

TÉRMINOS DE REFERENCIA

CONTRATACIÓN DE EXPERTO EN PROCESAMIENTO DE LENGUAJE NATURAL (NLP) PARA EL DESARROLLO DE NUEVOS COMPONENTES DEL SISTEMA DE SELECCIÓN A CONTROL ORDINARIO.

CONTRATO DE PRÉSTAMO N°4725/OC-PE	
Denominación del Componente	Componente 2. Mejora del control y cumplimiento tributario y aduanero
Sub-Componente	2.1 Modernización e Integración de los Procesos y Sistemas de Tributos Internos y Aduanas.
Plazo	13 meses y 10 días calendarios

1. DENOMINACIÓN DE LA CONTRATACIÓN

Contratación de experto en procesamiento de lenguaje natural (NLP) para el desarrollo de variables para un nuevo componente del sistema de selección a control ordinario, que se requiere para reemplazar uno de los componentes legacy.

2. DEFINICIONES

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

- BID: Banco Interamericano de Desarrollo
- DCA: División de Clasificación Arancelaria
- DGRA: División de Gestión de Riesgos Aduaneros
- DPC: División de Procesos de Control
- DV: División de Valoración
- GGRIA: Gerencia de Gestión de Riesgos e Investigaciones Aduaneras
- GPAFC: Gerencia de Procesos de Atención Fronteriza y Control
- GTA: Gerencia de Técnica Aduanera
- NLP: Procesamiento de Lenguaje Natural
- PGCI: Proyecto Componente Gestión de Control e Inteligencia
- SIVEP: Sistema de Verificación de Precios
- SNAA: Superintendencia Nacional Adjunta de Aduanas
- SUNAT: Superintendencia Nacional de Aduanas y de Administración Tributaria
- UEMSI: Unidad Ejecutora "Mejoramiento del Sistema de Información de la SUNAT"

3. ANTECEDENTES

El 15.02.2019, 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 BID 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.

Al respecto, el componente 2 comprende el subcomponente 2.1 Modernización e integración de los procesos y sistemas de tributos internos y aduanas, y el subcomponente 2.2 Innovación en los procesos de trazabilidad y control aduanero.

El subcomponente 2.1 Modernización e integración de los procesos y sistemas de tributos internos y aduanas, antes citado, comprende el literal i. Modernización e integración de los procesos y sistemas de tributos internos y aduanas, que incluye el inciso (ii) modernización de sistemas legacy mediante la eliminación de la arquitectura de bases de datos descentralizadas, a fin de reducir los riesgos operativos por obsolescencia tecnológica.

El 30.12.2020, mediante la Resolución de Superintendencia N° 000234-2020/SUNAT, se incorporó al portafolio de programas de la SUNAT, el Programa Fronteras SMART y sus 2 Proyectos componentes: (1) Puertos, Aeropuertos y Fronteras Terrestres Inteligentes (PAFTI) y (2) Gestión de Control e Inteligencia (en adelante PGCi).

Dentro de los alcances de este último proyecto se encuentra previsto el entregable 06: "Nuevo Sistema de selección con inteligencia artificial", en el marco del cual se ha realizado el análisis y se ha identificado que en el sistema de selección a control ordinario hay un componente legacy que requiere ser reemplazado con un nuevo componente utilizando técnicas modernas de machine learning para su optimización.

4. OBJETIVO

4.1 OBJETIVO GENERAL

Contratar a un experto en NLP para el desarrollo de las variables "índice de correspondencia de clasificación arancelaria" y "diferencia del valor unitario declarado versus referencias SIVEP", con el fin de que estas sean integradas en un nuevo componente del sistema de selección a control ordinario.

4.2 OBJETIVOS ESPECÍFICOS

- Fortalecer el control aduanero y mejorar la detección del incumplimiento aduanero, efectuando el análisis y desarrollo de la variable "índice de correspondencia de clasificación arancelaria", considerando la información de las descripciones comerciales y la subpartida arancelaria nacional más probable con el fin de optimizar el sistema de selección a control ordinario y reemplazar uno de los componentes legacy.
- Fortalecer el control aduanero y mejorar la detección del incumplimiento aduanero, efectuando el análisis y desarrollo de la variable "diferencia del valor unitario declarado versus referencias SIVEP", considerando la información del valor unitario declarado respecto de las referencias del SIVEP, con el fin de optimizar el sistema de selección a control ordinario y reemplazar uno de los componentes legacy.

5. MARCO DE REFERENCIA

El servicio 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 N° 000234-2020/SUNAT.
- Texto Integrado del Reglamento de Organización y Funciones (ROF) de la SUNAT, aprobado con Resolución de Superintendencia N° 000092-2023/SUNAT.
- Modelo de Gobernabilidad de Programas y Proyectos – MGPP v.3.2.
- Metodología de Desarrollo de Modelos de Minería de Datos (MDMD) – v.1.0.
- Ley de Gobierno Digital, aprobado con Decreto Legislativo N° 1412.
- Plan de Gobierno Digital 2022 – 2024, aprobado con Resolución de Superintendencia N° 054-2022/SUNAT.
- Ley N° 29733, Ley de Protección de Datos Personales.
- Texto Único Ordenado del Código Tributario aprobado mediante Decreto Supremo N° 135-99-EF.
- Ley General de Aduanas aprobada por el Decreto Legislativo N° 1053 y modificatorias.
- Texto Único Ordenado de la Ley N° 27806 - Ley de Transparencia y Acceso a la Información Pública aprobada por Decreto Supremo N° 021-2019-JUS.

