

TÉRMINOS DE REFERENCIA

CONTRATACIÓN DE CONSULTORES ESPECIALIZADOS EN DESARROLLO DE MODELOS ANALÍTICOS AVANZADOS DE CIENCIA DE DATOS

CONTRATO DE PRÉSTAMO N°4725/OC-PE	
Componente	Componente 3: Mejora de la Inteligencia fiscal y de la gestión del riesgo de la información tributaria y aduanera
Sub-Componente	Gestión del Riesgo mediante modelos analíticos predictivos y de ciencia de datos
Plazo	12 meses y 10 días para la entrega del informe final previsto en esta consultoría.

1. ANTECEDENTES

El 15 de febrero de 2019 la República del Perú suscribió con el BID el Contrato de Préstamo N° 4725/OC-PE, para financiar en la SUNAT el proyecto de inversión pública “Mejoramiento de los Servicios de Recaudación Tributaria y Aduanera a través de la Transformación Digital”, cuya ejecución está a cargo de la UEMSI.

Este proyecto de inversión pública tiene los siguientes componentes:

Componente 1. Mejora del modelo de gobernanza institucional en SUNAT.

Componente 2. Mejora del control y cumplimiento tributario y aduanero.

Componente 3. Mejora de la inteligencia fiscal y de la gestión del riesgo de la información tributaria y aduanera.

El componente 1, comprende acciones para una adecuada coordinación entre tributos internos y aduanas, especialmente con el uso de innovación tecnológica, tales como: a) modelo de gestión por procesos orientada a resultados, b) mejora de la estructura organizacional de SUNAT con integración tecnológica y c) nuevo modelo de capacidades de recursos humanos.

El componente 2, comprende acciones para lograr la integración de los procesos internos de la SUNAT, a través de la innovación tecnológica, tales como: a) modernización e integración de los procesos y sistemas de tributos internos y aduanas y b) innovación en los procesos de trazabilidad y control aduanero.

El componente 3, comprende acciones para fortalecer la capacidad analítica de la SUNAT, tales como: a) Gobernanza de la información de la SUNAT, b) **Gestión del Riesgo mediante modelos analíticos predictivos y de ciencia de datos** y c) ciberseguridad en el manejo de la información tributaria y aduanera.

En el marco del subcomponente b) se ejecutará la consultoría en mención.

Con Resolución de Superintendencia **RS N.º 000232-2020/SUNAT** en su anexo 5, se incorpora al portafolio de programas de la SUNAT al **Programa Fronteras SMART** y sus 2 Proyectos componentes: **(1) Gestión de Control e Inteligencia y (2) Puertos, Aeropuertos y Fronteras Terrestres Inteligentes**. En el marco del Proyecto 1 se ejecutará la consultoría en mención.

2. MARCO DE REFERENCIA

La consultoría deberá tomar en consideración lo siguiente:

- El Contrato de Préstamo N° 4725/OC-PE.
- Proyecto de Inversión Pública N° 2430225 “Mejoramiento de los Servicios de Recaudación Tributaria y Aduanera a Través de la Transformación Digital”.
- Proyecto Gestión de Control e Inteligencia del Programa Fronteras Inteligentes incorporado al portafolio de programas de la SUNAT mediante Resolución de Superintendencia RS N.º 000232-2020/SUNAT.
- Reglamento de Organización y Funciones de la SUNAT.
- Modelo de Gobernabilidad de Programas y Proyectos – MGPP v.3.2
- Metodología de Desarrollo de Modelos de Minería y otras normas internas de la SUNAT en lo que corresponda.
- Normas emitidas por la Presidencia del Consejo de Ministros del Gobierno Central relacionadas a la Transformación Digital del Estado Peruano, en lo que corresponda

3. OBJETIVO

El objetivo es contratar hasta cinco (5) consultores especializados en ciencia de datos, en el transcurso de los próximo ocho (08) meses, para la planificación, desarrollo e implementación de modelos analíticos avanzados para la identificación de riesgos y formulación de hipótesis de incumplimiento aduanero, con la finalidad de potenciar la detección de incidencias aduaneras y la formulación de tratamientos para mitigar fraudes que respondan a los distintos comportamientos de los viajeros, usuarios u operadores de comercio exterior; incorporando además métodos probabilísticos y predictivos a fin de dar continuidad al desarrollo evolutivo de la ciencia de datos como uno de los elementos o componentes de soporte del modelo de gestión de riesgos del incumplimiento aduanero de la SUNAT.

