**PROPUESTA IDEA ANTI CORRPUCION - CAZACORRUPTOS ANALYTICS**

# Preámbulo:

## Dudas:

¿Cómo una empresa puede ofrecer servicios tan diversificados?

¿Cuánto tiempo tienen ofreciendo estos servicios?

¿Por qué las fechas no coinciden?

# Problema

No existe una herramienta accesible para todo público que integre la información de los datos para combatir la corrupción y que permita la interacción y retro alimentación de la comunidad.

## Causas:

* Los datos que se encuentran publicados en los portales públicos contra la corrupción se encuentran con diferentes formatos, desactualizados, en diferentes repositorios, no existen las URLS y existe información que se podría considerar como open data como las consultas públicas.
* Relaciones entre las personas de las empresas, marcas atípicas y relaciones entre empresas cuya visualización no se puede contar a través de historias.
* Necesidades de agregar documentos, imposibilidad de acceder a la información. Existe un degaste de tiempo al acceder y procesar la información puede ser muy tedios el integrar varias fuentes de datos.
* Los índices de precios de los productos desde la oferta no son comparables contra las ofertas realizadas en los contratos y por tanto muchas de las veces están por encima de los precios del mercado. \*Debe existir un precio de referencia, organismo o entidad.
* Existir ambigüedad respecto de los productos, deben ser medible. Luego debe ser clara y medible los productos.
* Existen errores u omisiones en los propios documentos generados como reportes de las instituciones que deben reportar esta información. Luego al no estar los datos en línea se produce un efecto de congelamiento de datos cuya extracción posterior imposibilita mecanismos efectivos para combatir la corrupción.
* Ciertos números de contratos que tuvieron un problema en particular.

## Posibles soluciones:

* Calificaciones de proveedores con rankings. Puntuación de riesgo. Niveles.
* Unificar fuentes de información y análisis automático de estas fuentes.
* Personas que ayuden a identificar estos casos. Crowdsourcing la mejor referencia es quien pude ayudar agregar exactitud a estos modelos.
* Presentar reporte de empresas relacionado anónimo y mostrar cuales empresas están participando de uno o varios concursos.
* Empresas de geo referencia en un mismo lugar.
* Modelo de riesgo que siga una estructura y se exija un número que en función de los diferentes sistemas se pueda calificar en una matriz de riesgo.
* Enfoque al periodismo investigativos. Personas calificadas. Rankign de personas que colaboraron.

# Propuesta:

***Crea una herramienta accesible para todo público que integre la información de los datos para combatir la corrupción y que permita la interacción y retro alimentación de la comunidad.***

## Objetivo:

Mejorar la visualización de ofertas, recomendación vía aplicación web publica y bajo suscripción que permita calificar los riesgos de las compras para que la comunidad pueda realizar un control social efectivo.

* Recomendar sin necesidad de ser parte del grupo de proveedores calificados del estado.
* Ayudar a que las personas puedan levantar dudas o sospechas.
* Usar una comunidad para co-crear e implementar servicios de detección de casos de corrupción.
* Crear un asistente virtual que use NLP y Deep learning, para llenar los datos de participantes precios de referencia sea con paginas u otros sitios mejores datos y/o predicciones a futuros.
* Extraer información de diferentes fuentes de carácter público y ponerla en línea para cualquier persona en la sociedad.
* Crear un reposito de código fuente libre para la generación de modelos y visualizaciones de la información.
* Crear un reposito de código analítico libre para la implantación de modelos de detección de fraude.
* Utilizar contenedores y servidores de aplicación de código abierto para el despliegue de la solución.

A quien que para quien está dirigida:

* Sociedad: Para que puedan visualizar los procesos de compras que están ocurriendo en este momento.
* Gobierno: Para la toma de decisiones respecto de ciertos productos o proveedores que están realizando prácticas dudosas en los procesos de contratación. Crear nuevas leyes en función de los datos.
* Científicos: Crear nuevos modelos de caracterización de malas prácticas. Permitir la interacción comunitaria en el campo de la ciencia de datos aplicada a problemas prácticos.
* Empresa: Detección de malas prácticas entre socios o responsable de las compañías para que puedan realizar eximentes intra-institucionales.

# Tiempo y recursos:

## Entregables:

1. Web Comunitaria Cazacorruptos Analytics
2. Web Visualizador Cazacorruptos Analytics.
3. Extracción y limpieza de fuentes.
4. Extracción de características con ML.
5. Extracción de contexto con Deep Learning.
6. Implementación Arquitectura Streaming.
7. Servicios rest de datos limpios.
8. Conectores Instituciones Gubernamentales.
9. Pruebas de la aplicación.
10. Pruebas de carga.
11. Despliegue productos.

## Recursos:

Es importante mencionar que al estar los productos como fuentes open source y bajo una licencia GPL, estos permitirán que sea la comunidad quien pueda ir mejorando esta aplicación. No es necesario la adquisición de licenciamiento para todos aquellos entregables que se encuentren bajo un esquema de GLP.

1. Equipo de trabajo
   1. Equipo de científicos de datos.
   2. Equipo desarrollador.
   3. Equipo experiencia de usuario.
   4. Equipo de QA y liberaciones.
   5. Equipo de administradores de aplicaciones.
   6. Equipo funcional.
   7. Equipo de periodistas en analíticas.
   8. Equipo liberación y despliegue.
   9. Equipo instituciones públicas.
2. Hadware
3. Software Base
4. Licenciamiento
5. Comunicación post producción

## Inversión:

En caso de desear realizarse una primera versión funcional para luego realizar una publicación de las fuentes para mejorar se toma en consideración:

* Para la inversión se considera un tiempo de desarrollo inicial de seis meses con un sueldo de $1200 por recurso.

