

CONTRALORÍA GENERAL DE LA REPÚBLICA

PROYECTO “MEJORAMIENTO DE LOS SERVICIOS DE CONTROL GUBERNAMENTAL PARA UN CONTROL EFECTIVO, PREVENTIVO Y FACILITADOR DE LA GESTIÓN DE PÚBLICA” – BID3

	NOMBRE	CÓDIGO
NOMBRE COMPONENTE	ADECUADOS PROCESOS PARA UN CONTROL EFECTIVO Y EFICIENTE	1
PRODUCTO	SISTEMA INTEGRADO DE LOS SERVICIOS DE CONTROL DE LA CGR Y OCIS (SICG)	1.8
NOMBRE PROYECTO	“MÓDULO DE ANÁLISIS DE DATOS PARA DAR SOPORTE A LOS AUDITORES DURANTE LA EJECUCIÓN DE LOS SERVICIOS DE CONTROL” EN EL SISTEMA INTEGRADO DE CONTROL	1.8.2



Términos de Referencia

CONTRATACIÓN DE UN CONSULTOR CIENTÍFICO DE DATOS PARA EL DESARROLLO DEL “MÓDULO DE ANÁLISIS DE DATOS PARA DAR SOPORTE A LOS AUDITORES DURANTE LA EJECUCIÓN DE LOS SERVICIOS DE CONTROL” EN EL SISTEMA INTEGRADO DE CONTROL DEL PROYECTO INTERNO 1.8.2

1. Introducción
2. Antecedentes
3. Objetivo
4. Alcance de la Consultoría
5. Metodología de trabajo
6. Actividades a Realizar
7. Productos
8. Plazo de ejecución de la consultoría
9. Perfil del consultor
10. Monto, forma y condiciones de pago
11. Lugar de prestación del servicio
12. Coordinación, supervisión y conformidad
13. Accesos
14. Derechos de propiedad y confidencialidad de la información
15. Anticorrupción y Antisoborno
16. Penalidades
17. Anexo

MAYO 2026

CONTRATACIÓN DE UN CONSULTOR CIENTÍFICO DE DATOS PARA EL DESARROLLO DEL “MÓDULO DE ANÁLISIS DE DATOS PARA DAR SOPORTE A LOS AUDITORES DURANTE LA EJECUCIÓN DE LOS SERVICIOS DE CONTROL” EN EL SISTEMA INTEGRADO DE CONTROL DEL PROYECTO INTERNO 1.8.2

1. INTRODUCCIÓN

La Contraloría General de la República (CGR) es el órgano superior del Sistema Nacional de Control que cautela el uso eficiente, eficaz y económico de los recursos del Estado, la correcta gestión de la deuda pública, así como la legalidad de la ejecución del presupuesto del sector público y de los actos de las instituciones sujetas a control; coadyuvando al logro de los objetivos del Estado en el desarrollo nacional y bienestar de la sociedad peruana.

La Contraloría General de la República, en su rol fiscalizador y promotor de la buena gestión pública, requiere fortalecer sustancialmente sus capacidades analíticas para la detección proactiva de patrones irregulares y riesgos de corrupción en los procesos de contrataciones públicas. El objetivo primordial es la identificación temprana de proveedores favorecidos, la detección de posibles fraccionamientos indebidos del gasto público, y el descubrimiento de vínculos o redes de relación impropias entre proveedores y funcionarios.

En el marco del Proyecto de Inversión Pública “Mejoramiento de los Servicios de Control Gubernamental Para un Control Efectivo, Preventivo y Facilitador de la Gestión Pública”, se han elaborado los términos de referencia para la contratación del servicio de un consultor individual científico de datos para brindar soporte especializado en Big data durante el desarrollo del módulo de análisis de datos para dar soporte a los auditores durante la ejecución de los servicios de control en el sistema integrado de control del proyecto interno 1.8.2.

ÁREA USUARIA

El área usuaria del presente servicio es la Subgerencia de Sistemas de Información y Analítica de Datos.

2. ANTECEDENTES

El Proyecto de Inversión (CUI) N.º 2412703 “Mejoramiento de los Servicios de Control Gubernamental para un control Efectivo, Preventivo y Facilitador de la Gestión Pública-BID3” fue declarado viable por la Unidad Formuladora de la Contraloría General de la República (Gerencia Técnico Normativo) ahora Subgerencia de Formulación de Proyectos con fecha 28 de agosto de 2018 e inscrito en el Banco de Inversiones del MEF; el referido proyecto tiene como objetivo central brindar adecuados servicios de control gubernamental en el Estado”.

El plazo de ejecución del Proyecto es de 5 años, del 2019 al 2023, con financiamiento de 50% con recursos de endeudamiento con el Banco Interamericano de Desarrollo (BID) y 50% con recursos ordinarios, como contrapartida nacional, de acuerdo al Contrato de Préstamo (Nº 4724/OC-PE). El Proyecto contempla la ejecución de los siguientes componentes:

- Componente 1. Adecuados procesos para un control efectivo y eficiente.
- Componente 2. Adecuadas capacidades de los RRHH en temas de control gubernamental.
- Componente 3. Adecuado acceso a TIC en los procesos de control gubernamental.
- Componente 4. Adecuada capacidad operativa para la prestación de servicios de control desconcentrados.

En el marco del componente 1, se estableció el proyecto interno 1.8.2 "Sistema Integrado de Control".

Al respecto, dada la importancia del adecuado desarrollo de las actividades relacionadas al proyecto mencionadas líneas arriba, se hace necesaria la contratación del servicio de un consultor científico de datos, para el despliegue del proyecto Interno 1.8.2 Sistema Integrado de Control - Componente “Módulo de análisis de datos para dar soporte a los auditores durante la ejecución de los servicios de control”.

