

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 2023-2

1. 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.

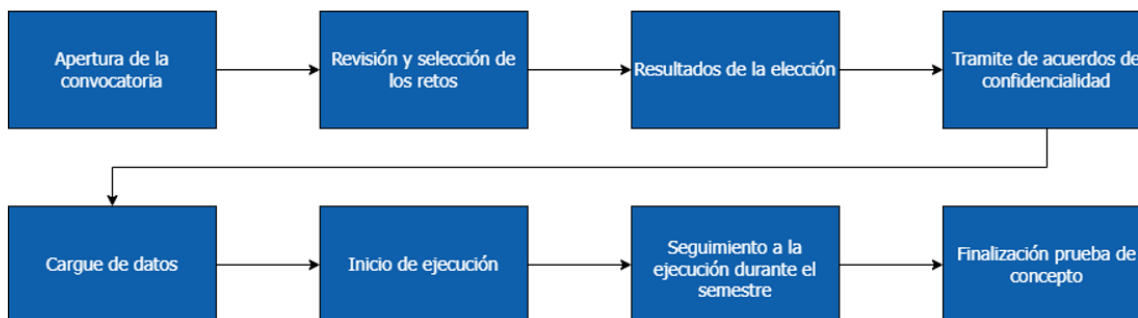
El Centro está constituido por: Grupo Bancolombia, Grupo Nutresa, IBM de Colombia, SAS Institute Colombia, DELL, CDT CREATIC, y las Universidades ICESI, EAFIT, los Andes y la Pontificia Universidad Javeriana.

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 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



