

TÉRMINOS DE REFERENCIA

CONTRATACIÓN DE UN CONSULTOR ESPECIALIZADO DS04 EN DESARROLLO DE MODELOS ANALÍTICOS AVANZADOS PARA EL PROCESO EVOLUTIVO DE LA CIENCIA DE DATOS

CONTRATO DE PRESTAMO BID N.º 4725/OC-PE	
Denominación del componente	Mejora de la inteligencia fiscal y la gestión de riesgo tributario y aduanero
Posición	Consultor Especializado DS04
Plazo	10 meses y 10 días calendarios, 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) Proyecto, "Mejoramiento de los Servicios de Recaudación Tributaria y Aduanera a través de la Transformación Digital (PE-L 1239)".
- (d) **CT:** Coordinación Técnica de la Unidad Ejecutora Mejoramiento del Sistema de Información de la SUNAT (UEMSI).
- (e) **UEMSI:** Unidad Ejecutora "Mejoramiento del Sistema de Información de la SUNAT" – MSI
- (f) **MEF:** Ministerio de Economía y Finanzas.
- (g) **INER:** Intendencia Nacional de Estrategias y Riesgos.
- (h) **GAIR:** Gerencia de Análisis de Información y Riesgos.
- (i) **DIA:** División de Información y Análisis.
- (j) **INSI:** Intendencia Nacional de Sistemas de Información.
- (k) **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.
- (l) **TEXT MINING:** La minería de textos (Text Mining) es una disciplina englobada dentro de las técnicas de acceso, recuperación y organización de información y consiste en un conjunto de técnicas que nos permiten extraer información relevante y desconocida de manera automática dentro de grandes volúmenes de información textual, normalmente en lenguaje natural y por lo general no estructurada.
- (m) **WEB SCRAPING:** es una técnica utilizada mediante programas de software para extraer información de sitios web. Usualmente, estos programas simulan la navegación de un humano en la World Wide Web ya sea utilizando el protocolo HTTP manualmente, o incrustando un navegador en una aplicación.
- (n) **BI:** Business Intelligence o inteligencia empresarial, se refiere a la utilización de datos en una empresa para facilitar la toma de decisiones dentro de la misma. Es un conjunto de tecnologías, herramientas y procesos que permiten transformar los datos almacenados en información, ésta en conocimiento y este conocimiento es dirigido a un plan o una estrategia comercial / financiera, etc.
- (o) **BIG DATA:** manejo de grandes volúmenes de datos (estructurados y no estructurados) que conducen a tomar decisiones estratégicas.



OMAR IVAN DAVILA
PERALES
GERENTE
09/02/2023 10:58:17



FREDDY MARCIAL
QUEVEDO CAJAS
JEFE DE DIVISION
07/02/2023 11:06:53



CECILIA DEL PILAR
LEON MESTANZA
SUPERVISOR
ENCARGADO
06/02/2023 17:56:38

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.º 4725/OC-PE, para financiar el proyecto de inversión pública “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 PROYECTO, 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 subcomponente “Gestión de riesgos mediante el uso de modelos analítico-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.

Los tratamientos a implementar producto de la aplicación del modelo de ciencia de datos se aplican para cerrar las brechas en cada fase del ciclo de vida de los contribuyentes, y que son las siguientes: 1. Inscripción, 2. Documentación, 3. Registro y Declaración, 4. Pago y 5. Veracidad, lo que incluyen aspectos transversales a las brechas indicadas.

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 Scraping, 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 PUBLICA

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 2018-2022, 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



OMAR IVAN DAVILA
PERALES
GERENTE
09/02/2023 10:58:17



FREDDY MARCIAL
QUEVEDO CAJAS
JEFE DE DIVISION
07/02/2023 11:06:53



CECILIA DEL PILAR
LEON MESTANZA
SUPERVISOR
ENCARGADO
06/02/2023 17:56:38

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 incumplimientos y la formulación de tratamientos que respondan a los distintos comportamientos.

3. OBJETIVO

3.1. OBJETIVO GENERAL

Contratar los servicios de un (01) Consultor Especializado DS04 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, liderando aquellos vinculados con la imposición directa de las actividades no empresariales y su vinculación con el sistema tributario peruano, con la finalidad de potenciar la detección de riesgos de incumplimientos y la formulación de tratamientos que respondan a los distintos comportamientos; 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 riesgo del incumplimiento tributario de la SUNAT.

3.2. OBJETIVOS ESPECIFICOS

- 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 y su ajuste para una mayor precisión, para optimizar el resultado de los tratamientos propuestos.
- Potenciar el equipo de ciencia de datos 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.
- Organizar el trabajo para implementar de manera progresiva una línea de producción con el fin de operativizar los modelos desarrollados a través de la INSI, e institucionalizar las iniciativas de analítica avanzada.
- 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 trabajo del equipo de ciencia de datos.
- 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 las brechas de incumplimiento en cada ciclo de vida del contribuyente.