6. JUSTIFICACIÓN

Una de las iniciativas previstas en la SNAA es la implementación, en el marco del proyecto Gestión de Control e Inteligencia, de un “Nuevo Sistema de Selección con Inteligencia Artificial”, el cual tiene como finalidad mejorar la selección a control ordinario en el régimen de importación para el consumo, considerando la reformulación y sustitución de los modelos de minería de datos legacy, empleando técnicas modernas de machine learning.

En ese contexto, con la finalidad de fortalecer el control aduanero y mejorar la detección del incumplimiento aduanero se requiere desarrollar la variable “índice de correspondencia de clasificación arancelaria”, utilizando técnicas de procesamiento de lenguaje natural aplicadas a la descripción de la mercancía, así como la variable “diferencia del valor unitario declarado versus referencias SIVEP”, utilizando técnicas de procesamiento de lenguaje natural aplicadas a la descripción de la mercancía, a fin de detectar inconsistencias respecto de lo señalado en el Arancel de Aduanas y de las referencias de valor en el SIVEP, respectivamente. Asimismo, se requiere integrar las citadas variables en un nuevo componente del sistema de selección a control ordinario.

Para este desarrollo, la SUNAT requiere contar con el servicio de un profesional experto con conocimiento especializado y experiencia en técnicas modernas de machine learning en procesamiento de lenguaje natural.

Lo expuesto contribuirá con el objetivo estratégico institucional OEI.01 Mejorar el cumplimiento tributario y aduanero de los administrados y la acción estratégica: AEI.01.04 Modernización y transformación digital de los procesos de la entidad.

7. ALCANCE DEL TRABAJO

Se requiere desarrollar las variables “índice de correspondencia de clasificación arancelaria” y “diferencia del valor unitario declarado versus referencias SIVEP” para el sistema de selección a control ordinario a implantarse en la nueva plataforma MLOps centralizado, con el uso de técnicas modernas de machine learning y procesamiento de lenguaje natural (NLP), obteniendo los siguientes productos:

7.1 Variable “índice de correspondencia de clasificación arancelaria”

Planificar, analizar y desarrollar la variable “índice de correspondencia de clasificación arancelaria”, considerando la identificación de riesgos a partir de la “lectura” de las descripciones comerciales y la subpartida arancelaria nacional más probable, empleando técnicas modernas de machine learning y NLP, con el fin de que dicha variable sea integrada en un nuevo componente del sistema de selección a control ordinario, que se requiere para reemplazar uno de los componentes legacy.

7.2 Variable “diferencia del valor unitario declarado versus referencias SIVEP”

Planificar, analizar y desarrollar la variable “diferencia del valor unitario declarado versus referencias SIVEP”, considerando la identificación de riesgos a partir de la “lectura” de las descripciones comerciales, la subpartida arancelaria nacional más probable y el valor unitario declarado respecto de las referencias del SIVEP, empleando técnicas modernas de machine learning y NLP, con el fin de que dicha variable sea integrada en un nuevo componente del sistema de selección a control ordinario, que se requiere para reemplazar uno de los componentes legacy.

8. ACTIVIDADES A DESARROLLAR

El experto por contratar efectuará las siguientes actividades:

8.1 Planificación del trabajo por realizar

Durante esta actividad el experto coordinará con personal de PGCI, DPC, DCA, DV y DGRA la elaboración del plan de trabajo, debiendo tomar conocimiento de los siguientes aspectos:

- Lineamientos para el desarrollo del plan de trabajo;
- Documentación necesaria para el desarrollo del plan de trabajo;
- Cronograma y descripción de las actividades;
- Descripción del contenido y presentación de los entregables detallados en los numerales del 8.2 al 8.14.

El PGCI, por encargo del área usuaria, efectuará la convocatoria, centralizará los resultados de las reuniones sostenidas y elaborará el acta de las reuniones realizadas.

Actividades relacionadas con el producto señalado en el numeral 7.1: Variable “índice de correspondencia de clasificación arancelaria”

8.2 Creación de la tabla para entrenamiento del modelo texto clasificador

- Reunión inicial con personal de PGCI, DPC, DCA y DGRA para tomar conocimiento sobre los lineamientos para el desarrollo de la actividad.
- Reuniones de avances que permitan verificar las actividades desarrolladas, de acuerdo a la frecuencia que será definida por el PGCI, por encargo del área usuaria.
- Análisis y desarrollo de la tabla para entrenamiento del modelo texto clasificador.
- Transferencia de conocimientos a través de charlas y/o presentaciones y/o talleres.
- Elaboración de manuales de uso y/o instructivos.

