

## TÉRMINOS DE REFERENCIA

### CONTRATACIÓN DE CONSULTORES ESPECIALIZADOS EN SOLUCIONES DE BIG DATA Y LANDING ZONE PARA EL PROCESO DE FORTALECIMIENTO DE LA CIENCIA DE DATOS

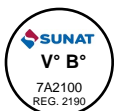
CONTRATO DE PRÉSTAMO BID N° 4725/OC-PE	
Denominación del Componente	Mejora de la inteligencia fiscal y la gestión de riesgo tributario y aduanero
Sub componente	3.2 Gestión de riesgos mediante el uso de modelos analíticos predictivos y de ciencia de datos
Plazo	18 meses y 10 días calendario, incluidos los plazos de aprobaciones y conformidades.

Cuando se mencionen en el presente documento los siguientes términos y expresiones, tendrán el significado que se indica a continuación:

- a) **SUNAT:** Superintendencia Nacional de Aduanas y de Administración Tributaria.
- b) **BID:** Banco Interamericano de Desarrollo.
- c) **CT:** Coordinación Técnica de la Unidad Ejecutora Mejoramiento del Sistema de Información de la SUNAT (UEMSI).
- d) **UEMSI:** Unidad Ejecutora "Mejoramiento del Sistema de Información de la SUNAT" - MSI
- e) **MEF:** Ministerio de Economía y Finanzas.
- f) **INER:** Intendencia Nacional de Estrategias y Riesgos de la SUNAT.
- g) **GAIR:** Gerencia de Análisis de Información y Riesgos de la INER.
- h) **DIA:** División de Información y Análisis de la GAIR.
- i) **INSI:** Intendencia Nacional de Sistemas de Información de la SUNAT.
- j) **MACHINE LEARNING (ML):** El aprendizaje automático o aprendizaje automatizado o aprendizaje de máquinas (del inglés, machine learning) es el subcampo de las ciencias de la computación y una rama de la inteligencia artificial, cuyo objetivo es desarrollar técnicas que permitan que las computadoras "aprendan".
- k) **BIG DATA:** Es el manejo de grandes volúmenes de datos (estructurados y no estructurados) que conducen a tomar decisiones estratégicas.
- l) **CONTENEDORES:** Un contenedor es un formato que empaqueta todo el código y las dependencias de una aplicación en un formato estándar que permite su ejecución rápida y fiable en todos los entornos informáticos.
- m) **INTELIGENCIA ARTIFICIAL (IA):** Se refiere a sistemas o máquinas que imitan la inteligencia humana para realizar tareas y pueden mejorar iterativamente a partir de la información que recopilan.
- n) **APPLICATION PROGRAMMING INTERFACE (API):** Interfaz de programación de aplicaciones.



OMAR IVAN DAVILA  
PERALES  
GERENTE  
18/04/2022 15:17:08



FREDDY MARCIAL  
QUEVEDO CAJAS  
JEFE DE DIVISION  
17/04/2022 12:24:22



CECILIA DEL PILAR  
LEON MESTANZA  
SUPERVISOR  
ENCARGADO  
14/04/2022 21:13:46

## 1. ANTECEDENTES

La República del Perú suscribió con el Banco Interamericano de Desarrollo – BID el 15 de febrero de 2019 el Contrato de Préstamo N° 4725/OC-PE, para financiar el Proyecto de Inversión Pública (PIP) “Mejoramiento de los Servicios de Recaudación Tributaria y Aduanera a través de la Transformación Digital”, en adelante el PROYECTO, cuya ejecución está a cargo de la Unidad Ejecutora Mejoramiento del Sistema de Información de la SUNAT – UEMSI.

El referido PIP tiene los siguientes componentes: i) Mejora del modelo de gobernanza institucional en SUNAT, ii) Mejora del control y cumplimiento tributario y aduanero y iii) Mejora de la inteligencia fiscal y la gestión de riesgo tributario y aduanero; este último tiene el sub-componente “Gestión de riesgos mediante modelos analíticos predictivos y de ciencia de datos”.

La ciencia de datos forma parte del Modelo Integral de Gestión de Riesgos que la institución viene desarrollando en la INER, proceso que es cíclico, y que se sustenta en siete (7) fases o etapas: (1) análisis del contexto estratégico y operativo, (2) identificación de incumplimientos, (3) valorización de incumplimientos, (4) análisis del comportamiento y riesgos, (5) establecimiento del tratamiento, (6) ejecución del tratamiento, y (7) monitoreo y evaluación.