2. Glosario

- **Autor:** Persona natural que realiza la creación intelectual o que ha contribuido, significativamente y con aporte intelectual, a la obtención de una obra. Se entiende como autor el desarrollador, diseñador, arquitecto, escritor o cualquier otra persona generadora de cualquier clase de obra protegida por la propiedad intelectual.
- **Backlog:** En una lista con todos los requerimientos iniciales del producto que se va a desarrollar.
- **Cesión de derechos patrimoniales:** Acto jurídico por el cual el autor, inventor o titular de los derechos patrimoniales de propiedad intelectual cede voluntariamente sus derechos a favor de la empresa de manera total o parcial, gratuita u onerosa. Para que la cesión de derechos patrimoniales tenga efectos, debe constar por escrito, en documento privado que, para efectos de las condiciones de publicidad y oponibilidad frente a terceros, deberá ser inscrito en el Registro de Actos y Contratos de la Dirección Nacional de Derechos de Autor. De acuerdo con lo establecido en la Ley 23 de 1982, “será inexistente toda estipulación en virtud de la cual el autor transfiera de modo general o indeterminable la producción futura, o se obligue a restringir su producción intelectual o a no producir”. La cesión de los derechos queda limitada a las modalidades de explotación previstas y al tiempo y ámbito territorial determinados contractualmente. La falta de mención del tiempo limita la transferencia a cinco años, y la del ámbito territorial, al país en el que se realice la transferencia.
- **DataOps:** Es la unión de Data Scientist y el equipo de operaciones.
- **Derechos de autor:** Se entiende por derecho de autor aquella prerrogativa de orden moral y patrimonial que el Estado otorga a todo creador de obras artísticas o científicas, incluido el soporte lógico (software): literarias, musicales, audiovisuales, de multimedia, de ingeniería civil, de artes escénicas, entre otras, y bases de datos desde el momento mismo de la creación, sin necesidad de registro, depósito o formalidad alguna, siempre y cuando la obra esté dotada de originalidad (esfuerzo intelectual), exteriorizada mediante alguna manifestación (letras, música...) y fijada en algún soporte material que permita la obtención de copias o la comunicación pública, sin que importe para ello su mérito, calidad o destinación. No son susceptibles de protección las ideas, ni su contenido conceptual, ideológico o técnico, de las obras científicas; tampoco su aprovechamiento industrial o comercial. Para que la protección de una obra se haga efectiva, no es necesario el registro en la Dirección Nacional de Derechos de Autor; sin embargo, te lo recomendamos como una buena práctica que garantiza mayor seguridad jurídica, otorga oponibilidad de los actos frente a terceros y sirve como medio probatorio de los actos sometidos a registro.
- **Derechos morales:** Son el conjunto de facultades y prerrogativas reconocidas en el artículo 30 de la Ley 23 de 1982, atribuidas al autor de la obra por la sola creación de esta, sin necesidad de formalidad o registro. Son derechos ilimitados en el tiempo y, por su carácter personalísimo, irrenunciables, inalienables, imprescriptibles e inembargables. Son derechos morales a) el de paternidad de la obra, b) el de integridad de la obra, c) el de ineditud o inédito, d) el de modificación de la obra, y e) el de arrepentimiento o retracto.
- **Derecho moral de paternidad:** Consiste en el derecho del autor a reivindicar la paternidad y exigir que su nombre y el título de la obra sean mencionados cada vez que esta sea utilizada, publicada o divulgada. Se presumirá que el autor es la persona cuyo nombre, seudónimo, iniciales u otro signo que lo identifique aparece indicado en la obra, en sus reproducciones, en sus divulgaciones, en sus representaciones o en cualquier forma de comunicación (artículo 30, literal a, Ley 23 de 1982).
- **Derecho moral de integridad:** Consiste en el derecho del autor a oponerse a toda deformación, mutilación o modificación que atente contra el decoro de la obra o contra la reputación del autor (artículo 30, literal b, Ley 23 de 1982).
- **Derecho moral de inédito:** Consiste en el derecho del autor de la obra a optar por publicarla o por dejarla inédita (artículo 30, literal c, Ley 23 de 1982).
- **Derecho moral de modificación:** Consiste en el derecho del autor a modificar su obra en cualquier tiempo, previo el pago de las indemnizaciones a que haya lugar a favor de terceros que pudieran verse afectados (artículo 30, literal d, Ley 23 de 1982).
- **Derecho moral de arrepentimiento o retracto:** Consiste en el derecho del autor a sacar de circulación su obra en cualquier tiempo, previo el pago de las indemnizaciones a que haya lugar a favor de terceros que pudieran verse afectados (artículo 30, literal e, Ley 23 de 1982).