- Elaboración del informe de resultados.
- Exposición de resultados en la reunión de cierre de la actividad.

El PGCI, por encargo del área usuaria, efectuará la convocatoria, centralizará los resultados de las reuniones sostenidas y elaborará el acta de las reuniones realizadas.

8.3 Desarrollo del preprocesamiento de los datos del modelo texto clasificador

- Reunión inicial con personal de PGCI, DPC, DCA y DGRA para tomar conocimiento sobre los lineamientos para el desarrollo de la actividad.
- Reuniones de avances que permitan verificar las actividades desarrolladas, de acuerdo a la frecuencia que será definida por el PGCI, por encargo del área usuaria.
- Análisis y desarrollo del preprocesamiento de los datos del modelo texto clasificador.
- Transferencia de conocimientos a través de charlas y/o presentaciones y/o talleres.
- Elaboración de manuales de uso y/o instructivos.
- Elaboración del informe de resultados.
- Exposición de resultados en la reunión de cierre de la actividad.

El PGCI, por encargo del área usuaria, efectuará la convocatoria, centralizará los resultados de las reuniones sostenidas y elaborará el acta de las reuniones realizadas.

8.4 Desarrollo del entrenamiento del modelo texto clasificador con técnica 1

- Reunión inicial con personal de PGCI, DPC, DCA y DGRA para tomar conocimiento sobre los lineamientos para el desarrollo de la actividad.
- Reuniones de avances que permitan verificar las actividades desarrolladas, de acuerdo a la frecuencia que será definida por el PGCI, por encargo del área usuaria.
- Análisis y desarrollo del entrenamiento del modelo texto clasificador con técnica 1 de NLP.
- Transferencia de conocimientos a través de charlas y/o presentaciones y/o talleres.
- Elaboración de manuales de uso y/o instructivos.
- Elaboración del informe de resultados.
- Exposición de resultados en la reunión de cierre de la actividad.

El PGCI, por encargo del área usuaria, efectuará la convocatoria, centralizará los resultados de las reuniones sostenidas y elaborará el acta de las reuniones realizadas.

8.5 Desarrollo del entrenamiento del modelo texto clasificador con técnica 2

- Reunión inicial con personal de PGCI, DPC, DCA y DGRA para tomar conocimiento sobre los lineamientos para el desarrollo de la actividad.
- Reuniones de avances que permitan verificar las actividades desarrolladas, de acuerdo a la frecuencia que será definida por el PGCI, por encargo del área usuaria.
- Análisis y desarrollo del entrenamiento del modelo texto clasificador con técnica 2 de NLP.

- Transferencia de conocimientos a través de charlas y/o presentaciones y/o talleres.
- Elaboración de manuales de uso y/o instructivos.
- Elaboración del informe de resultados.
- Exposición de resultados en la reunión de cierre de la actividad.

El PGCI, por encargo del área usuaria, efectuará la convocatoria, centralizará los resultados de las reuniones sostenidas y elaborará el acta de las reuniones realizadas.

8.6 Desarrollo del entrenamiento del modelo texto clasificador con técnica 3 y comparación

- Reunión inicial con personal de PGCI, DPC, DCA y DGRA para tomar conocimiento sobre los lineamientos para el desarrollo de la actividad.
- Reuniones de avances que permitan verificar las actividades desarrolladas, de acuerdo a la frecuencia que será definida por el PGCI, por encargo del área usuaria.
- Análisis y desarrollo del entrenamiento del modelo texto clasificador con técnica 3 de NLP y comparación con los modelos que consideran la técnica 1 y 2 de NLP.
- Transferencia de conocimientos a través de charlas y/o presentaciones y/o talleres.
- Elaboración de manuales de uso y/o instructivos.
- Elaboración del informe de resultados.
- Exposición de resultados en la reunión de cierre de la actividad.

El PGCI, por encargo del área usuaria, efectuará la convocatoria, centralizará los resultados de las reuniones sostenidas y elaborará el acta de las reuniones realizadas.

8.7 Desarrollo de los ajustes al modelo texto clasificador

- Reunión inicial con personal de PGCI, DPC, DCA y DGRA para tomar conocimiento sobre los lineamientos para el desarrollo de la actividad.
- Reuniones de avances que permitan verificar las actividades desarrolladas, de acuerdo a la frecuencia que será definida por el PGCI, por encargo del área usuaria.
- Análisis y desarrollo de los ajustes al modelo texto clasificador en función a la retroalimentación.
- Transferencia de conocimientos a través de charlas y/o presentaciones y/o talleres.
- Elaboración de manuales de uso y/o instructivos.
- Elaboración del informe de resultados.
- Exposición de resultados en la reunión de cierre de la actividad.

El PGCI, por encargo del área usuaria, efectuará la convocatoria, centralizará los resultados de las reuniones sostenidas y elaborará el acta de las reuniones realizadas.

8.8 Elaboración de informe final sobre modelamiento del modelo texto clasificador, documentación para la implementación de la variable “índice de correspondencia de clasificación arancelaria” y acompañamiento para su implementación.

