**¿Cuál es el problema que este proyecto propuesto busca abordar?**

El problema que este proyecto busca abordar es la falta de transparencia en el gasto público destinado a la publicidad estatal. Esta ausencia de información limita la capacidad de la ciudadanía para entender y supervisar el uso de los recursos públicos, afectando la rendición de cuentas y la confianza en la gestión estatal.

**¿Cuáles son los objetivos del proyecto y las actividades que se implementarán para lograr esos objetivos?**

**Objetivo:** Analizar a través del SAPO (herramienta que utiliza datos del SECOP que permite ver si hay sobre costos en la contratación pública del país).

**Objetivos específicos**:

* Implementar un algoritmo de inteligencia artificial que categoriza el lenguaje natural de manera semántica, para analizar la contratación de publicidad del Estado y lo que se invierte.
* Analizar el cambio que ha existido en el tiempo de la inversión en publicidad en el Estado del gobierno actual.

**Productos del modelo:**

El modelo provee información sobre el contrato, En primer lugar reporta un valor de referencia del precio, además da dos valores entre los que es mas probable que se encuentren este tipo de contratos. Por último si se le da un valor sugerido (o se prueba un contrato real) identifica que tanta verosimilitud tiene frente a las referencia del modelo.

**Utilidades de competencia:**

El modelo permite a nuevos contratistas tener información de precios asociados

**Utilidades para corrupción:**

Este modelo permite identificar sobrecostos y subcostos en algunos contratos públicos. La intuición que tenemos es que muchas de las estratégias que se usan para llegar a corrupción (contratos por debajo de puerta, licitaciones faltas, contratación directa) dan la libertad para que los contratos sean en valores más convenientes. De tal manera que queden recursos para repartir entre el licitador y el contratante.

Por otro lado también se puede identificar contratos a más bajo precio de lo normal. en estos casos de corrupción, aunque distintos. los contratos más bajos se pueden lograr para sacar a la competencia, recibir los primeros pagos e incumplir el contrato.

**Limitations:**

Hay una cantidad importante de limitaciones en este proceso. En primer lugar la información sobre la que se entrenó el modelo contiene casos de sobrecostos y subcostos. Aún así el modelo es capaz de reconocer estos casos (utilizando estratégias de regularización y con un error de entrenamiento específico, ver documentación).

Además la información en la que se constituyó es pasada, el mercado puede tener cambios importantes en los precios por información, tecnología, entre otros. Por lo tanto puede que el modelo de referencias de precios esté un poco atrasado. Esperamos reducir este problema con reentrenamientos periódicos sobre nuevos contratos.

Tal vez lo más importante es que el modelo no es perfecto, van a haber casos en los que el modelo levanta una alerta, o sugiere un precio y es errado. Estas herramientas nunca tienen una capacidad absoluta ni perfecta de predicción y por lo tanto deben verse como herramientas para priorizar casos de corrupción y acercarse a la realidad. Sobre el porcentaje de error que tienen realizaremos experimentos con expertos para determinarlo.

**Características técnicas:**

El primer modelo que se utiliza es un transformer tomaarsen/static-similarity-mrl-multilingual-v1. Este modelo es particularmente bueno para aprovechar los contenidos semánticos y aplicarlos a otros lenguajes como el español, además de ser capaz de tener un tuneo fino útil desde capas finales relativamente pequeñas [[2]](https://www.zotero.org/google-docs/?1hS5JU).

Los resultados de procesar la descripción del contrato con este modelo resulta en un valor de entrada compacto, continuo, que expresa los contenidos significativos del contrato para la tarea de identificar el valor que debería tener.

Red neuronal

El primer modelo que se utiliza es un transformer tomaarsen/static-similarity-mrl-multilingual-v1. Este modelo es particularmente bueno para aprovechar los contenidos semánticos y aplicarlos a otros lenguajes como el español, además de ser capaz de tener un tuneo fino útil desde capas finales relativamente pequeñas.

Los resultados de procesar la descripción del contrato con este modelo resulta en un valor de entrada compacto, continuo, que expresa los contenidos significativos del contrato para la tarea de identificar el valor que debería tener.

Red neuronal

El modelo de redes neuronales se entrenó utilizando los resultados de la última capa del modelo de lenguaje y las siguientes variables ["Ciudad", "Orden","Modalidad de Contratación", “Tipo de Contrato”, "Codigo de Categoría Principal", “duracion valor”]. Esta información da una idea de cuánto debería ser el nivel de precios del contrato público.

R2 en datos menores a 1000 millones de dolares en validación: 0.43

R2 en validación: 0.054219614264203075