Entre los objetivos específicos tenemos:

- Posibilitar el conocimiento especializado de analítica avanzada para el diseño y proyección de escenarios futuros, adoptando métodos y herramientas que integren la retroalimentación de los modelos probabilísticos (analíticos, predictivos y/o ciencia de datos) y su ajuste para una mayor precisión, para optimizar el resultado de los tratamientos propuestos.
- Impulsar la transferencia del “Know How” a los profesionales de la División de Gestión de Riesgos Aduaneros (DGRA) mediante asesoría especializada en la planificación, evaluación, planteamiento de requerimientos y logro de objetivos de las iniciativas de analítica avanzada para el soporte de la gestión de riesgos de cumplimiento tributario aduanero.
- Organizar el trabajo para implementar de manera progresiva una línea de producción oportuna con el fin de operativizar los modelos.
- Incorporar el uso de métodos y técnicas complejas de analítica avanzada, así como en la proyección y desarrollo de modelos probabilísticos y predictivos proponiendo las herramientas y plataformas de información más avanzadas que optimicen los resultados del equipo de trabajo de profesionales de la DGRA.
- Incorporar como forma de trabajo la participación cruzada de especialistas en las diferentes etapas del desarrollo de la metodología de modelos avanzados de ciencia de datos, como en la creación de

variables/indicadores para identificar y medir los incumplimientos aduaneros.

4. JUSTIFICACIÓN

La Superintendencia Nacional Adjunta de Aduanas (SNAA) viene evaluando el desarrollo e implementación de iniciativas para la mejora de la gestión de riesgos y control dentro del marco de las facultades del control aduanero, con el objetivo de mejorar la capacidad de la Aduana para detectar y corregir el incumplimiento aduanero, para los diversos procesos aduaneros en las diversas etapas: a través de la incorporación de nuevas herramientas tecnológicas para la captura de información durante toda la cadena de la distribución física de mercancías, en el ingreso y salida del país, lo cual proporcionará un gran volumen de datos para gestionar la información, evaluación de riesgos e implementación de acciones de control.

Una de las iniciativas previstas en la SNAA es la implementación de un “Nuevo Sistema de Control con Inteligencia Artificial” en el marco del proyecto Gestión de Control e Inteligencia, el cual tiene como finalidad desarrollar un nuevo sistema de selección al control aduanero, a fin de incrementar el nivel de incidencia sustancial en el canal rojo, en el régimen de importación para el consumo, considerando en la medida de lo posible, un nuevo ambiente para el procesamiento de Bigdata y la formulación de modelos de minería, empleando técnicas de machine learning que permitan predecir la comisión de nuevos tipos de fraude en las transacciones riesgosas.

Lo expuesto contribuirá con el objetivo estratégico institucional “OEI.03 Reducir el fraude Aduanero” y la acción estratégica: AEI.03.01 Mercancía seleccionada a control concurrente.

La SNAA asimismo tiene planificado el desarrollo e implementación de modelos analíticos tanto para el control concurrente como para el control previo y posterior, utilizando nuevas técnicas de análisis provenientes de la ciencia de datos a fin de disponer de herramientas de información y selección que permitan establecer estrategias de tratamiento de riesgos diferenciadas de acuerdo al perfil de riesgo del operador de comercio exterior, buscando aplicar acciones de control o facilitación proporcionales, equitativos y efectivos que promuevan el cumplimiento voluntario.

En esa línea, mientras los especialistas de aduana incorporados en el año pasado al aprobar el curso aduanero tributario adquieren la experiencia necesaria, 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 aduanero, con la finalidad potenciar la detección de riesgos de incumplimientos y la formulación de tratamientos que respondan a los distintos comportamientos.

5. ALCANCE DEL TRABAJO Y ACTIVIDADES A DESARROLLAR

Planificar, desarrollar e implementar algoritmos avanzados de machine learning y métodos complejos de análisis exploratorio, identificando atipicidades, cambios bruscos de comportamiento de empresas (monto declarado, giro, entre otros),

aplicando segmentación descriptiva y modelado probabilístico y predictivo, referidos a la identificación de riesgos aduaneros con la finalidad potenciar la detección y formulación de tratamientos, proporcionando soporte a la gestión de riesgos en la Aduana peruana.

6. ACTIVIDADES A DESARROLLAR

6.1. Planificación relativa a los modelos

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 las 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 a 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, precisando las tareas, plazos y recursos.

6.2. Comprensión del negocio y enfoque analítico

Durante esta actividad, se deberán identificar los atributos que se vinculan a los objetivos del modelo, 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. 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. Para el caso de los modelos predictivos, se deberá determinar las fuentes de información requeridas y disponibles para la obtención de las variables objetivo. Asimismo, se definen los métodos de extracción y recopilación de estos datos. 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 pueda 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. 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 minería de datos más apropiada de acuerdo con la solución planteada al problema y luego se construye los modelos analíticos. 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 para encontrar el mejor modelo para las variables disponibles.

6.6. Evaluación del modelo

Esta actividad comprende las tareas necesarias para evaluar los modelos obtenidos, en el ambiente de trabajo proporcionado por la administración, en base principalmente al nivel de precisión, comprensión y utilidad, lo cual es crucial para la aplicación adecuada de las técnicas de minería de datos a la realidad. Se deberá realizar primero, el diseño de la evaluación especificando las medidas y pruebas y luego se realiza la ejecución de la evaluación. La evaluación del modelo implica el cálculo de varias medidas de diagnóstico y de otros resultados, como tablas y gráficos, lo que permitirá interpretar la calidad y la eficacia del modelo en la resolución del problema, así como cuantificar los beneficios esperados para la administración aduanera.