- Reunión inicial con personal de PGCI, DPC, GTA, DCA y DGRA para tomar conocimiento sobre los lineamientos para el desarrollo de la actividad.
- Reuniones de avances que permitan verificar las actividades desarrolladas, de acuerdo a la frecuencia que será definida por el PGCI, por encargo del área usuaria.
- Análisis sobre el modelamiento del modelo texto clasificador, elaboración de la documentación para la implementación de la variable “Índice de correspondencia de clasificación arancelaria” y acompañamiento a la DGRA para su implementación e integración en los nuevos componentes del sistema de selección a control ordinario en la plataforma MLOps.
- Transferencia de conocimientos a través de charlas y/o presentaciones y/o talleres.
- Elaboración de manuales de uso y/o instructivos.
- Elaboración del informe de resultados.
- Exposición de resultados en la reunión de cierre de la actividad.

El PGCI, por encargo del área usuaria, efectuará la convocatoria, centralizará los resultados de las reuniones sostenidas y elaborará el acta de las reuniones realizadas.

Actividades relacionadas con el producto señalado en el numeral 7.2: Variable “diferencia del valor unitario declarado versus referencias SIVEP”

8.9 Creación de la tabla para entrenamiento del modelo referencias SIVEP

- Reunión inicial con personal de PGCI, DPC, DV y DGRA para tomar conocimiento sobre los lineamientos para el desarrollo de la actividad.
- Reuniones de avances que permitan verificar las actividades desarrolladas, de acuerdo a la frecuencia que será definida por el PGCI, por encargo del área usuaria.
- Análisis y desarrollo de la tabla para entrenamiento del modelo referencias SIVEP.
- Transferencia de conocimientos a través de charlas y/o presentaciones y/o talleres.
- Elaboración de manuales de uso y/o instructivos.
- Elaboración del informe de resultados.
- Exposición de resultados en la reunión de cierre de la actividad.

El PGCI, por encargo del área usuaria, efectuará la convocatoria, centralizará los resultados de las reuniones sostenidas y elaborará el acta de las reuniones realizadas.

8.10 Desarrollo del preprocesamiento de los datos del modelo referencias SIVEP

- Reunión inicial con personal de la PGCI, DPC, DV y DGRA para tomar conocimiento sobre los lineamientos para el desarrollo de la actividad.
- Reuniones de avances que permitan verificar las actividades desarrolladas, de acuerdo a la frecuencia que será definida por el PGCI, por encargo del área usuaria.
- Análisis y desarrollo del preprocesamiento de los datos del modelo referencias SIVEP.
- Transferencia de conocimientos a través de charlas y/o presentaciones y/o talleres.
- Elaboración de manuales de uso y/o instructivos.
- Elaboración del informe de resultados.

- Exposición de resultados en la reunión de cierre de la actividad.

El PGCI, por encargo del área usuaria, efectuará la convocatoria, centralizará los resultados de las reuniones sostenidas y elaborará el acta de las reuniones realizadas.

8.11 Desarrollo del entrenamiento del modelo referencias SIVEP con técnica 1

- Reunión inicial con personal de PGCI, DPC, DV y DGRA para tomar conocimiento sobre los lineamientos para el desarrollo de la actividad.
- Reuniones de avances que permitan verificar las actividades desarrolladas, de acuerdo a la frecuencia que será definida por el PGCI, por encargo del área usuaria.
- Análisis y desarrollo del entrenamiento del modelo referencias SIVEP con técnica 1 de NLP.
- Transferencia de conocimientos a través de charlas y/o presentaciones y/o talleres.
- Elaboración de manuales de uso y/o instructivos.
- Elaboración del informe de resultados.
- Exposición de resultados en la reunión de cierre de la actividad.

El PGCI, por encargo del área usuaria, efectuará la convocatoria, centralizará los resultados de las reuniones sostenidas y elaborará el acta de las reuniones realizadas.

8.12 Desarrollo del entrenamiento del modelo referencias SIVEP con técnica 2 y comparación

- Reunión inicial con personal de PGCI, DPC, DV y DGRA para tomar conocimiento sobre los lineamientos para el desarrollo de la actividad.
- Reuniones de avances que permitan verificar las actividades desarrolladas, de acuerdo a la frecuencia que será definida por el PGCI, por encargo del área usuaria.
- Análisis y desarrollo del entrenamiento del modelo referencias SIVEP con técnica 2 de NLP y comparación con los modelos que consideran la técnica 1 de NLP.
- Transferencia de conocimientos a través de charlas y/o presentaciones y/o talleres.
- Elaboración de manuales de uso y/o instructivos.
- Elaboración del informe de resultados.
- Exposición de resultados en la reunión de cierre de la actividad.

El PGCI, por encargo del área usuaria, efectuará la convocatoria, centralizará los resultados de las reuniones sostenidas y elaborará el acta de las reuniones realizadas.