- **Derechos patrimoniales:** Son el conjunto de facultades y prerrogativas atribuidas al autor de la obra de manera originaria por la sola creación de esta, sin necesidad de formalidad o registro, y que le permiten al autor o titular ejercer actos de explotación económica de su obra, o bien, autorizarlos a favor de terceros (artículo 12 de la Ley 23 de 1982). Pueden ser ejercidos directamente por el autor o titular, o por terceros facultados para ello. Admiten cesión, enajenación, transferencia o cualquier otro acto de disposición en todo o en parte. Pueden ser transferidos por causa de muerte, renunciarse y embargarse, son limitados en el tiempo, prescriptibles y expropiables. Son derechos patrimoniales a) el de reproducción, b) el de distribución de ejemplares, c) el de transformación y d) el de comunicación pública. Las facultades patrimoniales pueden consistir en cualquier forma de utilización que admita la obra, y cada una de ellas puede ser explotada independientemente.
- **Derecho patrimonial de reproducción:** Consiste en el derecho del autor o del titular a realizar o autorizar la fijación de la obra en cualquier medio que permita su divulgación y la obtención de copias (artículo 12 de la Ley 23 de 1982).
- **Derecho patrimonial de distribución:** Consiste en el derecho del autor o del titular a distribuir públicamente ejemplares o copias de la obra mediante venta o arrendamiento, y a autorizar tal conducta a terceros (artículo 12 de la Ley 23 de 1982).
- **Derecho patrimonial de transformación:** Consiste en el derecho del autor o del titular a realizar o autorizar la creación de obras derivadas como traducciones, adaptaciones, arreglos, compendios, extractos, colecciones (artículo 12 de la Ley 23 de 1982).
- **Derecho patrimonial de comunicación pública:** Consiste en el derecho del autor o del titular a realizar o autorizar la comunicación en público de la obra mediante ejecución, exposición, puesta a disposición, representación o transmisión, según sea el caso (artículo 12 de la Ley 23 de 1982).
- **Derecho de explotación exclusiva:** Son el conjunto de prerrogativas que el titular de los derechos adquiere por obtener la protección de los activos vía propiedad industrial. Los derechos de explotación exclusiva de las invenciones y modelos de utilidad son adquiridos con la obtención del título de patente de invención o del título de patente de modelo de utilidad, respectivamente; los de diseño o dibujo industrial con el registro de diseño industrial; los de las nuevas variedades vegetales con el certificado de obtentor vegetal; los de las marcas y lemas comerciales con el registro; los de las enseñas y nombres comerciales con su uso en el comercio, y los de las indicaciones geográficas con la declaración de protección. Los secretos industriales no requieren registro para su protección dada su naturaleza.
- **Desarrollo de modelos analíticos:** Es una representación matemática o estadística de un problema, con el objetivo de predecir algún evento, optimizar algún indicador o responder alguna pregunta de investigación. Los modelos analíticos tienen como resultado un valor, indicador o una respuesta al ser aplicados sobre nuevos datos. Se espera que este resultado se acerque lo mayor posible al resultado real.
- **Desarrollo de software:** El desarrollo de software se refiere a un conjunto de actividades informáticas dedicadas al proceso de creación, diseño, despliegue y compatibilidad de software.
- **DevOps:** Combinación de los términos ingleses Development (desarrollo) y Operations (operaciones). DevOps es un modo de abordar la cultura, la automatización y el diseño de las plataformas para generar mayor valor organizacional y capacidad de respuesta, mediante la prestación ágil de servicios de alta calidad. Todo ello es posible gracias a la prestación rápida y constante de servicios de TI. DevOps implica vincular las aplicaciones heredadas con las aplicaciones creadas en la nube y las infraestructuras más nuevas.
- **DevSecOps:** Esta es una forma de unión entre desarrolladores, seguridad y operación
- **Hardware:** Todos los componentes físicos de la computadora y sus periféricos.
- **Información:** Es el producto del procesamiento de datos que llega a ser útil para determinada persona o institución. En sentido general, la información es un conjunto organizado de datos procesados, que constituyen un mensaje que cambia el estado de conocimiento del sujeto o sistema que recibe dicho mensaje. Desde el punto de vista de la teoría general de sistemas cualquier señal o input capaz de cambiar el estado de un sistema se constituye en información.
- **Información confidencial:** Será secreto empresarial cualquier información no divulgada que una persona, natural o jurídica, legítimamente posea, que pueda usar en alguna actividad productiva, industrial o comercial, y que pueda transmitir a un tercero, en la medida en que dicha información sea: a. Secreta, en el sentido de que, como conjunto o en la configuración y reunión precisa de sus componentes, no sea generalmente conocida ni fácilmente accesible para quienes se encuentran en los círculos que normalmente manejan la información respectiva. b. Tenga un valor comercial por

ser secreta. c. Haya sido objeto de medidas razonables tomadas por su legítimo poseedor para mantenerla secreta. La información de un secreto empresarial podrá estar referida a la naturaleza, características o finalidades de los productos; a los métodos o procesos de producción, o a los medios de distribución-comercialización de productos o prestación de servicios (artículo 260 de la Decisión 486 de 2000).