6.7. 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 se requiera, efectuar el ajuste y la reimplementación del modelo para acelerar su proceso de actualización con el fin de obtener mejores resultados.
- Participar en la fase de despliegue y estabilización para la operativización de los modelos, proponiendo las mejoras pertinentes y la posible solución a los incidentes identificados. Al respecto, cabe precisar que esta etapa se dará cuando el modelo de minería requiera llevarse a producción o cuando esté implantado en el sistema operacional de SUNAT.
- 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.

- 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 proceso de desarrollo por otros profesionales, incluso respecto de aquellos ya implantados.
- 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.

6.8 Reporte de resultados

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 al líder normativo del proyecto Gestión de Control e Inteligencia a través de correo electrónico o en el depósito en el espacio virtual que se disponga, para su revisión y remisión a la División de Gestión de Riesgos Aduaneros (DGRA) en su calidad de área usuaria a fin de que evalúe la conformidad de los informes presentados por los consultores.

7. ENTREGABLES

El servicio de consultoría tendrá un plazo de duración de doce (12) meses y 10 días calendario, plazo que incluye la entrega de los informes siguientes:

- Informes mensuales detallando las actividades realizadas, resultados y recomendaciones, presentados dentro de los cinco (5) días calendario del mes siguiente.
- 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 DGRA del código o comandos estructurados para la ejecución de los modelos analíticos.
 - Resumen ejecutivo de principales logros y modelos analíticos desarrollados.
 - Conclusiones, recomendaciones y lecciones aprendidas.

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 a través de Mesa de Partes en el horario de 08:30am a 04:30pm, los mismos deberán estar dirigidos a la Coordinación Técnica de la UEMSI.

Las actividades se desarrollarán de manera presencial conforme a lo indicado en el numeral 9 del presente documento.

8. PERFIL DE LOS CONSULTORES Y COMPETENCIAS

Formación Académica Mínima	Egresado, Bachiller, titulado o licenciado en Estadística, Matemáticas, Investigación Operativa, Ingeniería Informática, Ingeniería de Sistemas, Ciencias de la Computación y/o afines.
-------------------------------	---

Estudios adicionales	De preferencia deberá contar con estudios de Maestría y/o Diplomados en Business Intelligence (BI), y/o en Data Science (DS), Y opcionalmente, con estudios de Estadística, Matemáticas, Informática, Sistemas, Ciencias de la Computación y/o afines
Experiencia general	Experiencia general, al menos cinco (05) años en entidades del sector público y/o privado.
Experiencia específica	Experiencia específica, al menos tres (03) años como Científico de Datos o Analista de Datos o Especialista de Datos 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.
Conocimientos	- Machine Learning - Lenguajes de programación: Python y/o R. Y opcionalmente: SQL, y/o C++, y/o Javascript. - Manejo de bases de datos: Teradata y/o Oracle y/o MongoDB. Y opcionalmente: SQL Server, y/o PostgreSQL, y/o MySQL y/o informix y/o afines. - De preferencia contar con conocimiento en herramientas: SAS, y/o SPSS modeler. Acreditada mediante certificados o constancias de capacitación
Entrevistas	Criterios de evaluación: - Dominio temático en la predicción de riesgos o fraudes - Capacidad de trabajo en equipo

9. LUGAR DE LA CONSULTORÍA

Los servicios de consultoría se realizarán de manera presencial en la Av. Gamarra 680 Chucuito de la provincia Constitucional del Callao Departamento de Lima, Perú. La SUNAT, brindará las facilidades que se requieran para el desarrollo de la consultoría.

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° 026-2020 en el marco de la emergencia sanitaria por el COVID-19.

10. UNIDAD ORGANIZACIONAL RESPONSABLE DE OTORGAR LA CONFORMIDAD PREVIA AL PAGO:

El área responsable de otorgar la conformidad será la Coordinación Técnica de la UEMSI, previa opinión favorable de la Gerencia de Gestión de Riesgos e Investigaciones Aduaneras en su calidad de área usuaria y opinión favorable de la Jefatura del Proyecto Gestión de Control e Inteligencia.

Para el pago del informe final se requerirá previamente la No Objeción del BID.

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

El costo total de los servicios asciende a S/.780,000.00 (setecientos ochenta mil y 00/100 soles) por los cinco (05) consultores, importe que está sujeto a la respectiva retención de impuesto a la renta.

El costo mensual para cada consultor de los servicios asciende a S/.13,000.00 (trece 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, desde la fecha de inicio de actividades prevista en el contrato, por un plazo de doce (12) meses calendario y tendrá un plazo de diez (10) días para la entrega del informe final, a efecto de que pueda obtener la aprobación y conformidad correspondiente.

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, contando con la opinión favorable de la Gerencia de Gestión de Riesgos e Investigaciones Aduaneras en su calidad de área usuaria y de la Jefatura del Proyecto Gestión de Control e Inteligencia.

Adicionalmente, previo al último pago, se requerirá de la No Objeción del BID al informe final del consultor.

12. 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 de este. Asimismo, el Consultor no podrá publicar las recomendaciones formuladas en el curso de, o como resultado de la prestación de los servicios. En caso de que 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, salvo autorización expresa de la SUNAT.