8.13 Desarrollo de los ajustes al modelo referencias SIVEP

- Reunión inicial con personal de PGCI, DPC, DV y DGRA para tomar conocimiento sobre los lineamientos para el desarrollo de la actividad.
- Reuniones de avances que permitan verificar las actividades desarrolladas, de acuerdo a la frecuencia que será definida por el PGCI, por encargo del área usuaria.
- Análisis y desarrollo de los ajustes al modelo referencias SIVEP en función a la retroalimentación.

- Transferencia de conocimientos a través de charlas y/o presentaciones y/o talleres.
- Elaboración de manuales de uso y/o instructivos.
- Elaboración del informe de resultados.
- Exposición de resultados en la reunión de cierre de la actividad.

El PGCI, por encargo del área usuaria, efectuará la convocatoria, centralizará los resultados de las reuniones sostenidas y elaborará el acta de las reuniones realizadas.

8.14 Elaboración del informe final sobre el modelamiento del modelo referencias SIVEP, documentación para la implementación de la variable “diferencia del valor unitario declarado versus referencias SIVEP” y acompañamiento para su implementación.

- Reunión inicial con personal de PGCI, DPC, GTA, DV y DGRA para tomar conocimiento sobre los lineamientos para el desarrollo de la actividad.
- Reuniones de avances que permitan verificar las actividades desarrolladas, de acuerdo a la frecuencia que será definida por el PGCI, por encargo del área usuaria.
- Análisis sobre el modelamiento del modelo referencias SIVEP, elaboración de la documentación para la implementación de la variable “diferencia del valor unitario declarado versus referencias SIVEP” y acompañamiento a la DGRA para su implementación e integración en los nuevos componentes del sistema de selección a control ordinario en la plataforma MLOps.
- Transferencia de conocimientos a través de charlas y/o presentaciones y/o talleres.
- Elaboración de manuales de uso y/o instructivos.
- Elaboración del informe de resultados.
- Exposición de resultados en la reunión de cierre de la actividad.

El PGCI, por encargo del área usuaria, efectuará la convocatoria, centralizará los resultados de las reuniones sostenidas y elaborará el acta de las reuniones realizadas.

9. ENTREGABLES, PLAZOS Y MONTO DE PAGO

Los entregables emitidos por el experto constituyen “información clasificada”; lo que implica que la información contenida en estos es confidencial con reserva tributaria, la cual se encuentra dentro de las excepciones previstas en el artículo 17 del Texto Único Ordenado de La Ley N.º 27806, Ley de Transparencia y Acceso a la Información Pública¹. Asimismo, cabe añadir que la Ley General de Aduanas, establece en su artículo 163 que, para ejercicio del control aduanero, la Administración Aduanera emplea, principalmente las técnicas de gestión de riesgo para focalizar las acciones de control en aquellas actividades o áreas de alto riesgo, respetando la naturaleza confidencial de la información obtenida para tal fin.

El servicio tendrá un plazo de duración de trece (13) meses y 10 días calendarios, plazo que incluye las aprobaciones y conformidades, del Plan de Trabajo e Informes que presentará el experto, en función al detalle contenido en el siguiente cuadro:

¹ La Ley de Transparencia y Acceso a la Información Pública, en su artículo 1, establece que tiene como finalidad promover la transparencia de los actos del Estado y regular el derecho fundamental del acceso a la información consagrado en la Constitución Política del Perú; asimismo, en su Artículo 17 establece las excepciones al ejercicio de la información confidencial.

Entregable	Detalle del entregable	Plazo máximo del entregable (*)
01	Entregable 01: Plan de Trabajo, detallado de acuerdo con lo indicado en el ítem 8.1.	A los cinco (5) días calendario contados a partir del día siguiente hábil de la comunicación de la UEMSI sobre el inicio del servicio.
02	Entregable 02: Informe sobre la creación de la tabla para entrenamiento del modelo texto clasificador, así como la tabla culminada.	A los treinta (30) días calendario contados a partir del día siguiente hábil de la comunicación de la UEMSI sobre el inicio del servicio.
03	Entregable 03: Informe sobre el desarrollo del preprocesamiento de los datos del modelo texto clasificador, culminado.	A los setenta (70) días calendario contados a partir del día siguiente hábil de la comunicación de la UEMSI sobre el inicio del servicio.
04	Entregable 04: Informe sobre el desarrollo del entrenamiento del modelo texto clasificador con técnica 1 de NLP, culminado.	A los cien (100) días calendario contados a partir del día siguiente hábil de la comunicación de la UEMSI sobre el inicio del servicio.
05	Entregable 05: Informe sobre el desarrollo del entrenamiento del modelo texto clasificador con técnica 2 de NLP, culminado.	A los ciento veinte (120) días calendario contados a partir del día siguiente hábil de la comunicación de la UEMSI sobre el inicio del servicio.
06	Entregable 06: Informe sobre el desarrollo del entrenamiento del modelo texto clasificador con técnica 3 de NLP y comparación con los modelos que consideran la técnica 1 y 2 de NLP, culminados.	A los ciento cincuenta (150) días calendario contados a partir del día siguiente hábil de la comunicación de la UEMSI sobre el inicio del servicio.
07	Entregable 07: Informe sobre el desarrollo de los ajustes al modelo texto clasificador en función a la retroalimentación, culminado.	A los ciento ochenta (180) días calendario contados a partir del día siguiente hábil de la comunicación de la UEMSI sobre el inicio del servicio.
08	Entregable 08: Informe final sobre el modelamiento culminado del modelo texto clasificador, documentación culminada para la implementación de la variable "índice de correspondencia de clasificación arancelaria" y acompañamiento para su implementación.	A los doscientos doce (212) días calendario contados a partir del día siguiente hábil de la comunicación de la UEMSI sobre el inicio del servicio.
09	Entregable 09: Informe sobre la creación de la tabla para entrenamiento del modelo referencias SIVEP, así como la tabla culminada.	A los doscientos treinta y siete (237) días calendario contados a partir del día siguiente hábil de la comunicación de la UEMSI sobre el inicio del servicio.
10	Entregable 10: Informe sobre el desarrollo del preprocesamiento de	A los doscientos sesenta y ocho (268) días calendario contados