3. OBJETIVO

3.1. Objetivo General

Contratar los servicios de un(a) consultor(a) altamente especializado(a) en ciencia de datos, con una sólida experiencia en Machine Learning y procesamiento de grandes volúmenes de datos, para el desarrollo del componente “Módulo de análisis de datos para dar soporte a los auditores durante la ejecución de los servicios de control”, del proyecto interno 1.8.2, con la finalidad de disponer datos procesados y filtrados (Extracción, Transformación, limpieza y carga) de fuentes externas e internas, para la identificación de casos atípicos en el que se transgrede una normativa vigente. Esta información será de utilidad para la labor del auditor.

3.2. Objetivos Específicos

- a) **Desarrollar y Optimizar Modelos de Machine Learning:** Diseñar, entrenar y optimizar modelos de aprendizaje supervisado y no supervisado utilizando Apache Spark (Spark MLlib), asegurando la eficiencia, escalabilidad y precisión en el manejo de grandes volúmenes de datos para la detección de anomalías y patrones de riesgo.
- b) **Asegurar la Calidad y Relevancia de los Datos:** Preparar y refinar conjuntos de datos mediante procesos exhaustivos de limpieza, transformación y enriquecimiento de información, garantizando su alta calidad y relevancia para el entrenamiento y validación efectiva de los modelos.
- c) **Implementar Estrategias de Sostenibilidad del Modelo:** Incorporar mecanismos de autoevaluación y, cuando sea pertinente, de autoentrenamiento para permitir la actualización continua de los modelos mediante el análisis de nuevos datos, asegurando que los algoritmos se mantengan precisos, pertinentes y adaptados a la evolución de los patrones de contratación.
- d) **Realizar Análisis de Datos para Optimización de Modelado:** Estudiar y analizar a profundidad los datasets para identificar patrones, tendencias y relaciones ocultas, determinando el algoritmo y la estrategia de modelado más adecuados para cada caso de uso y optimizando el proceso de modelado predictivo y de detección de anomalías.
- e) **Garantizar la Robustez y Precisión de los Modelos:** Realizar evaluaciones rigurosas y pruebas exhaustivas de los modelos desarrollados, validando su robustez, precisión, capacidad de generalización y desempeño en diferentes escenarios antes de su despliegue final.
- f) **Asegurar el Despliegue Operacional del Modelo:** Integrar y desplegar los modelos de Machine Learning en entornos productivos de la CGR, asegurando la implementación de estándares de calidad, seguridad, mantenimiento y monitorización continua de su desempeño.
- g) **Proveer Documentación Técnica Completa y Validada:** Elaborar y mantener actualizado el diccionario y diagrama del modelo de datos, así como documentar de forma exhaustiva todo el proceso de desarrollo, los algoritmos utilizados y los resultados obtenidos, garantizando que los productos entregados cumplan con los estándares de calidad y los objetivos establecidos.

4. ALCANCE DE LA CONSULTORÍA

El servicio del consultor Científico de Datos estará centrado en el desarrollo integral del proyecto y comprenderá las siguientes áreas de trabajo:

4.1. Análisis y Preparación de Datos Avanzada:

- 4.1.1. **Análisis Exploratorio de Datos (EDA):** Realizar un análisis exploratorio profundo de los datos para comprender su estructura, distribuciones, relaciones iniciales y la identificación de anomalías.
- 4.1.2. **Identificación y Adquisición:** Evaluar y obtener las fuentes de datos relevantes del SIAF y SEACE, asegurando la completitud y pertinencia para los objetivos del proyecto

- 4.1.3. **Integración y Consolidación:** Diseñar e implementar procesos para integrar y consolidar grandes volúmenes de datos de múltiples tablas y sistemas (SIAF, SEACE) en un formato unificado para análisis.
- 4.1.4. **Limpieza y Transformación:** Aplicar técnicas avanzadas de limpieza de datos para manejar valores atípicos, inconsistencias, datos faltantes, duplicidades y errores de formato. Realizar transformaciones necesarias (ej. normalización, estandarización, codificación de variables categóricas) para optimizar los datos para el modelado.
- 4.1.5. **Enriquecimiento y Generación de Características:** Crear nuevas variables o características a partir de los datos existentes que sean relevantes para la detección de patrones de favoritismo y fraccionamiento.

4.2. Desarrollo de Modelos Analíticos y de Machine Learning (ML) Avanzados:

- 4.2.1. **Diseño y Selección de Modelos:** Diseñar y seleccionar los algoritmos de Machine Learning más apropiados para los problemas específicos:
- 4.2.2. **Identificación de Favoritismo:** Modelos de clasificación (ej. Regresión Logística, Random Forest, Gradient Boosting) o de puntuación de riesgo.
- 4.2.3. **Detección de Fraccionamiento:** Técnicas de agrupamiento (clustering) o detección de anomalías para identificar patrones de compras repetitivas o divididas.
- 4.2.4. **Evaluación de Vínculos:** Análisis de grafos o redes para mapear y evaluar relaciones entre proveedores y funcionarios (ej. por DNI, RUC, direcciones, teléfonos, etc.).
- 4.2.5. **Entrenamiento y Optimización:** Entrenar los modelos utilizando metodologías robustas, aplicar técnicas de validación cruzada y optimizar hiperparámetros para maximizar la precisión, recall, F1-score u otras métricas relevantes para la detección de fraude/riesgo.
- 4.2.6. **Despliegue y Pruebas:** Colaborar en las pruebas de rendimiento y robustez de los modelos en diferentes entornos (desarrollo, calidad) para su futura puesta en producción. Se enfatiza el uso de entornos de procesamiento distribuido como **Apache Spark (especialmente Spark MLlib)** para el entrenamiento y aplicación de modelos en grandes volúmenes de datos.

4.3. Generación de Reportes Técnicos Estructurados:

- 4.3.1. **Diseño y Desarrollo de Reportes:** Crear plantillas y scripts para la generación automática de reportes técnicos. Estos reportes deberán presentar los hallazgos de los modelos y análisis de manera clara, concisa y comprensible para usuarios técnicos y no técnicos de la CGR.
- 4.3.2. **Contenido de Reportes:** Los reportes incluirán tablas de resultados, estadísticas clave, gráficos estáticos (generados programáticamente), resúmenes de patrones detectados, listas de proveedores identificados con riesgo, métricas de desempeño de los modelos, y cualquier otra información relevante para la acción de control.

