centro de Excelencia y Apropiación en *Big Data y Analytics* de Colombia, comprometido con innovar y generar nuevo conocimiento en Analítica, *Big Data*, Aprendizaje automático e Inteligencia Artificial, que se traduzca en soluciones a problemas relevantes para el desarrollo y la competitividad del país.

El Centro está constituido por: Grupo Bancolombia, Grupo Nutresa, IBM de Colombia, SAS Institute Colombia, DELL, CDT CREATIC, y las Universidades ICESI, EAFIT, los Andes y la Pontificia Universidad Javeriana.

En esta oportunidad, la Alianza desea invitarlo a participar en el desarrollo de pruebas de concepto académicas que le permitan generar valor a su organización a través del desarrollo de proyectos de ciencia de datos.

El presente documento describe el objetivo de la convocatoria, las características de las pruebas de concepto los requisitos para participar, las fechas y proceso de postulación, y los criterios de selección. De forma general, la figura 1 describe los pasos de la presente convocatoria. Las fechas de cada paso se muestran en el cronograma.

Prueba de Concepto (PoC) se entiende, como el desarrollo de un modelo de aprendizaje automático (*machine learning*) que intenta responder una pregunta analítica o requerimiento analítico de la organización. Las organizaciones postulan una pregunta o requerimiento analítico, que junto con información adicional se denomina reto analítico, y si son seleccionadas ingresan al proceso denominado PoC.

**Figura 1. Pasos de la presente convocatoria**

Diagrama

Descripción generada automáticamente

# OBJETIVO DE LA CONVOCATORIA

Seleccionar hasta 20 retos analíticos, que cumplan con la rúbrica de calificación, que permitan la exploración, identificación, análisis, modelamiento[[1]](#footnote-2) y visualización de la información presentada en datos estructurados y no estructurados, para la ejecución de las Pruebas de Concepto académicas (PoC).

El objetivo principal de esta convocatoria es impulsar la cercanía entre los problemas de las organizaciones, relacionados con la ciencia de datos y la inteligencia artificial, con la academia. Es un ejercicio académico dado que los equipos de científicos de datos estarán conformados por estudiantes de maestría relacionadas con analítica e inteligencia artificial de las Universidades JAVERIANA, ANDES, EAFIT e ICESI.

# CARACTERÍSTICAS DE LAS PRUEBAS DE CONCEPTOS (RETOS ANALÍTICOS)

1. Las PoC no tienen ningún costo para las organizaciones. Buscan acercar las soluciones de Alianza CAOBA a las necesidades de las organizaciones que participen.
2. Los retos analíticos, deben ser problemas reales e identificados por la empresa postulante, que involucre el análisis de un conjunto de datos y que estén claramente definidos dentro del alcance de la ciencia de datos.
3. Los retos deben contar con una descripción de la necesidad, objetivo general, pregunta o requerimiento analítico, e inventario de fuentes de datos disponibles. Ver punto 8 formulario de inscripción.
4. El resultado principal será un modelo analítico en lenguaje *“R” o “Python”.* Un modelo analítico es una función matemática[[2]](#footnote-3) o estadística programado en un ambiente de desarrollo (*Integrated Development Environment - IDE)* que permite apoyar una decisión compleja o fenómeno social.
5. El código y visualizaciones que se obtendrán serán en objetos y lenguaje *“R” o “Python*”. En algunos casos esto puede variar para el uso de aplicativos propios de la organización o por sugerencia del equipo de trabajo, lo cual dependerá de la decisión y conocimiento del mismo equipo (estudiantes y rol técnico representante de la organización).
6. No se admiten retos que incluyan desarrollo de software o despliegue de la solución en aplicativos propios de la organización.
7. Los datos deben estar disponibles para su consulta e inicio de las pruebas de concepto antes del 30 de julio de 2022, en la carpeta que se creará en *OneDrive* de CAOBA si el reto es elegido para ser desarrollado. Los datos preferiblemente deben estar en los siguientes formatos: .CSV .XLS. JSON. Cualquier otro formato se revisará en la convocatoria.
8. Las organizaciones podrán postular como máximo 3 retos analíticos por semestre.
9. La asignación de los retos a las Universidades es realizada por Alianza CAOBA.
10. Se conformarán equipos entre 2 y 4 estudiantes de las maestrías en analítica de las Universidades JAVERIANA, ANDES, EAFIT e ICESI quienes trabajarán en un reto durante 15 semanas. Aplicarán técnicas de exploración de datos y algoritmos de aprendizaje automático que aprendieron durante la maestría.
11. Dado que es un ejercicio académico existe un grado de incertidumbre sobre los resultados de la prueba de concepto.
12. Al finalizar la prueba de concepto, la persona con el rol técnico representante de la organización deberá entregar una evaluación, la cual será el insumo principal para que el profesor defina el 40% de la nota de los estudiantes.
13. Los estudiantes también entregaran una evaluación cualitativa de la disposición de la organización. La cual será tenida en cuenta para futuras PoC.
14. Los referentes teóricos bajo los cuales está basada la convocatoria que definen y aclaran los supuestos de las PoC analíticas son: (Domingos, 2012) y (Kotu & Deshpande, 2018). El primero permite definir cuáles son los principales supuestos para un adecuado modelo de aprendizaje automático. El segundo define cuales son las técnicas y tareas de aprendizaje automático (capitulo 1).

