

TÉRMINOS DE REFERENCIA

CONTRATACIÓN DE CONSULTORES ESPECIALIZADOS EN EL DESARROLLO DE MODELOS ANALÍTICOS AVANZADOS PARA EL PROCESO DE CONSOLIDACIÓN 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.

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.


V° B°
7A2000
REG. 5032
OMAR IVAN DAVILA
PERALES
GERENTE
18/04/2022 15:17:08


V° B°
7A2100
REG. 2190
FREDDY MARCIAL
QUEVEDO CAJAS
JEFE DE DIVISION
17/04/2022 12:24:22


V° B°
7A2102
REG. 0177
CECILIA DEL PILAR
LEON MESTANZA
SUPERVISOR
ENCARGADO
14/04/2022 21:13:46

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.

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.



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

3. OBJETIVO

3.1. OBJETIVO GENERAL

Contratar los servicios de seis (06) Consultores Especializados en Ciencia de Datos para la planificación, desarrollo e implementación de modelos analíticos avanzados para la identificación de riesgos y formulación de hipótesis de incumplimiento tributario vinculados a los diferentes tratamientos de la Administración Tributaria, con la finalidad de consolidar la detección de riesgos de incumplimientos y la formulación de tratamientos que respondan a los distintos comportamientos y procesos; incorporando además el análisis predictivo, prescriptivo y de aprendizaje profundo, 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.

3.2. OBJETIVOS ESPECÍFICOS

- Potenciar el conocimiento especializado de analítica avanzada para el diseño y proyección de escenarios futuros, adoptando métodos y herramientas que integren la implementación, monitoreo y la retroalimentación de los modelos de analítica avanzada y su ajuste para una mayor precisión, para optimizar el resultado de los tratamientos propuestos.
- Potenciar la implementación de manera significativa de una línea de producción oportuna que permita operativizar y desplegar los modelos analíticos.
- Incorporar el uso de métodos y técnicas de aprendizaje profundo en data no estructurada, mediante el uso de herramientas y plataformas de Big Data que optimicen los resultados del equipo de ciencia de datos.
- Incorporar sistemas de seguridad para modelos de Machine Learning en las diversas etapas del desarrollo, implementación y retroalimentación de los modelos de analítica, para el soporte de la gestión de riesgos de cumplimiento tributario.
- Incorporar métodos de “Model-agnostic” para la interpretabilidad de modelos de Machine Learning y Deep Learning con el fin de operativizar los modelos analíticos.
- Proponer mejoras metodológicas al modelo de ciencia de datos actual implementado en la INER.
- Diseñar e implementar métricas para back testing asociados a los riesgos tributarios de los modelos analíticos que están en producción.

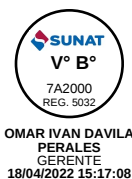
4. MARCO DE REFERENCIA

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.

5. ALCANCE DEL TRABAJO

Planificar, desarrollar e implementar algoritmos avanzados, adaptables y métodos complejos de análisis exploratorio, segmentación descriptiva, modelado probabilístico, y análisis descriptivo, predictivo y prescriptivo; referidos a la identificación de riesgos de



incumplimiento tributario y a la formulación de hipótesis de incumplimiento tributario, con la finalidad potenciar la detección y formulación de tratamientos, proporcionando soporte al Modelo Integral de Gestión de Riesgo en la SUNAT y en su vinculación con las diferentes acciones institucionales.

6. FUNCIONES

6.1. Planificación del ciclo de vida de los modelos analíticos

Durante esta actividad, los consultores deberán alinear expectativas con los objetivos de los modelos de analítica avanzada y la forma como se van a desarrollar, validando los entregables, alcance, plazos, participantes y la interacción mediante reuniones o entrevistas. Esta fase implica la participación que permita lograr las precisiones sobre los siguientes temas:

- Objetivo, señalando explícitamente qué riesgos busca identificar con cada uno de los modelos a elaborar, previa determinación de la situación actual, en relación con la eficacia en la predicción de los modelos en producción.
- Alcance, indicando los segmentos, dependencias y/o áreas en los que serán aplicados los modelos a desarrollar.
- Cronograma de actividades y el proceso de Machine Learning Operations, marco del sistema escalado de todo el ciclo de vida de los modelos analíticos.