Entregable	Detalle del entregable	Plazo máximo del entregable (*)
	los datos del modelo referencias SIVEP, culminado.	a partir del día siguiente hábil de la comunicación de la UEMSI sobre el inicio del servicio.
11	Entregable 11: Informe sobre el desarrollo del entrenamiento del modelo referencias SIVEP con técnica 1 de NLP, culminado.	A los trescientos (300) días calendario contados a partir del día siguiente hábil de la comunicación de la UEMSI sobre el inicio del servicio.
12	Entregable 12: Informe sobre el desarrollo del entrenamiento del modelo referencias SIVEP con técnica 2 de NLP y comparación con los modelos que consideran la técnica 1 de NLP, culminados.	A los trescientos treinta (330) días calendario contados a partir del día siguiente hábil de la comunicación de la UEMSI sobre el inicio del servicio.
13	Entregable 13: Informe sobre la culminación del desarrollo culminado de los ajustes al modelo referencias SIVEP en función a la retroalimentación.	A los trescientos sesenta (360) días calendario contados a partir del día siguiente hábil de la comunicación de la UEMSI sobre el inicio del servicio.
14	Entregable 14: Informe final sobre el modelamiento culminado del modelo referencias SIVEP, documentación culminada para la implementación de la variable "diferencia del valor unitario declarado versus referencias SIVEP" y acompañamiento para su implementación.	A los trescientos noventa y seis (396) días calendario contados a partir del día siguiente hábil de la comunicación de la UEMSI sobre el inicio del servicio.

(*) Si la fecha de entrega coincide con un sábado, domingo o feriado, esta fecha se trasladará al día hábil siguiente.

Los plazos máximos en días para la revisión, levantamiento de observaciones y conformidad de cada entregable, por el experto o por la SUNAT, según corresponda serán los siguientes:

Entregable	Detalle del entregable	Revisión SUNAT (*)	Levantamiento Observaciones (*)	Conformidad de la SUNAT (*)
01	Entregable 01: Plan de Trabajo, detallado de acuerdo con lo indicado en el ítem 8.1.	3	2	2
02	Entregable 02: Informe sobre la creación de la tabla para entrenamiento del modelo texto clasificador, así	10	5	5

Entregable	Detalle del entregable	Revisión SUNAT (*)	Levantamiento Observaciones (*)	Conformidad de la SUNAT (*)
	como la tabla culminada.			
03	Entregable 03: Informe sobre el desarrollo del preprocesamiento de los datos del modelo texto clasificador, culminado.	10	5	5
04	Entregable 04: Informe sobre el desarrollo del entrenamiento del modelo texto clasificador con técnica 1 de NLP, culminado.	10	5	5
05	Entregable 05: Informe sobre el desarrollo del entrenamiento del modelo texto clasificador con técnica 2 de NLP, culminado.	10	5	5
06	Entregable 06: Informe sobre el desarrollo del entrenamiento del modelo texto clasificador con técnica 3 de NLP y comparación con los modelos que consideran la técnica 1 y 2 de NLP, culminados.	10	5	5
07	Entregable 07: Informe sobre el desarrollo de los ajustes al modelo texto clasificador en función a la retroalimentación, culminado.	10	5	5
08	Entregable 08: Informe final sobre el modelamiento culminado del modelo texto clasificador,	10	5	5

Entregable	Detalle del entregable	Revisión SUNAT (*)	Levantamiento Observaciones (*)	Conformidad de la SUNAT (*)
	documentación culminada para la implementación de la variable “índice de correspondencia de clasificación arancelaria” y acompañamiento para su implementación.			
09	Entregable 09: Informe sobre creación de la tabla para entrenamiento del modelo referencias SIVEP, así como la tabla culminada.	10	5	5
10	Entregable 10: Informe sobre el desarrollo del preprocesamiento de los datos del modelo referencias SIVEP, culminado.	10	5	5
11	Entregable 11: Informe sobre el desarrollo del entrenamiento del modelo referencias SIVEP con técnica 1 de NLP, culminado.	10	5	5
12	Entregable 12: Informe sobre el desarrollo del entrenamiento del modelo referencias SIVEP con técnica 2 de NLP y comparación con los modelos que consideran la técnica 1 de NLP, culminados.	10	5	5
13	Entregable 13: Informe sobre el desarrollo	10	5	5