En ese contexto, hasta la fecha se viene aplicando la metodología de ciencia de datos en el desarrollo de modelos analíticos, que han permitido implementar productos que atienden parte de los riesgos de incumplimiento tributario, entre otros, los vinculados con: (i) selección de contribuyentes por niveles de riesgo, (ii) detección de incumplimientos e inconsistencias a partir de la evaluación masiva de grandes volúmenes de datos, (iii) prevención del incumplimiento, y (iv) optimización en la priorización de contribuyentes para tratamientos posteriores. Esto mediante técnicas como web scrapping, text mining, algoritmos de simulación y optimización, y algoritmos de machine learning aplicados sobre data estructurada y no estructurada, acompañadas del conocimiento tributario de los especialistas del negocio.

## 2. FINALIDAD PÚBLICA

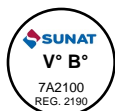
La SUNAT tiene la misión de servir al país proporcionando los recursos necesarios para la sostenibilidad fiscal y la estabilidad macroeconómica. En ese contexto, la Administración tiene como prioridad incrementar la presión tributaria reduciendo las brechas de incumplimiento, manteniendo la confianza en el sistema tributario y su administración.

Dentro de su Plan Estratégico Institucional para el período 2018-2024, con la finalidad de alcanzar la efectividad tributaria, la SUNAT estableció entre uno de sus principales objetivos estratégicos institucionales el de Mejorar el Cumplimiento Tributario y Aduanero.

Alineado a dicho objetivo estratégico, la SUNAT pretende ampliar su capacidad de explotación de la información, para identificar riesgos y formular hipótesis de incumplimiento tributario, con la finalidad potenciar su detección y formulación de tratamientos, siendo necesaria la incorporación de recursos.



OMAR IVAN DAVILA  
PERALES  
GERENTE  
18/04/2022 15:17:08



FREDDY MARCIAL  
QUEVEDO CAJAS  
JEFE DE DIVISION  
17/04/2022 12:24:22



CECILIA DEL PILAR  
LEON MESTANZA  
SUPERVISOR  
ENCARGADO  
14/04/2022 21:13:46

En esa línea, la SUNAT busca incorporar recursos para el desarrollo de los modelos analíticos avanzados, con uso de la ciencia de datos, para la identificación de riesgos y formulación de hipótesis de incumplimiento tributario, con la finalidad potenciar la detección de riesgos de incumplimiento y la formulación de tratamientos que respondan a los distintos comportamientos.

### 3. OBJETIVO

#### 3.1. OBJETIVO GENERAL

Contratar los servicios de dos (02) Consultores Especializados en Arquitectura de Big Data para el desarrollo e implementación de modelos de analítica avanzada de la INER, y el monitoreo de los módulos del ambiente de Big Data en el cual se desarrollan, implementan y despliegan dichos modelos, a fin de dar continuidad al ciclo de la ciencia de datos y optimizar los diferentes procesos dentro del equipo de analítica avanzada que viene trabajando en la INER desde hace más de cuatro (4) años, en el marco del Modelo Integral de Gestión de Riesgos de la SUNAT. Asimismo, facilitar la transición al nuevo entorno de trabajo Landing Zone a ser adquirido por la SUNAT.

#### 3.2. OBJETIVOS ESPECÍFICOS

- Potenciar el conocimiento especializado de analítica avanzada adoptando métodos y herramientas que integren técnicas de preprocesamiento, implementación, monitoreo y la retroalimentación de los modelos de analítica avanzada.
- Desarrollar soluciones analíticas por lotes para la implementación de proyectos analíticos avanzados con enfoque en la recopilación, análisis y visualización de datos.
- Diseñar e implementar una línea de producción que permita operativizar y desplegar los modelos de analítica avanzada desarrollados por la INER en el ambiente de Big Data.
- Incorporar el uso de métodos, técnicas, herramientas y plataformas para la automatización, despliegue, monitoreo de los modelos de analítica avanzada que optimicen los resultados del trabajo del equipo de ciencia de datos.
- Monitorear que los módulos del ambiente de Big Data se encuentren operativos de acuerdo con las necesidades del equipo de ciencia de datos y su ejecución se realice en los tiempos establecidos.
- Potenciar el equipo de ciencia de datos mediante asesoría especializada en el uso de los módulos que brindan soporte al ambiente Big Data en el cual se desarrollan, implementan y despliegan los modelos de analítica avanzada.
- Mejorar el performance de la plataforma de información para la implementación de modelos analíticos, predictivos y prescriptivos basados en ciencia de datos.