4.4. Documentación Técnica Completa y Código Fuente:

- 4.4.1. **Documentación de Modelos:** Elaborar una documentación técnica para cada modelo desarrollado, incluyendo la descripción del algoritmo, la lógica de negocio, las fuentes de datos utilizadas, las métricas de evaluación, las supuestas limitaciones y las recomendaciones para su mantenimiento y mejora.
- 4.4.2. **Diccionario y Diagrama de Datos:** Crear un diccionario de datos detallado para los datasets transformados en las capas de consumo, y un diagrama de modelo de datos que ilustre la estructura de las relaciones clave para el proyecto.
- 4.4.3. **Código Fuente Comentado:** Entregar todo el código fuente del proyecto (scripts de extracción, transformación, carga (ETL), código de

los modelos ML, scripts de reportes) debidamente comentado, organizado y versionado en los repositorios de la Contraloría.

- 4.4.4. Metodología y Hallazgos:** Documentar la metodología de ciencia de datos seguida, los principales hallazgos del análisis exploratorio, los patrones detectados y las recomendaciones derivadas del proyecto.

4.5. Transferencia de Conocimiento:

- 4.5.1. Realizar** sesiones de transferencia de conocimiento al personal interno para asegurar la sostenibilidad y escalabilidad del proyecto.

5. METODOLOGIA DE TRABAJO

La consultoría se desarrollará conforme a los lineamientos establecidos por la Gerencia de Tecnologías de la Información y Gobierno Digital y bajo supervisión de la Subgerencia de Sistemas de Información y Analítica de Datos, en coordinación con la Subgerencia de Gestión de Inversiones.

El trabajo se ajustará a la metodología institucional de la CGR para la implementación del proyecto, aplicando el Procedimiento "Desarrollo de Soluciones", Estándares de base de datos y otros lineamientos técnicos vigentes.

La Subgerencia de Sistemas de Información y Analítica de Datos establecerá los canales de coordinación y comunicación, y validará los entregables del consultor conforme a los estándares de calidad definidos para el proyecto.

6. ACTIVIDADES A REALIZAR

El consultor Científico de Datos ejecutará las siguientes actividades, las cuales no son limitativas:

- **Fase de Planificación y Requerimientos:**
 - Levantamiento Detallado de Requerimientos Funcionales y Técnicos: Participar en reuniones y talleres con los usuarios finales y el equipo técnico de la CGR para comprender a fondo los problemas de negocio, los procesos de contratación pública, las necesidades de información y las expectativas del producto a implementar.
 - Documentar los requerimientos funcionales y no funcionales (rendimiento, seguridad) del producto, así como los indicadores clave de éxito.
 - Identificación de Fuentes de Datos: Identificar y evaluar las fuentes de datos externas e internas que serán necesarias para el entrenamiento y validación de los modelos de machine learning.
- **Fase de Ingesta y Procesamiento de Datos (Lakehouse - Capas Plata y Oro):**
 - Carga de Fuentes al Lakehouse: Proceso ETL para la carga de las fuentes de datos externas e internas identificadas al Lakehouse de la CGR, el cual está implementado bajo el ecosistema de Apache Hadoop.
 - Realizar procesos de limpieza y transformación de los datos en la capa Plata y Oro de la plataforma de minería de datos de la CGR, asegurando la consistencia, completitud y adecuación para el análisis y modelado.
 - Orquestación de Procesos ETL (DAGs): Diseñar, desarrollar y probar los Directed Acyclic Graphs (DAGs) para orquestar los procesos ETL, siguiendo los estándares y buenas prácticas definidas por la Gerencia de Tecnologías de Información y Gobierno Digital de la CGR.
 - Ejecutar análisis exploratorios y estadísticos para detectar patrones, tendencias y casos atípicos, lo que facilitará la selección del algoritmo óptimo para el modelado
- **Fase de Análisis, y desarrollo de modelos de machine learning:**
 - Desarrollar, entrenar y validar modelos de machine learning utilizando Apache Spark (Spark MLlib), implementando técnicas supervisadas y no supervisadas,

- e incorporando mecanismos de autoevaluación para optimizar su desempeño con datos futuros.
- Análisis Profundo de Pagos y Modalidades de Contratación: Ejecutar un análisis estadístico y exploratorio de los patrones de pagos, montos contractuales y las diferentes modalidades de contratación (licitación pública, adjudicación simplificada, contratación directa, etc.) presentes en los datos.
- Desarrollo de Modelos para la Identificación de Proveedores Favoritos: Definir criterios técnicos y aplicar modelos (ej. basados en regresión, clasificación o reglas) para identificar proveedores que recurrentemente son contratados sin procesos competitivos claros o bajo condiciones que sugieren favoritismo. Calcular métricas de "riesgo de favoritismo".
- Implementación de Modelos para la Detección de Patrones de Compras Anómalas: Desarrollar algoritmos y técnicas (ej. series temporales, detección de anomalías, agrupamiento) para identificar patrones de fraccionamiento de compras, frecuencia inusual de contratación con ciertos proveedores, y contratación de objetos o servicios sospechosamente similares.
- Realizar pruebas de rendimiento y robustez de los modelos en entornos de desarrollo, calidad y producción, asegurando su correcta integración y funcionamiento en la solución final.
- **Fase de Certificación y puesta a producción:**
 - Elaborar documentación técnica necesaria para el proceso de certificación (modelos, variables consideradas en cada modelo, configuraciones, parámetros, procesos ETL, entre otros).
 - Levantar las observaciones presentadas en el proceso de certificación.
 - Colaborar en el despliegue y seguimiento en producción de los modelos, aplicando buenas prácticas y atendiendo incidencias durante la implementación.
- **Fase de Transferencia de Conocimiento:**
 - Realizar sesiones de transferencia de conocimiento al personal interno para asegurar la sostenibilidad y escalabilidad del proyecto

7. PRODUCTOS

El Consultor deberá generar y entregar los siguientes productos a través de mesa de partes virtual de la CGR (<https://mesadepartesvirtual.contraloria.gob.pe/mpvirtual/>) o de forma física, dirigido a la Subgerencia de Sistemas de Información y Analítica de Datos, con copia a la Subgerencia de Gestión de Inversiones o la que haga sus veces, en las fechas que corresponda.