6.2. Comprensión del negocio y enfoque analítico

Durante esta actividad, se deberán analizar los incumplimientos o riesgos que fueron planteados en la fase de planificación, se evalúa la factibilidad de los posibles planteamientos y se ajustan algunas definiciones. Asimismo, se deberá definir la granularidad o grado de desagregación con la que se debe extraer la información disponible.

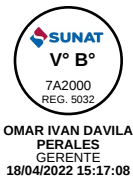
Se deberán identificar los atributos que se vinculan a las hipótesis planteadas por los especialistas del negocio, así como también las fuentes de información de donde se pueden obtener dichos atributos, con el fin de ser utilizados para el análisis de datos referidos al desarrollo de los modelos analíticos avanzados priorizados.

En esta actividad se definen tanto los datos primitivos como los atributos calculados a partir de datos primitivos que serían útiles para alcanzar el objetivo planteado, concluyendo con la expresión del objetivo bajo el contexto de las técnicas estadísticas y de aprendizaje automático, para que se puedan identificar las más adecuadas para el resultado deseado.

6.3. Requisitos, recopilación y la comprensión de datos

Se deberá determinar los requisitos de los datos en base al enfoque analítico elegido, referidos al desarrollo de los modelos analíticos avanzados priorizados.

Para cada modelo analítico planteado se establecerá el contenido de datos, formatos y representaciones. En la etapa inicial de recopilación de datos deberán identificar y reunir los recursos de datos disponibles (estructurados, no estructurados y semiestructurados) y relevantes para el dominio del objetivo a ser atendido.



Para el caso de los modelos predictivos, se deberá determinar las fuentes de información requeridas y disponibles para la obtención de las variables objetivo o target. Asimismo, se definen los métodos de extracción y recopilación de estos datos.

Al obtener los datos, se deberán utilizar estadísticas descriptivas y técnicas de visualización para alcanzar un grado de comprensión adecuado de la data a partir de la exploración de sus características en los ambientes disponibles, evaluar la calidad y descubrir las ideas iniciales, así como también esbozar el manejo más conveniente que será aplicado de acuerdo con la técnica de minería de datos a ser utilizada.

6.4. Preparación de los datos

Durante esta etapa, se deberá tener los datos preparados para iniciar el modelado de la solución de minería de datos. Esto implica realizar las actividades de limpieza de datos, de selección de datos de los atributos, construcción de atributos, transformación y tratamiento de atributos, así como obtener el nivel de agregación y desagregación requerido.

Se deberá ejecutar el proceso de ingeniería de características para crear variables explicativas adicionales, también conocidas como indicadores o características, a través de una combinación de conocimiento en el dominio y de variables estructuradas existentes.

6.5. Diseño del modelo

El objetivo de esta actividad es realizar la selección de la técnica y algoritmo de machine learning o aprendizaje profundo más apropiado de acuerdo con la solución planteada al problema, considerando el potencial desequilibrio de clases en los conjuntos de datos para luego construir el modelo respectivo.

En esta actividad se utiliza la primera versión del conjunto de datos preparado y se enfoca en desarrollar modelos predictivos o descriptivos según el enfoque analítico previamente definido. Para una técnica determinada, podrán probarse múltiples algoritmos con sus respectivos parámetros con el fin de encontrar el mejor modelo con las variables disponibles mediante una gestión de experimentos y modelos.

6.6. Evaluación del modelo

Esta actividad comprende las tareas necesarias para evaluar los modelos desarrollados, en base principalmente a las métricas necesarias, comprensión y utilidad, lo cual es crucial para la aplicación adecuada de las técnicas machine learning o de aprendizaje profundo a la realidad.

Por cada modelo desarrollado se deberá realizar la evaluación mediante métodos como “Model-agnostic”, especificando las medidas y pruebas. La evaluación del modelo implica el cálculo de distintas medidas de performance a partir de indicadores o gráficamente, asimismo se deberá realizar pruebas de stress del modelo desarrollado, lo que permitirá garantizar la calidad y la eficacia del modelo en la resolución del problema planteado minimizando los posibles sesgos.