### 4. MARCO DE REFERENCIA

La consultoría deberá tomar en consideración lo siguiente:

- El Contrato de Préstamo N° 4725/OC-PE.
- Documento de Organización y Funciones Provisional de la SUNAT.
- Normas internas de la SUNAT en lo que corresponda.



OMAR IVAN DAVILA  
PERALES  
GERENTE  
18/04/2022 15:17:08



FREDDY MARCIAL  
QUEVEDO CAJAS  
JEFE DE DIVISION  
17/04/2022 12:24:22



CECILIA DEL PILAR  
LEON MESTANZA  
SUPERVISOR  
ENCARGADO  
14/04/2022 21:13:46

## 5. ALCANCE DEL TRABAJO

Desarrollar e implementar modelos de analítica avanzada, así como monitorear los módulos del ambiente Big Data y la disponibilidad de los datos con los cuales se desarrollan, implementan y despliegan los modelos de analítica avanzada, a efectos de potenciar la detección y formulación de tratamientos, proporcionando soporte a la gestión de riesgos de incumplimiento tributario en la SUNAT. Asimismo, realizar el acompañamiento y capacitación al equipo de trabajo y usuarios finales sobre los modelos implementados.

## 6. FUNCIONES

### 6.1. Planificación relativa a los proyectos

Durante esta actividad, la consultoría especializada deberá alinear las expectativas de proyecto de ciencia de datos con respecto a la infraestructura y tecnología disponible en la SUNAT, validando los entregables, alcance y plazos mediante la interacción en reuniones o entrevistas. Asimismo, debe incluir la elaboración de un diagnóstico inicial del estado de la actual plataforma de Big Data de la SUNAT.

### 6.2. Desarrollo de estrategia de datos

Deberá desarrollar e implementar estrategias de datos para los proyectos de analítica avanzada. Estas estrategias deben incluir al menos lo siguiente: diseños de modelos de datos, estándares de desarrollo, la implementación, gestión de almacenes de datos, calidad de datos y otros módulos que darán soporte al ambiente Big Data.

### 6.3. Recopilación y la comprensión de datos

Deberá identificar y preparar los datos disponibles (estructurados, no estructurados y semiestructurados) y relevantes para el dominio del objetivo a ser atendido. Al obtener los datos, se deberán utilizar técnicas que aseguren su calidad, confiabilidad y trazabilidad.

### 6.4. Despliegue y monitoreo del modelo

El objetivo de esta actividad es facilitar la puesta en producción de los distintos modelos desarrollados por los equipos de ciencia de datos, entre otros, mediante el uso de APIs en el ambiente Big Data. Además, monitorear el rendimiento de los modelos puestos en producción mediante distintas métricas.

### 6.5. Auditoria del ambiente Big Data

Se realizarán procesos de auditoría periódicas para valorar el rendimiento de los distintos módulos del ambiente Big Data y elaborar informes de rendimiento y estado situacional. Lo que permitirá perfeccionar el trabajo constante del equipo de ciencia de datos.



OMAR IVAN DAVILA  
PERALES  
GERENTE  
18/04/2022 15:17:08



FREDDY MARCIAL  
QUEVEDO CAJAS  
JEFE DE DIVISION  
17/04/2022 12:24:22



CECILIA DEL PILAR  
LEON MESTANZA  
SUPERVISOR  
ENCARGADO  
14/04/2022 21:13:46

## 6.6. Otras actividades

- En los casos que se requiera, deberá plantear soluciones para el monitoreo, retroalimentación de los modelos y mejorar su precisión y utilidad.
- Automatizar, a requerimiento, algunos o todos los pasos del desarrollo de los modelos y de la recopilación de retroalimentación.
- Otras actividades que se requieran como presentaciones, participación en eventos y sesiones de acuerdo con las competencias relacionadas al perfil y a los requerimientos de las iniciativas y proyectos de la ciencia de datos.
- La entrega de los informes preliminares y las presentaciones que forman parte del desarrollo de las actividades y que no constituyen entregables, deberán ser entregados a la DIA a través de correo electrónico o el depósito en el espacio virtual que disponga.

En todos los casos, como parte del objetivo específico de participación cruzada, en el desarrollo de todas las actividades, el consultor especializado podrá participar también asesorando en la creación de variables/indicadores, en el modelado o en la evaluación de los modelos priorizados, desarrollados y/o en desarrollo por otros profesionales, incluso respecto de aquellos ya implantados.