- **Invencción:** Activo intangible protegido vía propiedad industrial como nueva creación. Consistente en una solución técnica a un problema técnico: artefactos, dispositivos, máquinas, utensilios, moléculas, composiciones químicas, etc; metodologías, procesos o procedimientos que solucionan el problema técnico. Para su protección son necesarios su registro y la obtención del derecho mediante el otorgamiento de la patente de invención en la Superintendencia de Industria y Comercio.
- **Inventor:** Persona natural que realiza una invención o nueva creación o que ha contribuido significativamente, y con aporte intelectual, a la obtención de una. En el inventor están radicados de manera originaria los derechos de explotación sobre la invención, pero puede cederlos a favor de terceros mediante un contrato de cesión de derechos.
- **Machine Learning (ML):** Es una disciplina del campo de la Inteligencia Artificial que, a través de algoritmos, dota a ordenadores de la capacidad de identificar patrones en datos masivos y elaborar predicciones. Este aprendizaje permite a los computadores realizar tareas específicas de forma autónoma, es decir, sin necesidad de ser programados.
- **MLOPS:** MLOps o Machine Learning Operations, es una extensión de la metodología DevOps que busca incluir los procesos de aprendizaje automático y ciencia de datos en la cadena de desarrollo y operaciones para hacer que el desarrollo del ML sea más confiable y productivo.
- **Modelo de utilidad:** Activo intangible protegido por la vía de la propiedad industrial como nueva creación. Se considera modelo de utilidad toda nueva forma, configuración o disposición de elementos de algún artefacto, herramienta, instrumento, mecanismo u otro objeto, o parte de este, que permita un mejor o diferente funcionamiento, utilización o fabricación del objeto que le incorpore o que le proporcione alguna utilidad, ventaja o efecto técnico que antes no tenía (artículo 81 de la Decisión 486 de 2000). Para su protección es necesario el registro ante la autoridad nacional competente, que para el caso de Colombia es la Superintendencia de Industria y Comercio.
- **NoOps:** Es un caso donde ya no se requiere que exista el equipo de operaciones. Aquí los equipos deben ser capaces de autoorganizarse para desplegar.
- **Patente de invención:** Título de propiedad que reconoce los derechos de propiedad industrial sobre las invenciones. Para su protección es necesario el registro ante la autoridad nacional competente, que para el caso de Colombia es la Superintendencia de Industria y Comercio. Para alcanzar la protección, la creación debe estar dotada de: Novedad, nivel inventivo y aplicación industrial.
- **Propiedad industrial:** Es el conjunto de derechos reconocidos al inventor de una invención o mejora; al diseñador de un dibujo o diseño industrial; al diseñador de un esquema de trazado de circuito integrado; al titular de un secreto empresarial; a los titulares de signos distintivos. Dentro de las formas en que se expresa la propiedad industrial, están: las patentes de invención y las patentes de modelo de utilidad, los diseños industriales, los esquemas de trazado de circuitos integrados, la información confidencial-secreto empresarial o industrial y los signos distintivos como marcas, lemas comerciales, nombres, enseñas comerciales e indicaciones geográficas. De la propiedad industrial surgen los derechos de reconocimiento a los creadores y de explotación.
- **Product Owner:** Profesional encargado de maximizar el valor de una solución generada en las empresas a partir de las metodologías ágiles.
- **Prueba de Concepto (PoC):** Se entiende, como el desarrollo de un modelo de aprendizaje automático (machine learning) que intenta responder una pregunta o requerimiento analíticos 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.
- **QAOps :** es la unión de QA y operaciones
- **Software:** Término general que designa los diversos tipos de programas, aplicaciones y sistemas de información usados en computación.
- **Scrum:** Es un framework que permite trabajar en una serie de interacciones en equipo.
- **Sistemas de Información:** Conjunto de programas de software que cumplen un objetivo específico dentro del negocio.