Entregable	Detalle del entregable	Revisión SUNAT (*)	Levantamiento Observaciones (*)	Conformidad de la SUNAT (*)
	culminado de los ajustes al modelo referencias SIVEP en función a la retroalimentación.			
14	Entregable 14: Informe final sobre el modelamiento culminado del modelo referencias SIVEP, documentación culminada para la implementación de la variable "diferencia del valor unitario declarado versus referencias SIVEP" y acompañamiento para su implementación.	10	5	5

(*) Si la fecha de entrega coincide con un sábado, domingo o feriado, esta fecha se trasladará al día hábil siguiente.

En el caso de entregables incompletos, se considerarán como no presentados y empezará a computar el plazo de la penalidad correspondiente.

Los entregables serán presentados en forma digital (en PDF y archivos en formato original editable) y/o físico, indicando su carácter de confidencial, en mesa de partes virtual de la SUNAT (<https://www.sunat.gob.pe/ol-at-ittramitedoc/registro/iniciar>) 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 08:30am a 04:30pm, los mismos deberán estar dirigidos a la Coordinación Técnica de la UEMSI.

Los pagos serán de acuerdo a la conformidad de los entregables, importe que está sujeto a la respectiva retención de impuesto a la renta.

10. PERFIL DEL EXPERTO Y COMPETENCIAS

Formación Académica Mínima	Bachiller o titulado o licenciado en Estadística, Matemáticas, Investigación Operativa, Ingeniería Informática, Ingeniería de Sistemas, Ciencias de la Computación y/o afines. Estudios concluidos de Maestría o Magister en Estadística, Matemáticas, Investigación Operativa, Ingeniería Informática, Ingeniería de Sistemas, Ciencias de la Computación y/o afines.
Estudios adicionales	Curso o diplomado o maestría con relación a inteligencia artificial o procesamiento de lenguaje

	natural (NLP), de al menos veinte (20) horas lectivas.
Experiencia general	Experiencia general de al menos tres (03) años como Científico de Datos o Especialista de Datos y/o afines en el sector público y/o privado.
Experiencia específica	Experiencia específica de al menos un (01) año en procesamiento de lenguaje natural (NLP).
Conocimientos	<ul style="list-style-type: none"> - Machine Learning - Lenguajes de programación: Python, PySpark, Spark u otros. - Manejo de bases de datos y big data: Oracle, MongoDB u otros
Entrevistas	Criterios de evaluación: <ul style="list-style-type: none"> - Dominio temático en procesamiento de lenguaje natural. - Capacidad de trabajo en equipo - Habilidades de comunicación - Capacidad analítica - Creatividad y resolución de problemas - Ética en el manejo de datos
Acreditación	<ul style="list-style-type: none"> - La formación académica se acreditará mediante copia simple del diploma o constancias o certificados o cualquier otra documentación que de manera fehaciente demuestre la formación académica requerida. - La experiencia general y específica se acreditará mediante copia simple de contratos y su respectiva conformidad o constancias o certificados o cualquier otra documentación que de manera fehaciente demuestre la experiencia requerida; asimismo se considera válido demostrar la experiencia laboral mostrando el portafolio de proyectos desarrollados en algún repositorio.

11. LUGAR DEL SERVICIO

El servicio se desarrollará con un enfoque mixto, donde determinadas actividades se realizarán de manera presencial y otras de manera remota, de acuerdo con las acciones previstas en el plan de trabajo aprobado por la SUNAT, teniendo en cuenta la propuesta del plan de trabajo efectuada por el experto.

El experto debe contar con su propio equipo de cómputo que le permita brindar el servicio solicitado, que tenga como mínimo 32 GB de memoria RAM, procesador i7 o ryzen7 y disco duro de 1TB.

Las actividades presenciales se realizarán en la Av. Garcilaso de la Vega 1472 Cercado de Lima - Perú y excepcionalmente en otras sedes de la SUNAT en el Departamento de Lima - Perú y/o en la Provincia Constitucional del Callao - Perú. La SUNAT brindará las facilidades que se requieran para el desarrollo del servicio, tales como: espacio de trabajo y la información necesaria para la prestación del servicio.

12. UNIDAD ORGANIZACIONAL RESPONSABLE DE OTORGAR LA CONFORMIDAD PREVIA AL PAGO

El área responsable de otorgar la conformidad de los entregables indicados en el numeral 9 del presente documento será la Coordinación Técnica de la UEMSI, previa opinión favorable de la GPAFC, GGRIA, y la Jefatura del PGCI.

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

El costo total de los servicios asciende a S/.208 000.00 (dos cientos ocho mil y 00/100 soles), importe que está sujeto a la respectiva retención de impuesto a la renta.