El desarrollo y atención de todas las actividades por parte del consultor especializado deberán contar con la opinión favorable de la DIA.

## 7. ENTREGABLES

El servicio de consultoría iniciará a partir del día siguiente útil de la suscripción del contrato y tendrá un plazo de duración de dieciocho (18) meses y diez (10) días calendario, plazo que incluye las aprobaciones y conformidades.

- Informes mensuales detallando las actividades realizadas, resultados, recomendaciones y lecciones aprendidas.
- Un informe de diagnóstico de la situación actual de la plataforma de trabajo.
- Un informe de propuesta sobre Metodologías de implementación de soluciones de Big Data.
- Un informe final que contenga:
  - Detalle de las actividades realizadas.
  - Entrega ordenada de la documentación generada durante el desarrollo de sus actividades, incluyendo la constancia de entrega a la DIA del código o comandos estructurados para la ejecución de los modelos analíticos.
  - Conclusiones y recomendaciones.

Los plazos estimados para la conformidad de cada entregable por parte de la Coordinación Técnica de la UEMSI serán de diez (10) días calendarios.

## 8. REQUISITOS MÍNIMOS DEL CANDIDATO Y PERFIL DE COMPETENCIAS

N°	Requisito	Criterio
01	Formación Académica	Bachiller, titulado o licenciado en Ingeniería Informática y/o Ingeniería de Sistemas y/o Ciencias de la Computación y/o afines.

02	Estudios adicionales	<p>De preferencia contar con estudios y/o certificaciones en:</p> <ul style="list-style-type: none"> <li>- Maestría y/o Diplomados en Ciencias de la Computación y/o Big Data y/o Procesamiento Distribuido y/o afines; y/o</li> <li>- CCA Spark and Hadoop Developer; y/o</li> <li>- CCA Administrator; y/o</li> <li>- Azure Databricks Certified Associate Platform Administrator; y/o</li> <li>- Databricks Certified Associate Developer for Apache Spark 3.0; y/o</li> <li>- AWS Certified Big Data</li> </ul> <p>Se otorgará puntaje adicional.</p>
03	Experiencia general	Experiencia general, al menos cinco (05) años en entidades del sector público y/o privado.
04	Experiencia específica	<p>Experiencia específica, al menos tres (03) años como Arquitecto de Datos y/o Ingeniero de Datos y/o afines en el sector público y/o privado, acreditada mediante:</p> <ul style="list-style-type: none"> <li>- Constancias de prestación de servicio.</li> <li>- Certificado de trabajo.</li> <li>- Orden de servicio o contrato con su respectiva conformidad.</li> </ul>
05	Conocimientos	<ul style="list-style-type: none"> <li>- Lenguajes de programación: R y/o Python y/o Scala; y/o</li> <li>- Manejo de acceso a fuentes de datos en Teradata y/o Oracle y/o Informix y/o MongoDB y/o Bus de eventos Kafka; y/o</li> <li>- Conocimiento de Flujo de Datos por lotes y Streaming para ingesta de datos; y/o</li> <li>- Conocimiento en arquitectura on premise y/o cloud: Hadoop y/o Cloudera y/o GCP y/o AWS, Oracle Cloud y/o IBM Analytics; y/o</li> <li>- Conocimiento en contenedores: Kubernetes y/o Docker y/o openshift y/o afines; y/o</li> <li>- Optimización de plataformas con procesamiento distribuido; y/o</li> <li>- Herramientas de BI</li> </ul> <p>La forma de acreditar los conocimientos será mediante Declaración Jurada en la Hoja de Vida del consultor.</p>
06	Entrevista	<p>Se evaluará:</p> <ul style="list-style-type: none"> <li>- Dominio temático.</li> <li>- Capacidad analítica.</li> <li>- Capacidad para trabajar en equipo.</li> <li>- Iniciativa.</li> </ul>



OMAR IVAN DAVILA PERALES  
GERENTE  
18/04/2022 15:17:08



FREDDY MARCIAL QUEVEDO CAJAS  
JEFE DE DIVISION  
17/04/2022 12:24:22



CECILIA DEL PILAR LEON MESTANZA  
SUPERVISOR ENCARGADO  
14/04/2022 21:13:46

## 9. LUGAR DE TRABAJO

La prestación será realizada en las instalaciones de la Av. Garcilaso de la Vega 1456 - Cercado de Lima, Edificio Sulamerica Piso 10, en las oficinas de la División de Información y Análisis, así como en los demás ambientes de dicha sede, que sean necesarios para el cumplimiento de la prestación. La SUNAT a efectos de salubridad adopta las especificaciones señaladas por la Ley N° 31246 y de manera excepcional o



temporal a través del área usuaria modifica o autoriza el trabajo remoto y/o el lugar de prestación de servicios, según lo previsto en el Decreto de Urgencia N° 115-2021 en el marco de la emergencia sanitaria por el COVID-19.