N°	Productos	Contenido
1	Primer Producto	Plan de Trabajo, detallando productos específicos derivados del servicio solicitado y las actividades requeridas para alcanzarlos. Para ello, el consultor deberá presentar el Plan de Trabajo de las actividades definidas en el presente documento, según como se indica en el Anexo 01, literal II.1.
2	Segundo Producto	Documento técnico que comprenda la limpieza y preprocesamiento de los datos para la identificación de proveedores Favoritos: <ul style="list-style-type: none"> • Imputación de valores faltantes (mean, median, KNN, etc.). • Eliminación o tratamiento de outliers. • Codificación de variables categóricas (OneHotEncoder, LabelEncoder, TargetEncoding). • Normalización o estandarización de variables numéricas. • Crear nuevas variables a partir de las existentes (Feature engineering).

		<ul style="list-style-type: none"> • División de los datos
3	Tercer Producto	<p>Documento técnico que comprenda la selección del algoritmo y desarrollo del modelo de Machine Learning para la identificación de proveedores Favoritos:</p> <ul style="list-style-type: none"> • Documentación de patrones y tendencias identificadas en los datos. • Documento de análisis estadístico detallado de las variables clave. • Propuesta fundamentada de algoritmos de machine learning a implementar según el problema. • Documentación del diseño y arquitectura de los modelos de machine learning seleccionados. • Reporte de implementación de modelos usando Apache Spark MLlib • Documentación de los parámetros de entrenamiento y configuraciones utilizadas
4	Cuarto Producto	<p>Entrenamiento y validación del modelo de Machine Learning para la identificación de proveedores Favoritos:</p> <ul style="list-style-type: none"> • Informe de métricas de evaluación inicial de los modelos. • Informe con el registro de experimentos y pruebas realizadas. • Reporte de validación cruzada y evaluación exhaustiva de los modelos. • Ajuste de hiperparámetros. • Persistencia del modelo entrenado • Análisis comparativo del rendimiento en diferentes escenarios • Documentación técnica completa de los modelos desarrollados para el pase a certificación.
5	Quinto Producto	<p>Documento técnico que comprenda la limpieza y preprocesamiento de los datos para la detección de Fraccionamiento:</p> <ul style="list-style-type: none"> • Imputación de valores faltantes (mean, median, KNN, etc.). • Eliminación o tratamiento de outliers. • Codificación de variables categóricas (OneHotEncoder, LabelEncoder, TargetEncoding). • Normalización o estandarización de variables numéricas. • Crear nuevas variables a partir de las existentes (Feature engineering). • División de los datos
6	Sexto Producto	<p>Documento técnico que comprenda la selección del algoritmo y desarrollo del modelo de Machine Learning para la detección de Fraccionamiento:</p> <ul style="list-style-type: none"> • Documentación de patrones y tendencias identificadas en los datos. • Documento de análisis estadístico detallado de las variables clave. • Propuesta fundamentada de algoritmos de machine learning a implementar según el problema. • Documentación del diseño y arquitectura de los modelos de machine learning seleccionados. • Reporte de implementación de modelos usando Apache Spark MLlib • Documentación de los parámetros de entrenamiento y configuraciones utilizadas

7	Septimo Producto	<p>Entrenamiento y validación del modelo de Machine Learning para la detección de Fraccionamiento:</p> <ul style="list-style-type: none"> · Informe de métricas de evaluación inicial de los modelos. · Informe con el registro de experimentos y pruebas realizadas. · Reporte de validación cruzada y evaluación exhaustiva de los modelos. · Ajuste de hiperparámetros. · Persistencia del modelo entrenado · Análisis comparativo del rendimiento en diferentes escenarios · Documentación técnica completa de los modelos desarrollados para el pase a certificación. <p>Documento Técnico que comprenda:</p> <ul style="list-style-type: none"> · Documentación técnica actualizada de los modelos desarrollados. · Manual de implementación y despliegue de los modelos · Informe con el registro de pruebas de integración en los diferentes ambientes · Informe de validación del pipeline completo de datos · Plan de monitoreo y mantenimiento de los modelos · Levantamiento de incidencias en los ambientes de certificación. · Documentación técnica actualizada de los modelos desarrollados para el pase a certificación y/o producción. <p>Documento Técnico que comprenda:</p> <ul style="list-style-type: none"> · Informe de transferencia de conocimiento desarrollado a los usuarios técnicos y funcionales. · Levantamiento de incidencias en los ambientes de certificación y/o producción. · Repositorio organizado con código, datasets y reportes (ej.: GitLab) · Listado de mejoras identificadas en la marcha blanca.
---	-------------------------	--

8. PLAZO DE EJECUCIÓN DE LA CONSULTORÍA

El plazo total del servicio contratado será de ciento ochenta (180) días calendario y se contabilizará a partir del siguiente de la firma de su contrato.

En caso de que el día de entrega del producto corresponda a un día no laborable, el consultor presentará el informe correspondiente al día hábil siguiente.

Productos	Plazos
Primer Producto	Será presentado hasta los siete (7) días calendario a partir del día siguiente de la suscripción del contrato.
Segundo Producto	Será presentado hasta los treinta (30) días calendario a partir del día siguiente de la suscripción del contrato.
Tercer Producto	Será presentado hasta los sesenta (60) días calendario a partir del día siguiente de la suscripción del contrato.
Cuarto Producto	Será presentado hasta los noventa (90) días calendario a partir del día siguiente de la suscripción del contrato.
Quinto Producto	Será presentado hasta los ciento veinte (120) días calendario a partir del día siguiente de la suscripción del contrato.
Sexto Producto	Será presentado hasta los ciento cincuenta (150) días calendario a partir del día siguiente de la suscripción del contrato.