- **Tecnología:** es el conjunto ordenado de conocimientos y los correspondientes procesos que tienen como objetivo la producción de bienes y servicios, teniendo en cuenta la técnica, la ciencia y los aspectos económicos, sociales y culturales involucrados
- **Transferencia de tecnología o de conocimiento:** Mecanismo por medio del cual un conocimiento pasa de una persona natural o jurídica a otra. En materia de propiedad intelectual, la transferencia de conocimiento o de tecnología puede ser llevada a cabo a través de estudios, diseños, asesorías, consultorías, cesión o licencia de derechos de propiedad intelectual o industrial, creación de nuevas empresas o de servicios de asistencia técnica, entre otras.
- **Tecnologías de Información y Comunicaciones TIC:** Las tecnologías de la información y comunicación (TIC) son un término genérico que se utiliza actualmente para hacer referencia a una amplia gama de tecnologías y aplicaciones, que permiten la adquisición, producción, almacenamiento, tratamiento, registro, transferencia y presentación de información, que utilizan diversos tipos de aparatos, equipos, sistemas y programas informáticos, y que a menudo se transmiten a través de las redes de telecomunicaciones
- **Usuario:** Individuo que utiliza una computadora, sistema operativo, servicio o cualquier sistema informático. Por lo general es una única persona.

3. Objetivo General

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 pregrado y maestría relacionadas con analítica e inteligencia artificial de las Universidades JAVERIANA, ANDES, EAFIT e ICESI.

4. Objetivos Específicos

1. Seleccionar retos analíticos, que cumplan con la rúbrica de calificación, que permitan la exploración, identificación, análisis, modelamiento y visualización de la información presentada en datos estructurados y no estructurados, para la ejecución de las Pruebas de Concepto académicas (PoC).
2. Brindar un espacio de aplicación de la ciencia de datos que fortalezca las habilidades necesarias para abordar el desarrollo de proyectos en analítica que beneficien a organizaciones públicas y privadas.

5. Selección Prueba de Concepto

Se seleccionarán hasta 20 retos analíticos, que cumplan con la rúbrica de calificación, que permitan la exploración, identificación, análisis, modelamiento¹ y visualización de la información presentada en datos estructurados y no estructurados, para la ejecución de las Pruebas de Concepto académicas (PoC).

Con las PoC se busca entre otros:

- Realizar pruebas analíticas a escala en laboratorio, orientadas a demostrar la factibilidad técnica de los conceptos tecnológicos.
- Demostración teórica de un producto/proceso/concepto.
- Determinar si una idea puede hacerse realidad.
- Diseñar un esbozo de algoritmos de Machine Learning y/o Modelo Analítico.
- Probar si la idea es viable y explorar el potencial de la idea para ser desarrollada o construida a nivel de laboratorio.
- Identificar y estimar las características y medidas preliminares de desempeño del sistema.
- Realizar experimentos de laboratorio que verifiquen la viabilidad de la aplicación.

¹ Aplicación de técnicas de aprendizaje automático y estadística avanzada.

- Realizar experimentos con pequeños conjuntos de datos representativos
- Identificar los posibles problemas técnicos y logísticos que podrían interferir con el éxito del proyecto.
- Abordar cómo el producto o servicio propuesto apoyará las metas y objetivos de la organización u otros requisitos empresariales como objetivo secundario.
- Evaluar modelos analíticos y/o algoritmos existentes para su posible reutilización.
- Evaluar las limitaciones de los programas informáticos disponibles.

Características de las Pruebas de Conceptos (retos analíticos)