No obstante, el consultor especializado podrá desplazarse a las diferentes sedes de la SUNAT (Lima metropolitana y el Callao), previa autorización de la DIA, para el cumplimiento de lo solicitado en la consultoría.

Los entregables serán presentados en forma digital (en PDF y archivos en formato original editable) y/o físico, en mesa de partes virtual y/o en las instalaciones de la SUNAT (Av. Garcilaso de la Vega N° 1472 – Lima 1), a través de Mesa de Partes en el horario de 08:30 a.m. a 23:59 p.m., los mismos deberán estar dirigidos a la DIA, quien informará lo pertinente a la Coordinación Técnica de la UEMSI.

## 10. UNIDAD ORGANIZACIONAL RESPONSABLE DE OTORGAR CONFORMIDAD PREVIA AL PAGO

La coordinación y la aprobación de los informes del Consultor Especializado en Arquitectura de Big Data y Analítica Avanzada contratado estarán a cargo del Coordinador Técnico de la UEMSI, previa opinión favorable de la DIA de la GAIR de la INER.



OMAR IVAN DAVILA  
PERALES  
GERENTE  
18/04/2022 15:17:08

## 11. COSTO DEL SERVICIO, FINANCIAMIENTO, PLAZO Y FORMA DE PAGO

El costo total de los servicios asciende a **S/ 360 000.00** (trescientos sesenta mil y 00/100 soles), importe que está sujeto a la respectiva retención de impuesto a la renta.

El costo mensual de los servicios asciende a **S/ 10 000.00** (diez mil y 00/100 soles), importe que está sujeto a la respectiva retención de impuesto a la renta.

El consultor especializado prestará los servicios objeto del presente contrato a partir del día siguiente útil de la fecha de la suscripción del contrato, por un plazo de dieciocho (18) meses y diez (10) días calendario, incluidos los plazos de aprobación y conformidad.

La forma de pago será mensual por los servicios prestados, previa aprobación del respectivo informe del consultor por parte de la Coordinador Técnico de la UEMSI, previa opinión de la DIA.



FREDDY MARCIAL  
QUEVEDO CAJAS  
JEFE DE DIVISION  
17/04/2022 12:24:22



CECILIA DEL PILAR  
LEON MESTANZA  
SUPERVISOR  
ENCARGADO  
14/04/2022 21:13:46

## 12. MATERIALES Y EQUIPOS

La SUNAT podrá proporcionar un equipo informático y habilitar una cuenta de correo electrónico institucional y demás accesos de manera temporal al consultor especializado, con la finalidad de que pueda realizar las coordinaciones que sean necesarias; así como material y útiles necesarios con la finalidad de que pueda realizar las prestaciones.

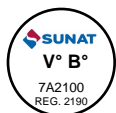
### 13. CONFIDENCIALIDAD Y PROPIEDAD INTELECTUAL

El consultor especializado se obliga a no difundir, aplicar ni comunicar a terceros información, base de datos, documentos ni cualquier otro aspecto relacionado a la SUNAT a la que tenga acceso, durante la ejecución del servicio y después de la finalización del mismo. Asimismo, el consultor especializado no podrá publicar las recomendaciones formuladas en el curso de la contratación, o como resultado de la prestación de los servicios. En caso el consultor especializado incumpla con la confidencialidad, la SUNAT a su sola discreción podrá rescindir el contrato y además adoptar las acciones legales que correspondan.

Los entregables, documentos, archivos y en general cualquier información o conocimiento generados durante el servicio, serán de propiedad única y exclusiva de la SUNAT, quedando prohibido su uso por parte del consultor especializado, salvo autorización expresa de la SUNAT.



OMAR IVAN DAVILA  
PERALES  
GERENTE  
18/04/2022 15:17:08



FREDDY MARCIAL  
QUEVEDO CAJAS  
JEFE DE DIVISION  
17/04/2022 12:24:22



CECILIA DEL PILAR  
LEON MESTANZA  
SUPERVISOR  
ENCARGADO  
14/04/2022 21:13:46