Séptimo Producto	Será presentado hasta los ciento ochenta (180) días calendario a partir del día siguiente de la suscripción del contrato.
------------------	---

En caso de que el día de entrega del producto corresponda a un día no laborable, El consultor presentará el informe correspondiente al día hábil siguiente.

Sin perjuicio, de lo antes señalado, y de presentarse situaciones justificadas que demanden un mayor plazo para la ejecución del trabajo a realizar por el Consultor, este podrá solicitar una ampliación de plazo a la SINFA, quien evaluará la pertinencia de dicha solicitud y procederá a dar o no su conformidad; sin que esto signifique el reconocimiento de mayores pagos a los establecidos en el presente documento.

En caso de existir observaciones a los productos la Entidad notificará al consultor dentro de un plazo máximo de cinco (05) días hábiles posteriores a la recepción de cada producto, para lo cual el consultor tendrá un plazo máximo de tres (03) días hábiles para levantar estas observaciones; dicho plazo comenzará a partir del primer día siguiente de notificada la comunicación.

Los plazos de revisión no son contabilizados dentro del plazo efectivo del servicio. Los requerimientos de revisión a las observaciones realizadas a los productos podrán ser solicitados hasta en dos (02) oportunidades por cada producto.

9. PERFIL DEL CONSULTOR

A. Formación Académica

- Profesional titulado en Ingeniería Sistemas o Ingeniería Informática o Ingeniero en Computación o Ingeniería Industrial o en Ciencias de la Computación o Matemático o Estadístico o afines u otras carreras afines a las Tecnologías de la Información o equivalente en el país origen.
- Deseable con estudios de posgrado concluidos o egresado en la especialidad de Ciencia de datos o Ciencias de la Computación o afines o equivalente en el país origen.

Se acreditará con copia del diploma respectivo de la formación académica.

B. Experiencia General

Experiencia mínima de ocho (8) años en el ejercicio de su profesión, en el sector privado o en el sector público.

La experiencia será contabilizada a partir de la condición de BACHILLER de la carrera profesional correspondiente.

C. Experiencia Específica

- i. Experiencia profesional comprobable de cuatro (4) años en el sector público o privado relacionado a actividades de ciencia de datos o Machine Learning, o similares.
- ii. Experiencia comprobable de haber participado como mínimo en dos (2) proyectos referidos a implementaciones de Modelos de Datos y/o Modelos Predictivos y/o Modelos de Machine Learning en el sector público o privado, en los roles de científico de datos o Ingeniero de Machine Learning o especialista o consultor relacionado a ciencia de datos.

La experiencia será contabilizada a partir de la condición de BACHILLER de la carrera profesional correspondiente.

Durante la evaluación curricular, con documentación técnica, deberá acreditar las experiencias específicas (i) y (ii).

D. Capacitación

- Diplomado o programa de especialización en Ciencia de Datos, Machine Learning o área relacionada (mínimo 80 horas acumuladas).
- Cursos en Python para ciencia de datos y/o análisis de datos.
- Certificación en Apache Spark o Spark MLlib (por ejemplo: Databricks Certified Associate Developer for Apache Spark o Cloudera Certified Professional, Azure, AWS, GCP)

Deseable capacitación en:

- Deseable Certificación en plataformas de ciencia de datos (como IBM Data Science Professional Certificate, Google Data Analytics Professional Certificate u otros)
- Deseable Cursos en metodologías ágiles aplicadas a proyectos de ciencia de datos (Scrum, Kanban)
- Deseable curso en estadística para ciencia de datos y/o para análisis de datos.

Se acreditará con copia simple de constancias, certificados u otros documentos según corresponda, de las capacitaciones requeridas.

10. MONTO, FORMA Y CONDICIONES DE PAGO

La retribución económica que pagará la Entidad por concepto de honorarios profesionales por el servicio contratado total asciende a la suma de S/. 72,000 (setenta y dos mil 00/100 soles), incluyendo los impuestos de Ley y todo costo o retención que recaiga en el servicio. No se procederá con pagos a cuenta por servicios no efectuados, ni adelanto alguno.

No incluye los costos por concepto de pasajes y viáticos que resulten necesarios para el desarrollo de la consultoría.

Cada uno de los pagos al consultor se realizará luego de la presentación y aprobación de cada uno de los productos entregados en los plazos previstos para la consultoría dispuestos en el numeral 8, los cuales deberá adjuntar el respectivo recibo de honorarios profesionales.

El pago se realizará en siete (7) armadas después de ejecutada la prestación y otorgada la conformidad, de acuerdo con el siguiente detalle:

N.º de pago	N.º de Producto	Retribución por producto S/
Primer pago	Primer Producto	S/. 0.00
	Segundo Producto	S/. 12,000
Segundo pago	Tercer Producto	S/. 12,000
Tercer pago	Cuarto Producto	S/. 12,000
Cuarto pago	Quinto Producto	S/. 12,000
Quinto pago	Sexto Producto	S/. 12,000
Sexto pago	Séptimo Producto	S/. 12,000

El consultor deberá hacer entrega de su recibo por honorarios a la entrega de cada producto asignado.

El pago se efectuará dentro de los quince (15) días calendario siguientes de la presentación de los entregables señalados en los presentes términos de referencia, para ello la Subgerencia de Sistemas de Información y Analítica de Datos deberá adjuntar la conformidad,

el producto correspondiente, el respectivo recibo de honorarios profesionales y demás documentos que se hayan estipulado.

Sin perjuicio, de lo antes señalado, y de presentarse situaciones justificadas que demanden un mayor plazo para la ejecución del trabajo a realizar por el Consultor, este podrá solicitar una ampliación de plazo a la Subgerencia de Sistemas de Información y Analítica de Datos, quien evaluará la pertinencia de dicha solicitud y otorgará o denegará su conformidad; sin que esto signifique el reconocimiento de mayores pagos a los establecidos en el presente documento. De caer el día de entrega del entregable, un día no laborable, el consultor presentará el producto al día siguiente hábil.

En caso de existir observaciones a los productos la Entidad notificará al consultor dentro de un plazo máximo de cinco (05) días hábiles posteriores a la recepción de cada producto, para lo cual el consultor tendrá un plazo máximo de tres (03) días hábiles para levantar estas observaciones; dicho plazo comenzará a partir del primer día siguiente de notificada la comunicación.