- a. Las PoC no tienen ningún costo para las organizaciones y/o empresas. Buscan acercar las soluciones de Alianza CAOBA a las necesidades de las organizaciones que participen; sin embargo, pueden existir convocatorias especiales que tendrán un costo según los acuerdos establecidos.
- b. 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.
- c. Los retos deben contar con una descripción de la necesidad, objetivo general, pregunta o requerimiento analítico, e inventario de fuentes de datos disponibles. Ver formulario de inscripción respectivo.
- d. Alianza CAOBA será la encargada de 1) convocar a las organizaciones para que postule retos y a las universidades para que postulen equipos de trabajo; 2) habilitar las herramientas que se utilizan en el desarrollo de los modelos y especificar las buenas prácticas que deben seguir los estudiantes (Protocolo de pruebas de concepto); 3) evaluar el desarrollo de las PoC para determinar cuáles pueden ir a la plataforma de analítica como servicio y alimentar el ecosistema con modelos.
- e. La revisión del modelo por parte de CAOBA y cargue del modelo en la plataforma de analítica de CAOBA como servicio no implica que se conocerán los datos de las diferentes empresas u organizaciones, sólo que el modelo recorre el PIPELINE que se ha creado para evaluar la calidad del desarrollo y hacer pruebas de despliegue en la plataforma.
- f. Las empresas, organizaciones y los estudiantes asignados se ven beneficiados, porque pueden tener una realimentación del resultado de la prueba de concepto tanto técnica como del despliegue.
- g. El hecho de subir el modelo a la plataforma representa un beneficio para las organizaciones y/o empresas, en el sentido de que podrán probar el modelo entregado en el marco de la prueba de concepto a través de la plataforma.
- h. Al tratarse de un ejercicio académico, se entiende que tanto la información como los modelos son de conocimiento únicamente de los estudiantes y profesor asignado y equipo técnico de las PoC de CAOBA.
- i. Si el cliente lo desea, el modelo puede estar en una capa 100% privada sin exposición pública, siempre y cuando esté dispuesto a suscribir el pago por el uso de la infraestructura que consuma el modelo, por lo que debe indicar si desea suscribir este servicio; de lo contrario podrá probar y utilizar el modelo sin costo en la capa gratuita que es pública.
- j. Tanto en la capa privada como en la capa gratuita solo se permitirá el uso del modelo, en ningún momento se expondrá para descargar el modelo o algún elemento asociado al mismo.
- k. El resultado principal será un modelo analítico en lenguaje “Python”. Un modelo analítico es una función matemática o estadística programado en un ambiente de desarrollo (Integrated Development Environment - IDE) que permite apoyar una decisión compleja o fenómeno social.
- l. El código del modelo y las visualizaciones que se obtendrán serán en lenguaje “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).
- m. No se admiten retos que incluyan desarrollo de software o despliegue de la solución en aplicativos propios de la organización.

- n. Los datos deben estar disponibles para su consulta e inicio de las pruebas de concepto según cronograma, 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.
- o. Las organizaciones podrán postular como máximo 3 retos analíticos por semestre.
- p. La asignación de los retos a las universidades es realizada por Alianza CAOBA.
- q. Se conformarán equipos entre 2 a 4 estudiantes de pregrado en ingeniería y/o 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 sus estudios de pregrado o maestría.
- r. Dado que es un ejercicio académico existe un grado de incertidumbre sobre los resultados de la prueba de concepto.
- s. 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 porcentaje de la nota de los estudiantes.
- t. Los modelos desarrollados en las PoC desarrollados en esta convocatoria serán cargados en la plataforma de analítica como servicio de la Alianza CAOBA; dichos datos no pueden ser utilizados por la Alianza para un ejercicio distinto a la prueba de concepto y pueden servir como insumo para el entrenamiento de herramientas.
- u. Al tratarse de un ejercicio académico, cuando los resultados de las PoC sean susceptibles de explotación se definirán entre las partes los criterios de propiedad intelectual con CAOBA, estudiantes, empresas y/o universidades según aplique.
- v. Los estudiantes firmarán acuerdos de confidencialidad con las empresas asignadas o que ellos mismos postulen y resulten aceptadas por la Convocatoria CAOBA según lo establecido en las condiciones de las PoC.
- w. La organización deberá entregar una carta de recomendación para cada uno de los estudiantes, destacando las habilidades y logros obtenidos en el desarrollo de la PoC.
- x. Los derechos morales de autor sobre las creaciones intelectuales desarrolladas por EL ESTUDIANTE, con ocasión de su participación, en el marco de la Prueba de Concepto y la investigación pertenecerán de manera perpetua, inalienable, inembargable e irrenunciable a EL ESTUDIANTE.
- y. A nivel de pruebas de concepto se tendrán en cuenta los derechos morales de la obra para los autores y titulares del modelo, respetando las reglas de divulgación y reconocimiento de los productos de Investigación en los resultados obtenidos.
- z. El modelo desarrollado será cargado en la plataforma de analítica como servicio de la Alianza CAOBA. Cuando se estructure un proyecto que sea susceptible de explotación se definirán entre las partes los criterios de propiedad intelectual.
- aa. Los datos recopilados en el repositorio de CAOBA serán tratados según la Política de Protección de Datos personales de la Pontificia Universidad Javeriana, que se puede consultar en el siguiente link: <https://www.javeriana.edu.co/documents/10179/9567882/Acuerdo+No+657.pdf>