El experto prestará el servicio objeto del presente contrato, desde la fecha de inicio de actividades prevista en el contrato, por un plazo de duración de trece (13) meses y 10 días calendarios para la entrega del informe final, a efecto de que pueda obtener la aprobación y conformidad correspondiente.

La forma de pago será según el cumplimiento de los entregables señalados en el numeral 9 del presente documento, previa aprobación del respectivo informe del experto por parte de la Coordinación Técnica de la UEMSI, contando con la opinión favorable de la GPAFC, GGRIA y la Jefatura del PGCI.

La forma de pago se considerará de la siguiente manera:

Entregable	Detalle del entregable	% de pago
01	Entregable 01: Plan de Trabajo, detallado de acuerdo con lo indicado en el ítem 8.1.	-
02	Entregable 02: Informe sobre la creación de la tabla para entrenamiento del modelo texto clasificador, así como la tabla culminada.	7
03	Entregable 03: Informe sobre el desarrollo del preprocesamiento de los datos del modelo texto clasificador, culminado.	7
04	Entregable 04: Informe sobre el desarrollo del entrenamiento del modelo texto clasificador con técnica 1 de NLP, culminado.	7
05	Entregable 05: Informe sobre el desarrollo del entrenamiento del modelo texto clasificador con técnica 2 de NLP, culminado.	7
06	Entregable 06: Informe sobre el desarrollo del entrenamiento del modelo texto clasificador con técnica 3 de NLP y comparación con los modelos que consideran la técnica 1 y 2 de NLP, culminados.	10
07	Entregable 07: Informe sobre el desarrollo de los ajustes al modelo texto clasificador en función a la retroalimentación, culminado.	7
08	Entregable 08: Informe final sobre el modelamiento culminado del modelo texto clasificador, documentación culminada para la implementación de la variable “índice de correspondencia de clasificación arancelaria” y acompañamiento para su implementación.	10

Entregable	Detalle del entregable	% de pago
09	Entregable 09: Informe sobre creación de la tabla para entrenamiento del modelo referencias SIVEP, así como la tabla culminada.	7
10	Entregable 10: Informe sobre el desarrollo del preprocesamiento de los datos del modelo referencias SIVEP, culminado.	7
11	Entregable 11: Informe sobre el desarrollo del entrenamiento del modelo referencias SIVEP con técnica 1 de NLP, culminado.	7
12	Entregable 12: Informe sobre el desarrollo del entrenamiento del modelo referencias SIVEP con técnica 2 de NLP y comparación con los modelos que consideran la técnica 1 de NLP, culminados.	7
13	Entregable 13: Informe sobre el desarrollo culminado de los ajustes al modelo referencias SIVEP en función a la retroalimentación.	7
14	Entregable 14: Informe final sobre el modelamiento culminado del modelo referencias SIVEP, documentación culminada para la implementación de la variable “diferencia del valor unitario declarado versus referencias SIVEP” y acompañamiento para su implementación.	10

14. CONFIDENCIALIDAD Y PROPIEDAD INTELECTUAL

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

Todo producto o materiales (impresos, estudios, informes, gráficos, programas, software de computación u otros), que se genere por el servicio, no constituye títulos de propiedad, derechos de autor y otro tipo de derechos para el experto, el mismo que a mérito del presente TDR, cede en forma exclusiva, sin generar retribución adicional a lo estipulado en el presente documento.

15. PENALIDADES

PENALIDAD POR RETRASO INJUSTIFICADO DE LAS PRESTACIONES

Si el experto incurre en retraso injustificado de las prestaciones objeto del presente término de referencia, la SUNAT le aplicará una penalidad por cada día calendario de retraso, deducible previa comunicación, de los pagos pendientes.

La penalidad se empezará a aplicar al día siguiente calendario desde la fecha límite para la presentación del entregable correspondiente.

La penalidad se aplicará y se calculará de acuerdo con la siguiente fórmula:

Penalidad diaria = $0.10 \times \text{Monto del Entregable} / 0.25 \times \text{Plazo en días calendario del Entregable (*)}$.

(*) El monto del entregable equivale al valor que se obtiene del porcentaje de pago definido en el numeral 13 de estos términos de referencia.

En caso de que hubiera observaciones al entregable, la penalidad se aplicará a partir del día siguiente de transcurrido el plazo para el levantamiento de observaciones, si el experto no hubiera levantado las mismas.

En el caso de entregables incompletos, se considerarán como no presentados y empezará a computar el plazo de la penalidad correspondiente.

La penalidad será aplicada hasta por un monto máximo equivalente al diez por ciento (10%) del monto total del contrato. Cuando se alcance el monto máximo de la penalidad, la SUNAT podría resolver el contrato por incumplimiento.

La penalidad establecida en la presente cláusula se aplicará sin perjuicio de la obligación del experto de responder por los daños y perjuicios que pudieran derivarse de su incumplimiento o de las demás sanciones que pudieran corresponder; debiéndose aplicar lo establecido en el capítulo IV-Incumplimiento del contrato del Reglamento.