Los plazos de revisión no son contabilizados dentro del plazo efectivo del servicio. Los requerimientos de revisión a las observaciones realizadas a los productos podrán ser solicitados hasta en dos (02) oportunidades por cada producto.

11. LUGAR DE PRESTACIÓN DEL SERVICIO

La prestación del servicio podrá realizarse de manera presencial en las oficinas de la Contraloría General de la República (CGR) y/o de forma remota, en coordinación con la Subgerencia de Sistemas de Información y Analítica de Datos (SINFA) y la Gerencia de Tecnologías de la Información y Gobierno Digital (GTIGD), según lo que se acuerde para el cumplimiento del desarrollo de los productos requeridos.

La prestación del servicio no implica vínculo laboral alguno con la CGR, en tanto se trata de una consultoría de naturaleza independiente.

Durante el servicio, el consultor es el único responsable de proveer cualquier herramienta de trabajo que necesite y que la CGR no disponga.

12. COORDINACIÓN, SUPERVISIÓN Y CONFORMIDAD

El consultor deberá reportar y coordinar sus actividades con la Subgerencia de Sistemas de Información y Analítica de Datos, quien otorgará la conformidad a los entregables e informes que presente.

Asimismo, por la naturaleza de sus actividades, el consultor deberá coordinar con el personal de la Subgerencia de Sistemas de Información y Analítica de Datos y otras unidades orgánicas intervinientes en el Proyecto, siendo potestad de la Subgerencia de Sistemas de Información y Analítica de Datos la coordinación o enlace entre el consultor y otras unidades orgánicas.

13. ACCESOS

El consultor tendrá acceso a:

- Modo consulta a las bases de datos que se utilizarán para la implementación de la solución.
- Herramientas colaborativas que se usan en el proceso de desarrollo de software en la CGR.
- Otros accesos al consultor según necesidad del servicio.

14. DERECHOS DE PROPIEDAD Y CONFIDENCIALIDAD DE LA INFORMACIÓN

EL CONSULTOR deberá declarar que en la medida en que el servicio prestado es por encargo, y el costo de su ejecución es asumida por la CGR; todo producto o materiales (impresos, estudios, informes, gráficos, programas, software de computación u otros), que se genere por el servicio, es de propiedad de la CGR, no constituyéndose títulos de propiedad, derechos de autor y otro tipo de derechos para el Consultor; el mismo que a mérito del presente TDR, cede en forma exclusiva y gratuita, sin generar retribución adicional a lo estipulado en el presente documento.

EL CONSULTOR se obliga a mantener la confidencialidad y reserva absoluta en el manejo de información a la que se tenga acceso y que se encuentre relacionada con la prestación, quedando prohibido revelar dicha información a terceros (excepto al BID), además del cumplimiento de las Políticas de Seguridad de la Información de la CGR y sus procedimientos vigentes, establecidos para asegurar su cumplimiento.

En tal sentido, EL CONSULTOR deberá dar cumplimiento a todas las políticas y estándares definidos por la Entidad, en materia de seguridad de la información. Dicha obligación comprende la información que se entrega, como también la que se genera durante la realización de las actividades y la información producida una vez que se haya concluido la prestación. Dicha información puede consistir en mapas, dibujos, fotografías, mosaicos, planos, informes, recomendaciones, cálculos y demás documentos e información compilados o recibidos por EL CONSULTOR.

La CGR establece los lineamientos de seguridad de la información con proveedores en el siguiente enlace:

https://doc.contraloria.gob.pe/SGSI/Lineamientos_Seguridad_Informacion_con_Proveedores_2025.pdf

15. ANTICORRUPCIÓN Y ANTISOBORNO

A la suscripción de este contrato, EL CONSULTOR declara y garantiza no haber ofrecido, negociado, prometido o efectuado ningún pago o entrega de cualquier beneficio o incentivo ilegal, de manera directa o indirecta, a los evaluadores del proceso de contratación o cualquier servidor de la entidad contratante.

Asimismo, EL CONSULTOR se obliga a mantener una conducta proba e íntegra durante la vigencia del contrato, y después de culminado el mismo en caso existan controversias pendientes de resolver, lo que supone actuar con probidad, sin cometer actos ilícitos, directa o indirectamente.

Aunado a ello, EL CONSULTOR se obliga a abstenerse de ofrecer, negociar, prometer o dar regalos, cortesías, invitaciones, donativos o cualquier beneficio o incentivo ilegal, directa o indirectamente, a funcionarios públicos, servidores públicos, locadores de servicios o proveedores de servicios de la Unidad Orgánica Responsable (UOR), de la dependencia encargada de la contratación, actores del proceso de contratación y/o cualquier servidor de la entidad contratante, con la finalidad de obtener alguna ventaja indebida o beneficio ilícito. En esa línea, se obliga a adoptar las medidas técnicas, organizativas y/o de personal necesarias para asegurar que no se practiquen los actos previamente señalados.

Adicionalmente, EL CONSULTOR se compromete a denunciar oportunamente ante las autoridades competentes los actos de corrupción o de inconducta funcional de los cuales tuviera conocimiento durante la ejecución del contrato con LA ENTIDAD CONTRATANTE.

Tratándose de una persona jurídica, lo anterior se extiende a sus accionistas, participacionistas, integrantes de los órganos de administración, apoderados, representantes legales, funcionarios, asesores o cualquier persona vinculada a la persona jurídica que representa; comprometiéndose a informarles sobre los alcances de las obligaciones asumidas en virtud del presente contrato.

Finalmente, el incumplimiento de las obligaciones establecidas en esta cláusula, durante la ejecución contractual, otorga a LA ENTIDAD CONTRATANTE el derecho de resolver total o

parcialmente el contrato. En ningún caso, dichas medias impiden el inicio de las acciones civiles, penales y administrativas a que hubiera lugar.