EMPRESAS

1. Apertura de la convocatoria empresas

La convocatoria se realizará de acuerdo con el cronograma planeado y se espera de los participantes que cumplan los siguientes requisitos:

- 1. Enviar el reto que se desea postular por medio de la siguiente link: <https://www.javeriana.edu.co/encuestador/index.php?r=survey/index&sid=192837>
- 2. Designar, por escrito, a una persona que puede ser funcionario, contratista o empleado de la empresa. Esta persona tendrá el rol de apoyo técnico y se le asignarán las siguientes responsabilidades:

- Participar en las reuniones semanales y/o quincenales requeridas durante la ejecución del proyecto.
 - 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. Presentar los documentos de constitución (registro en Cámara de Comercio) en caso de ser empresas privadas; para entidades públicas presentar la Ley, acuerdo o norma con el cual se crea. En ambos casos se deben adjuntar los documentos de Representación Legal y documentos de identidad de quien ejerza esta condición para la entidad o empresa.
 4. 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.

Se seleccionarán los retos que permitan la exploración, identificación, análisis, modelamiento y visualización de la información presentada en datos estructurados y no estructurados, para la ejecución de las Pruebas de Concepto (PoC) que involucre el análisis de un conjunto de datos y que estén claramente definidos dentro del alcance de la Ciencia de Datos.

Los retos enviados después de la fecha y hora indicada en el cronograma de la convocatoria y por medio distinto al establecido, no se tendrán en cuenta dentro del proceso de selección del banco de retos. Una vez inscritos no se podrán hacer modificaciones a los retos y formularios.

Cronograma

Fechas importantes:	
ACTIVIDAD	FECHA INICIO
Apertura de la convocatoria	2 de mayo de 2023
Resolución de inquietudes sobre la convocatoria	23 de mayo de 2023 – 3:00 pm
Cierre de la inscripción	28 de julio de 2023
Asignación de retos	1 y 2 de agosto de 2023
Confirmación elección del reto vía correo electrónico	3 de agosto de 2023
Trámite de acuerdos de confidencialidad	3 y 4 de agosto de 2023
Carga de datos en repositorio definido por CAOBA	4 de agosto de 2023
Inicio de ejecución de la prueba de concepto	8 de agosto de 2023
Finalización de la prueba de concepto	24 de noviembre de 2023

Resolución dudas link: <https://www.alianzacaoba.co/organizaciones>

Fecha: 23 de mayo de 2023

Hora: 4:00 p.m.

2. Revisión y selección de los 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.

Se seleccionarán teniendo en cuenta los siguientes elementos:

- Diligenciamiento del formulario 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.
- Envío de los documentos del Representante Legal.
- 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
Entrega de la documentación del Representante Legal y constitución de la organización	0= No entrega la documentación del Representante Legal o Constitución de la Organización 1= Entrega la documentación del Representante Legal y Constitución de la Organización
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 o no se encuentra correctamente diligenciada 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 y variables. 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.	<p>0= No se encuentra relación con ninguna tarea analítica.</p> <p>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).</p>
---	---

3. Resultados de la elección

Los retos seleccionados, serán notificados a las personas de contacto relacionadas en el formulario, con quienes se establecerá comunicación directa o por correo según cronograma.