**Personas:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Entregables** | **Nro Recursos** | **Costo x Entregable** | **Tiempo en Meses** |
| 1.       Web Comunitaria Cazacorruptos Analytics | 1 | 2400 | 1 |
| 2.       Web Visualizador Cazacorruptos Analytics. | 2 | 4800 | 2 |
| 3.       Extracción y limpieza de fuentes. | 2 | 4800 | 2 |
| 4.       Extracción de características con ML. | 1 | 2400 | 0.2 |
| 5.       Extracción de contexto con Deep Learning. | 1 | 2400 | 0.2 |
| 6.       Implementación Arquitectura Streaming. | 1 | 2400 | 0.2 |
| 7.       Servicios rest de datos limpios. | 1 | 2400 | 0.2 |
| 8.       Conectores Instituciones Gubernamentales. | 1 | 2400 | 0.2 |
| 9.       Pruebas de la aplicación. | 1 | 2400 | 0.2 |
| 10.   Pruebas de carga. | 1 | 2400 | 0.2 |
| 11.   Despliegue productos. | 1 | 2400 | 0.2 |
|  | 13 | 31200 | 6.6 |

Recursos:

|  |  |  |
| --- | --- | --- |
| **Recurso** | **Numero Recursos** | **Costo** |
| 1.       Equipo de trabajo | 13 | 31200 |
| 2.       Hadware | 3 | 7000 |
| 3.       Software Base | 3 | 7000 |
| 4.       Licenciamiento | 1 | 1200 |
| 5.       Comunicación post producción | 2 | 3000 |
|  |  | 49400 |

## Tiempo:

En función de las estimaciones se puede desarrollar este producto en un periodo estimado de tres meses.

IMACTO DEL NEGIO Y COMPLEDIAD DE DESARROLLO

Se puede crear una dirección y la cooperación internacional para la validación y verificación en función de las salidas de dividas tanto de compañías como sociedades, en este caso el entere rector es el Foro Global.

Ayuda colaborativa con las Instituciones bajo la lógica de gobierno abierto.

En este caso se evidenciaron las siguientes instituciones para dar un valor agregado al proceso de contratación y transparencia:

Senae: relacionadas con la valoración de productos, desde su entrega y recepción en puerto

SRI: La inclusión de validación de valores de estos productos en el mercado. Extracción de información relevante a productos desde facturación electrónica. Presentación de impuestos de las personas que intervienen las compras públicas.

INEC: la conexiones entre empresas y accionadas que participan en los procesos de contracción.

Secorp: Generar reportes de compras publica y estado de contratos buscando maximizar la información contenida en los contratos.

Impacto como vamos a llegar modificar y mejorar como puede ser

\*\*IMPACTO\*\*

¿Cuál es el enfoque elegido?

Esta solución se encuentra enfocada tomado como base los principales derechos humanos:

Justica: es necesario para evidenciar los casos que actualmente están siendo ejecutados en los procesos de contratación y por ende pueden generar o producir nuevas notificaciones a los entes de control.

Igualdad: El reconocer que existen contratos de orden publica con una ponderación mayor al 40% para persona naturales involucrara una evaluación de los principios de igualdad y responsabilidad en los procesos de contratación. Por otro lado, se puede evidenciar una marcada desigualdad respecto de los ingresos producidos por contratos contra los valores repostados en la administración tributaria.

Participación Ciudadana: La comunidad tendría el derecho y acceso a información pública en línea. Los desarrolladores y/o científicos tendrán acceso a los repositorios de código fuente en donde podrán seguir mejorando los modelos estadísticos, predictivos y generando nuevas visualizaciones que generen un verdadero valor a la sociedad y pueda estar informada y sacar conclusiones de los datos.

Esta solución se encuentra enfocada tomado como base otros sectores:

Territorio: La data que se encuentra de-centralizada, el objetivo es crear un repositorio común y filtrado con la información de las diversas fuentes que componen a un territorio para crear nuevos mecanismos correspondiente a organización territorial y distribución de la riqueza.

Transparencia: Debido a que se podría visualizar cuales son los proveedores que más ingresos tienen. Los datos abiertos no solo serán abiertos sino vivos para que su análisis pueda ser procesado y utilizado incluso para generar nuevas fuentes de empleo.

Sector en temas de salud y de emergencia: Evidenciar la necesidad de detectar cuales son los productos que mayor ingreso representan para el estado y llegar a realizar convenios internacionales para la reducción de aranceles. Conectar precios con los valores del mercado para conocer los valores reales.

Gobierno: Toma de decisiones y acciones en todo nivel, ahorro general debido a que el mostrar precios del mercado contra precios de contratación obliga a los participantes abajar los costos de sus licitaciones. Mayor contrato de las entidades al saber cuáles tiene mayores gastos con valores reales en contratos y cantidad de productos.

Resultados y productos:

* Prototipo Web Herramienta **CAZACORRUPTOS ANALYTICS**
* Visualizadores de casos de corrupción en TableaU.
* Note books para la conversión de los archivos Open Data Contracting para convertirlos en archivos CSV.
* Herramienta Web Open Source para la visualización a través de historias de usuario para la identificación de casos a típicos con servicios Rest y con Flask con Python utilizando los RUCS del SRI y la información.