EL CONSULTOR se compromete en cumplir la política antisoborno del programa aprobada y regulada mediante resolución la cual se encuentra disponible en el siguiente enlace: web: <https://cdn.www.gob.pe/uploads/document/file/3248111/Resoluci%C3%B3n%20de%20Contralor%C3%ADa%20N%C2%B0%20229-2022-CG.pdf.pdf?v=1654888578>.

16. PENALIDADES

a) **Penalidad diaria por mora**

- La penalidad se aplica por cada día de retraso injustificado en la prestación de la ejecución del servicio.
- Penalidad diaria:
Penalidad diaria = $0.10 \times \text{Monto del contrato} / (F \times \text{Plazo en días})$
Donde $F = 0.40$

a) **Consideraciones generales:**

- Cuando el Consultor no cumpla con **la presentación de los productos/entregables de acuerdo con las características y condiciones estipuladas**, se **considera como no presentado**, aplicándose la penalidad por cada día de retraso hasta su presentación, en este escenario, **la Entidad** comunicará al Consultor dicho incumplimiento.
- La penalidad se deducirá de pagos parciales o finales.
- Se exceptúa la penalidad si el Consultor acredita objetivamente que el retraso no le es imputable.
- Al alcanzar el límite máximo del 10 % acumulado de las penalidades por mora, la Entidad podrá resolver el contrato por incumplimiento.

a) **Identificación del retraso injustificado:**

- Cuando el Consultor **subsana las observaciones dentro del plazo otorgado**, no se aplicará penalidad.
- Cuando el Consultor **cumple con levantar las observaciones fuera del plazo otorgado en las dos (02) rondas de observaciones debidamente notificadas**, se considera la aplicación de penalidad por cada día de retraso contabilizado desde el día siguiente de la fecha de vencimiento del plazo otorgado.
- **Cuando vencido el plazo otorgado** para subsanar las dos (02) rondas de observaciones y el Consultor/Proveedor/Contratista presenta dentro del plazo otorgado, pero **persisten las observaciones¹**, La Unidad Orgánica responsable (UOR) en mérito a su evaluación técnica, podrá otorgar periodos adicionales **para el levantamiento de la persistencia de observaciones**, sin perjuicio de las penalidades respectivas. En este caso, se aplica la penalidad computando los días de retraso **desde el vencimiento del plazo otorgado con la primera notificación de observación, hasta que la subsane**, incluyéndose el plazo formal con el que cuenta la Unidad Orgánica responsable (UOR) para su evaluación.

17. ANEXOS:

Anexo N.º 01: Lineamientos para la presentación de documentos

Anexo N.º 02: Arquitectura de la Plataforma de Minería de Datos

Anexo N.º 03: Criterio de Aceptación (Checklist)

¹ Las Observaciones formuladas contarán con las rondas que se establezcan en el contrato, en caso surjan **nuevas observaciones que no hayan podido ser identificadas anteriormente, considerando la complejidad y la naturaleza de cada contratación, se procederá con reiniciar las rondas establecidas para estos casos específicos.**

Anexo N.º 01: Lineamientos para la presentación de documentos

I. ESPECIFICACIONES GENERALES

Los informes deben redactarse teniendo en cuenta las siguientes especificaciones:

1. Letra arial 11.
2. Espacio simple.
3. Carátula indicando, entre otros, nombre de consultoría, nombre de consultor y número de producto.
4. Impresión a doble cara.
5. Páginas numeradas en la parte inferior derecha.
6. Índice numerado de páginas.

II. ESPECIFICACIONES POR TIPO DE PRODUCTO

II.1. Plan de Trabajo

Tendrá la siguiente estructura:

1. Carátula
2. Índice
3. Introducción
4. Objetivo de consultoría
5. Productos a alcanzar
6. Actividades a cumplir por cada producto
7. Cronograma
8. Anexo(s)

II.2. Informe del Producto

Tendrá la siguiente estructura:

1. Carátula
2. Resumen ejecutivo
3. Índice
4. Introducción
5. Objetivo de consultoría
6. Productos alcanzados
7. Actividades realizadas
8. Grado de cumplimiento del producto
9. Dificultades y limitaciones encontradas
10. Conclusiones y Recomendaciones
11. Anexo(s)

II.3. Informe Final

Tendrá la siguiente estructura:

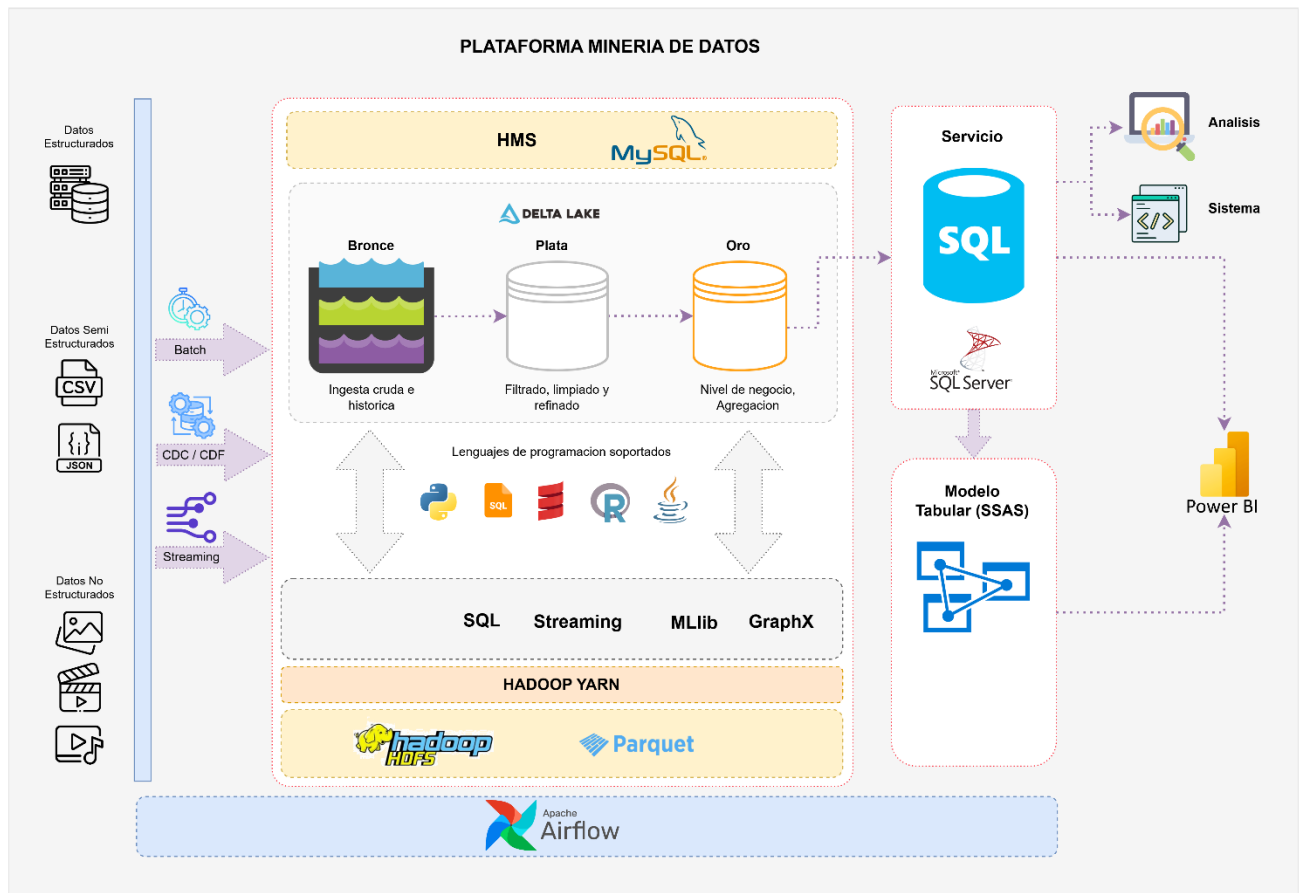
1. Carátula
2. Resumen ejecutivo
3. Índice
4. Introducción
5. Objetivo de consultoría
6. Productos alcanzados
7. Conclusiones y Recomendaciones
8. Anexo(s)

III. CONSIDERACIONES GENERALES DEL PRODUCTO, A SER TENIDAS EN CUENTA:

- a) Tapa del documento en el que se precisa el nombre de la consultoría, nombre del producto, el nombre del autor, la fecha de presentación y el nombre y logo de la CGR.
- b) Incluir índice de capítulos, así como de tablas o cuadros y de gráficos cuando corresponda.
- c) Incluir una lista de abreviaturas o acrónimos, en caso de que se usen siglas en el documento.
- d) Incluir un glosario de términos que requieran de explicación inicial para facilitar la lectura del documento.
- e) El resumen ejecutivo dará cuenta de los aspectos más relevantes del trabajo encargado.
- f) De acuerdo con la naturaleza y características del producto a entregar, el documento se dividirá en capítulos, los que estarán debidamente numerados.
- g) Las páginas del documento estarán debidamente numeradas.
- h) Las referencias bibliográficas deberán incluirse al final del documento.
- i) El consultor debe presentar sus entregables en Mesa de Partes Virtual de la Contraloría General de la República (<https://mesadepartesvirtual.contraloria.gob.pe/mpvirtual/>).
- j) El pago correspondiente a cada producto se detalla en el numeral 10 "MONTOS, FORMAS Y CONDICIONES DE PAGO".
- k) El Consultor se compromete a ceder los derechos patrimoniales de autor de los productos y documentos elaborados.
- l) El Consultor se compromete a guardar reserva de toda aquella información interna a la que tenga acceso para la ejecución de esta consultoría, cualquier uso de esta información, deberá ser autorizada previamente por la CGR.
- m) Todos los productos deberán de ser entregados y sustentados en la forma y plazos que se indican en estos Términos de Referencia.
- n) A la entrega del último producto, se adjuntarán las bases de datos, códigos de programación u otros materiales utilizados por el Consultor o que le hayan sido entregados a este por la CGR durante el proceso de ejecución de la consultoría.

ANEXO 2

ARQUITECTURA DE LA PLATAFORMA DE MINERÍA DE DATOS



Anexo N.º 03: Criterio de Aceptación (Checklist)

Para los productos que implican la entrega de código fuente se considerará el siguiente checklist de criterios de aceptación:

N.º	Criterio de Aceptación	Categoría	SI	NO
1	Exploración del análisis de Datos	Documentación		
2	Calidad e Ingeniería de Características (Feature Engineering): Uso exclusivo de capas Plata/Oro del Datamart institucional. Justificación estadística de nulos y outliers.	Preparación de Datos		
3	Estándares Institucionales de Extracción SQL: Cumplimiento de Reglas Técnicas de la CGR (uso de LEFT JOIN, lógica de "cortocircuito" y prohibición de <i>Full Table Scans</i> en tablas de hechos)	Estándares de Código		
4	Performance y Validación del Modelo: Superar umbrales mínimos de Accuracy, F1-Score y AUC-ROC.	Analítica Avanzada		
5	Interpretabilidad Orientada a Auditoría (Caja Blanca): Entrega de artefactos de explicabilidad (Valores SHAP, LIME o Feature Importance) para sustentar los hallazgos ante auditores.	Valor de Negocio		
6	Despliegue, Autenticación y MLOps: Código versionado en Git institucional	Infraestructura		
7	Integración para Ecosistema SSRS (SQL Server Reporting Services): Resultados del modelo (predicciones y puntajes de riesgo)	Integración		
8	Documentación de Linaje y Diccionario de Datos: Documentación con el rastreo del dato desde la fuente hasta el modelo final.	Documentación Técnica		
9	Hiperparámetros: Documentación de los hiperparámetros de los modelos y sus respectivas explicaciones	Analítica Avanzada		
10	Resultados de Métricas: Reporte completo de los resultados obtenidos en las métricas de desempeño de los modelos	Analítica Avanzada		
11	Oportunidades de Mejora: Identificación y documentación de mejoras futuras para la evolución del modelo	Valor de Negocio		