4. Tramite de acuerdos de confidencialidad

Con las personas que se establezca la comunicación de las entidades seleccionadas se tramitarán los acuerdos de confidencialidad y definirán, las políticas de derechos de autor que se apliquen en este caso particular. (Ver documento modelo)

5. Cargue de datos

Los retos seleccionados deberán realizar el cargue de datos que se utilizarán para el desarrollo de la prueba de concepto PoC según cronograma.

6. Ejecución de pruebas de concepto

- El objetivo principal de las Pruebas de Concepto es impulsar la cercanía entre los problemas de las entidades públicas y la empresa privada con la academia relacionada con la Ciencia de Datos e Inteligencia Artificial. Es un ejercicio académico dado que los equipos de científicos de datos estarán conformados por estudiantes de Pregrado en ingeniería o Maestría en Analítica e Inteligencia Artificial.
- No incluye despliegue, desarrollo de software o conexión entre aplicativos de sistemas de información de la organización.
- Un supuesto importante es que los estudiantes simulan ser consultores en analítica para adquirir dichas capacidades de consultoría.

Reto, Plan de Proyecto y su contenido

El proyecto presentado por las organizaciones y/o empresas debe tener como mínimo la siguiente información:

Título de Prueba de concepto o solución: El título abrirá el proyecto de prueba de concepto a realizar por tal motivo debe ser claro expresando la intención de la Prueba de Concepto, con palabras claves y dando precisión al marco en el que se va a desarrollar (espacio temporal).

Alcance: En este elemento se espera una precisa descripción del límite al que se llegará, inicio- fin que comprende, el problema a solucionar y los compromisos específicos que comprende la elaboración del proyecto de prueba de concepto.

Descripción de la necesidad y contexto: Debe responder, entre otras, a las siguientes preguntas: ¿Por qué se quiere realizar un proyecto de analítica, Big Data, ¿Machine Learning, o Inteligencia Artificial? ¿Qué problemáticas va a resolver? ¿Quiénes son los directamente beneficiados? ¿Qué información desea lograr para las decisiones de negocio? ¿Qué procesos están o estarían interesados en resolver la necesidad?

Objetivo general: El objetivo general debe ser uno solo, el cual representa “el fin último del proyecto”. No debe presentar detalles sobre la realización.

Inventario de fuentes disponibles: Listado de las fuentes de datos, descripción y características de la fuente. Detallar cualquier situación especial que pueda demorar o dificultar la disponibilidad de la información. (Permisos, trámites ante terceras entidades, anonimización de los datos, preparación previa, etc.) (Si hay necesidad, escriba cualquier comentario pertinente debajo de la tabla.)

Fuente	Descripción y características de la fuente
Nombre y descripción de la fuente de datos	Tipo de dato, formato, tamaño, origen, período que cubre, etc.

Acercamientos previos al problema: ¿El problema ya ha sido previamente afrontado por el cliente? ¿Qué aprendizaje se ha tenido? ¿Se puede tener acceso a dichos resultados o productos? Restricciones: Información adicional sobre cualquier posible restricción que se deba tener en cuenta para la ejecución del proyecto (datos, formatos, confidencialidad, tiempos, recursos, etc).
Inventario de modelos, algoritmos y metodologías preliminares: ¿Qué posibles modelos, algoritmos, metodologías o estrategias particulares pueden ser utilizadas para abordar alguna de las tareas identificadas en los objetivos generales y específicos de la problemática?

Modelo/Algoritmo/Metodología	Tarea con potencial de ser abordada

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.