# REQUISITOS PARA PARTICIPAR

1. Diligenciar el formulario de postulación de proyectos. [Academicas PoC | Alianzacaoba](https://www.alianzacaoba.co/organizaciones)
2. Designar, por escrito, a una persona que puede ser funcionario o contratista o empleado de su organización. Esta persona tendrá el rol de apoyo técnico y se le asignarán las siguientes responsabilidades:
   1. Participar en las reuniones semanales requeridas durante la ejecución del proyecto.
   2. Apoyar y facilitar el acceso a las personas, la información y los datos necesarios dentro de la organización para el abordaje y comprensión del reto analítico. (Ver modelo carta de asignación).
3. Firmar acuerdo de confidencialidad entre cada estudiante asignado al reto y la organización (Ver modelo de acuerdo). Las organizaciones tienen derecho a solicitar a los estudiantes que la prueba de concepto desarrollada no sea de libre consulta y sea totalmente confidencial. Por tiempos administrativos no se contempla la firma de acuerdos de confidencialidad entre las universidades y la organización.

# CRONOGRAMA

|  |  |
| --- | --- |
| ACTIVIDAD | FECHA INICIO |
| Apertura de la convocatoria | 1 junio 2022 |
| Resolución de inquietudes sobre la convocatoria. | 6 y 13 junio 4 pm (virtual) |
| Fecha límite postulación de retos | 20 junio 2022 |
| Revisión y selección de retos | 21 al 24 junio 2022 |
| Resultado de elección del reto vía correo electrónico | 28 al 30 junio 2022 |
| Trámite de acuerdos de confidencialidad | 1 al 30 julio 2022 |
| Carga de datos en repositorio definido por CAOBA | 1 al 30 julio 2022 |
| Inicio de ejecución de la prueba de concepto | 1 al 30 julio 2022 |
| Finalización de la prueba de concepto | 25 noviembre 2022 |

# PROCESO DE POSTULACIÓN

* Enviar el reto que se desea postular como prueba de concepto por medio del formulario
* Adjuntar la carta de designación del rol técnico.
* Los retos enviados después de la fecha indicada en el cronograma de la convocatoria y por medio distinto al establecido, no se tendrán en cuenta dentro del proceso de selección. Una vez inscritos no se podrán hacer modificaciones a los retos y formulados.

# SELECCIÓN DE RETOS

Se seleccionarán los retos que estén inscritos y cumplan con todos los requisitos definidos para formar parte del banco de pruebas de concepto. Su selección será determinada por el equipo técnico de Alianza CAOBA y los profesores de las universidades aliadas, quienes los evaluarán y seleccionarán teniendo en cuenta los siguientes criterios y rúbrica. En caso de empate se seleccionarán los retos por fecha de inscripción.

* Diligenciamiento del formulario en línea en su totalidad.
* Cercanía de la pregunta analítica o el objetivo general con tareas de aprendizaje automático.
* Descripción detallada de las fuentes de datos que utilizará el proyecto. Descripción del número de registros y variables.
* Disponibilidad del rol técnico (carta de designación).

Los retos seleccionados, serán notificados a las personas de contacto relacionadas en el formulario por medio de correo electrónico, con quienes se establecerá comunicación directa a fin de suscribir los acuerdos de confidencialidad organización-estudiantes y acceso a la carpeta para el cargue de datos.

## Rúbrica de calificación

|  |  |
| --- | --- |
| Criterio | Forma de evaluar |
| El reto se encuentra totalmente diligenciado. No hay preguntas sin diligenciar en el formulario. | **0=** existe uno o más campos vacíos del formulario  **1=** el formulario esta diligenciado en su totalidad |
| Los datos de la persona rol técnico se encuentran expresados en la carta de designación. | **0=** existe uno o más campos vacíos en los datos de la carta de designación  **1=** se encuentran todos los datos expresados de forma correcta en la carta de designación |
| Descripción detallada de las fuentes de datos. | **0=** alguno de los siguientes criterios no se evidencia en la descripción del formulario en línea: registros y variables, acceso a los datos, descripción de los datos, diccionario de datos.  **1=** cumple con todos los criterios establecidos: las fuentes de datos son descritas según número de registros y variables. Al revisar la descripción se entienden de forma clara y precisa los datos. Se puede entender que la organización tiene acceso y cuenta con los datos (incluyendo un diccionario de datos) |
| Cercanía de la pregunta analítica o el objetivo general con tareas de aprendizaje automático. | **0=** No se encuentra relación con ninguna tarea analítica.  **1=** Al leer la pregunta analítica o el objetivo analítico se puede inferir que se debe realizar alguna de las siguientes tareas de aprendizaje automático: regresión, clasificación, agrupamiento, segmentación, *clustering*, reglas de asociación, detección de anomalías, sistema de recomendación, aprendizaje profundo, minería de texto, minería de imágenes o selección de características (Kotu & Deshpande, 2018). |
| La organización presentó retos analíticos en versiones anteriores de la convocatoria y uno o más retos fueron beneficiados como PoC | **0=** Si obtuvo retos analíticos beneficiados  **1 =** No obtuvo retos analíticos beneficiados |

# BIBLIOGRAFÍA

1. Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, *55*(10), 78–87.
2. Kotu, V., & Deshpande, B. (2018). *Data science: Concepts and practice*. Morgan Kaufmann.

1. Aplicación de técnicas de aprendizaje automático y estadística avanzada. [↑](#footnote-ref-2)
2. Entiéndase como una fórmula matemática. [↑](#footnote-ref-3)