4. MARCO DE REFERENCIA

La consultoría deberá tomar en consideración lo siguiente:



OMAR IVAN DAVILA
PERALES
GERENTE
09/02/2023 10:58:17



FREDDY MARCIAL
QUEVEDO CAJAS
JEFE DE DIVISION
07/02/2023 11:06:53



CECILIA DEL PILAR
LEON MESTANZA
SUPERVISOR
ENCARGADO
06/02/2023 17:56:38

- El Contrato de Préstamo N.º 4725/OC-PE.
- Reglamento de Organización y Funciones 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 y modelado probabilístico y predictivo, 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 a la gestión de riesgos de incumplimiento tributario en la SUNAT, liderando aquellos vinculados con la imposición directa de las actividades no empresariales y su vinculación con el sistema tributario peruano.

6. ACTIVIDADES PARA DESARROLLAR

6.1. Planificación relativa a los proyectos

Durante esta actividad, la consultoría especializada deberá alinear expectativas con los objetivos de los proyectos 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 que se busca alcanzar con cada uno de los modelos a elaborar.
- Alcance, indicando los segmentos, dependencias y/o áreas en los que serán aplicados los modelos a desarrollar.
- Justificación, identificando los incumplimientos o riesgos que el modelo va a evaluar.

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



OMAR IVAN DAVILA
PERALES
GERENTE
09/02/2023 10:58:17



FREDDY MARCIAL
QUEVEDO CAJAS
JEFE DE DIVISION
07/02/2023 11:06:53



CECILIA DEL PILAR
LEON MESTANZA
SUPERVISOR
ENCARGADO
06/02/2023 17:56:38

técnicas estadísticas y de aprendizaje automático, para que la organización pueda identificar las más adecuadas para el resultado deseado.

6.3. Requisitos, recopilación y 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 método 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, 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 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 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 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.



OMAR IVAN DAVILA
PERALES
GERENTE
09/02/2023 10:58:17



FREDDY MARCIAL
QUEVEDO CAJAS
JEFE DE DIVISION
07/02/2023 11:06:53



CECILIA DEL PILAR
LEON MESTANZA
SUPERVISOR
ENCARGADO
06/02/2023 17:56:38

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.

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 requiera, efectuar el ajuste y la reimplementación del modelo para acelerar el proceso de actualización de este, con el fin de obtener mejores resultados.
- Participar en la fase de despliegue y estabilización para la operativización de los modelos a cargo de la INSI. 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.
- 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 diez (10) 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.



OMAR IVAN DAVILA
PERALES
GERENTE
09/02/2023 10:58:17



FREDDY MARCIAL
QUEVEDO CAJAS
JEFE DE DIVISION
07/02/2023 11:06:53



CECILIA DEL PILAR
LEON MESTANZA
SUPERVISOR
ENCARGADO
06/02/2023 17:56:38

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.º	Requisito	Criterio
01	Formación académica mínima	Egresado y/o Bachiller en Estadística, Matemáticas, Física, Ingeniería Informática, Ingeniería de Sistemas, Ciencias de la Computación y/o afines.
02	Estudios adicionales	De preferencia deberá contar con estudios de Maestría y/o Diplomados en Business Intelligence (BI), y/o en Data Science (DS), Estadística, Matemáticas, Física, Informática, Sistemas, Ciencias de la Computación y/o afines
03	Experiencia general	Experiencia general en los últimos cuatro (04) años en entidades del sector público o privado.
04	Experiencia específica	Experiencia específica en los últimos dos (02) años como Científico de Datos o Analista/Especialista de Datos en el sector público y/o privado.
05	Conocimientos	<ul style="list-style-type: none"> - Minería de Datos - Lenguajes de programación: Python y/o R, y/o SQL, y/o C++, y/o Javascript y/o Scala. - Manejo de bases de datos: Teradata y/o Oracle, y/o SQL Server, y/o PostgreSQL, y/o MySQL, y/o MongoDB, y/o afines. - De preferencia contar con conocimiento en herramientas: SAS, y/o SPSS, y/o Stata, y/o herramientas open source. <p>La forma de acreditar los conocimientos será mediante Declaración Jurada en la Hoja de Vida del consultor</p>
06	Entrevistas	<p>Se evaluará:</p> <ul style="list-style-type: none"> - Dominio Temático. - Capacidad Analítica. - Capacidad para trabajar en equipo. - Iniciativa.



OMAR IVAN DAVILA
PERALES
GERENTE
09/02/2023 10:58:17



FREDDY MARCIAL
QUEVEDO CAJAS
JEFE DE DIVISION
07/02/2023 11:06:53



CECILIA DEL PILAR
LEON MESTANZA
SUPERVISOR
ENCARGADO
06/02/2023 17:56:38

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. 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 (MPV) y/o en las instalaciones de la SUNAT (Av. Garcilaso de la Vega N.º 1472 – Lima 1), a través de Mesa de Partes en el horario de 08:30 a.m. a 04:30 p.m. o en el horario de 08:30 a.m. a 11:59 p.m. en MPV, 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 LA 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/ 100 000.00 (cien 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 diez (10) meses y diez (10) días calendarios, 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.



OMAR IVAN DAVILA PERALES
GERENTE
09/02/2023 10:58:17

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.



FREDDY MARCIAL QUEVEDO CAJAS
JEFE DE DIVISION
07/02/2023 11:06:53

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 de este. 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.

Firmado por:

OMAR IVAN DAVILA PERALES

Gerente

Gerencia de Análisis de Información y Riesgos

Intendencia Nacional de Estrategias y Riesgos



CECILIA DEL PILAR LEON MESTANZA
SUPERVISOR ENCARGADO
06/02/2023 17:56:38