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

6.7. Puesta en producción y monitoreo de la solución analítica

Esta actividad comprende un proceso de canalización de datos usados en el desarrollo del modelo, adicionalmente se debe garantizar el empaquetamiento de las bibliotecas y dependencias propias del modelo desarrollado en un servidor y/o contenedor.

Esta canalización debe probarse con el lenguaje de la aplicación y validarlo para permitir la implementación en el entorno de Big Data.

El consultor deberá plantear las métricas idóneas de monitoreo de la solución analítica a fin de garantizar la continuidad a través del tiempo de la predictibilidad del modelo.

6.8. Otras actividades

- En los casos que se requiera, deberá analizar la retroalimentación para ajustar el modelo y mejorar su precisión y utilidad.
- Automatizar, a requerimiento, algunos o todos los pasos de la evaluación del modelo y de la recopilación de retroalimentación.
- De acuerdo con lo que requiera, efectuar el ajuste y la reimplementación del modelo para acelerar el proceso de actualización del mismo, con el fin de obtener mejores resultados.
- Dado que el flujo de la metodología para la ciencia de datos ilustra la naturaleza iterativa del proceso de resolución de problemas, el consultor especializado podría volver frecuentemente a etapas previas para realizar ajustes a medida que van aprendiendo más sobre los datos y el modelado.
- 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 entregadas a la DIA a través de correo electrónico o el depósito en el espacio virtual que disponga.

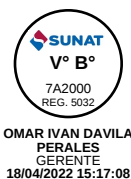
En todos los casos, como parte del objetivo específico de participación cruzada, en el desarrollo de todas las actividades, el consultor especializado podrá participar también asesorando en la creación de variables/indicadores, en el modelado o en la evaluación de los modelos priorizados, desarrollados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

- a. Informes mensuales detallando las actividades realizadas, resultados, recomendaciones y lecciones aprendidas.



- b. 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 Estadística y/o Matemática y/o Investigación Operativa y/o Ingeniería Informática y/o Ingeniería de Sistemas y/o Ciencias de la Computación y/o Electrónica y/o Economía y/o Ingeniería Económica y/o afines
02	Estudios adicionales	De preferencia contar con estudios de Maestría y/o Diplomados en Inteligencia Artificial y/o en Data Science, estudios de Estadística y/o Matemática y/o Informática y/o Sistemas y/o Ciencias de la Computación y/o afines. Se otorgará puntaje adicional.
03	Experiencia general	Experiencia general, al menos cinco (05) años en entidades del sector público y/o privado.
04	Experiencia específica	Experiencia específica, al menos tres (03) años como: Científico de Datos y/o Analista de Datos y/o Especialista de Datos y/o afines en el sector público y/o privado, acreditada mediante: – Constancias de prestación de servicio. – Certificado de trabajo. – Orden de servicio o contrato con su respectiva conformidad.
05	Conocimientos	– Lenguajes de programación: R y/o Python y/o Scala para el tratamiento de grandes volúmenes de datos usando herramientas como Apache Hadoop y/o Spark; y/o – Explotación de datos desde Data Lakes y Data Warehouse; y/o – Conocimiento de Estadística y/o Matemática y/o minería de datos y/o simulación y/u optimización; y/o – Desarrollo de proyectos analíticos construyendo modelos de data science y/o machine learning y/o inteligencia artificial con grandes sets de data estructurada y no estructurada. La forma de acreditar los conocimientos será mediante Declaración Jurada en la Hoja de Vida del consultor.
06	Entrevista	Se evaluará: – Dominio temático. – Capacidad analítica.



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

		– Capacidad para trabajar en equipo. – Iniciativa.
--	--	---

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 de los consultores contratados estarán a cargo del Coordinador Técnico de la UEMSI, previa opinión favorable de la DIA de la GAIR de la INER.

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

El costo total de los servicios asciende a **S/ 1 080 000.00** (un millón ochenta mil y 00/100 soles) por los seis (06) consultores, 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.

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



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

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

Los consultores se obligan 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, no podrán publicar las recomendaciones formuladas en el curso de la contratación, o como resultado de la prestación de los servicios. En caso el consultor 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