|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Decorative | | | | |
|  |  |  | |  |
| MODELO DE identificación de licitaciones públicas | | |
| Benjamin kutz | | **Stefani Pereira** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Decorative | | | | |
|  |  |  | |  | |
|  | presentación El modelo que se detalla a continuación tuvo como objetivo general clasificar las licitaciones públicas de Chile en las que las empresas o personas naturales puedan participar según las probabilidades en las que sus ofertas puedan ser aceptadas. | |  | |

|  |  |  |  |
| --- | --- | --- | --- |
|  | |  | |
| CONTEXTO EMPÍRICO **CARACTERIZACIÓN DEL OBJETO DE INVESTIGACIÓN**  Las licitaciones públicas son un método que utiliza la administración para la obtención de productos y servicios, es decir, cuando tiene una necesidad, publica una licitación con características específicas y de acuerdo con los proveedores que apliquen, selecciona al mejor postor en base a un baremo determinado  Se puede acceder a la plataforma de licitaciones públicas en Chile mediante el navegador escribiendo la dirección [www.mercadopublico.cl](http://www.mercadopublico.cl).  Para efectuar una búsqueda, se puede realizar ingresando el nombre, el ID o el título de la licitación o por el buscador avanzado que filtra según organismo público, proveedor, región, nombre de licitación, estado o fechas, arrojando como resultado de búsqueda el ID de la licitación, el nombre o título, el comprador, la fecha de cierre, el estado y más información acerca de la licitación, siendo este el parámetro más importante ya que posee en detalle los productos solicitados y sus cantidades respectivas. (Ver figura1)  Lamentablemente ocurre una dificultad al aplicar los filtros de búsqueda ya que no se puede filtrar por la información de la licitación y en muchos casos, la descripción y el título presentan diferencias, por lo que puede ocurrir que se filtre por título un producto y en la descripción se requiera un producto totalmente diferente o al contrario, que en la descripción se detallen productos que el proveedor tenga en inventario y en el título muestre un producto o servicio desigual y al no poder filtrar por descripción, esas licitaciones son altamente difíciles de ubicar por una persona ya que diariamente se presenta un alto volumen de las mismas que se van actualizando y acumulando, lo que implica mucha comprensión lectora, mucho tiempo de búsqueda y tareas repetitivas que pueden resultar aburridas y agobiantes. | | | |
|  |  | |  |
| Description  Lorem Ipsum is simply dummy text of the printing and typesetting industry. | Description  Lorem Ipsum is simply dummy text of the printing and typesetting industry. | | Description  Lorem Ipsum is simply dummy text of the printing and typesetting industry. |

|  |  |
| --- | --- |
|  |  |
| OBJETIVO GENERAL Crear un modelo que identifique las licitaciones públicas en Chile en las que el proveedor interesado pueda participar según su catálogo de productos. OBJETIVOS ESPECÍFICOS Identificar los procedimientos actuales de participación en las licitaciones públicas de Chile.  Establecer las variables que afectan al proceso de adquisición de licitaciones.  Estudiar el proceso de obtención de licitaciones en Chile. JUSTIFICACIÓN Participar en licitaciones públicas suele generar muchas dudas e incertidumbres sobre los requisitos que se deben cumplir para ganarlas, encontrar la licitación que se adapte al proveedor según producto y cantidades, estar en constante proceso de búsqueda de nuevas licitaciones y una de las más importantes, no perder oportunidades para licitar.  Para las personas naturales o jurídicas que desean participar en las licitaciones públicas, este modelo les permitirá obtener información sobre las licitaciones en las que puedan optar teniendo probabilidades altas de éxito, lo que ayuda a reducir el trabajo manual que lleva al desgaste físico y emocional de la persona quien realiza la búsqueda, disminuir el tiempo de exploración y errores que se puedan cometer en el proceso.  CONTEXTO METODOLÓGICO  Esta etapa tuvo como finalidad describir y definir la metodología que se aplicó para el funcionamiento del modelo la cual se divide en dos ciclos de prueba y el último ciclo como el definitivo que es el modelo final con los datos modificados de acuerdo a las conclusiones que se iban obteniendo al finalizar cada ciclo, es decir, a partir de las conclusiones obtenidas del primer ciclo, se realizaron modificaciones con espera de mejoras para el segundo ciclo y en el tercer ciclo se corrigieron los errores de los ciclos previos y se implementó el modelo en el que se basa el presente trabajo.  A continuación se pueden observar las técnicas implementadas para cada ciclo.  **OBTENCIÓN DE LOS DATOS**  Mediante la Plataforma del Mercado Público, se extrajeron los detalles de las licitaciones públicas de Chile y los productos contenidos en cada licitación, por otro lado, de obtuvieron los datos de un proveedor, su catálogo de productos y sus participaciones en las licitaciones. **COMPLETAR,** se aplicaron métodos de SQL para la fusión de los datoscreando así, un Datasetcon 51.000 registros para entrenar con los campos que se detallan a continuación. **COMPLETAR** (Ver figura 2)   * Tender\_id * Item\_Key * Rubros * Nombre linea adquisición * Descripcion línea adq. * Kupfer   **ANÁLISIS DE LOS DATOS**  En el primer ciclo se realizó un análisis superficial de los datos, **COMPLETAR**  Por otro lado, se pudo observar que se manifiesta un desbalance significativo en la variable “label”, presentándose en la categoría “0” un total del 90% de la muestra y en la “1” solo el 10%.    **PREPROCESAMIENTO DE DATOS**  **Limpieza de Datos**  Una vez realizado el análisis de nuestros datos, en el primer ciclo se procedió a efectuar algunas modificaciones para mejorar su lectura y manejo, como lo fue el cambio de nombre de las columnas por unos de fácil entendimiento para el usuario, el reemplazo del index por el ID de los productos (Item\_Key), eliminación de 113 columnas cuyas descripciones eran valores “NaNs” y retiro de aproximadamente 3.000 registros duplicados.  **Procesamiento de texto**  Natural Language Processing (**NLP**) es el área de estudio centrada en cómo los ordenadores entienden el lenguaje humano, lo interpretan y procesan, por este motivo, se hace necesario modificar el texto de los datos para que el ordenador pueda comprenderlo.  Se analizaron las descripciones del Dataset y se pudo notar que se presentan muchas irregularidades, textos en mayúsculas, minúsculas, muchos símbolos y caracteres especiales, tildes y palabras vacías, por lo que se procedió a eliminar estas últimas y convertir todo el texto en minúscula.  En el primer ciclo transformamos las palabras de las descripciones a su raíz aplicando la técnica de “Stemming” utilizando SnowballStemmer de NLTK, luego, empleando la matriz TF-IDF pudimos obtener la frecuencia en la que aparece cada palabra en las descripciones y al crearse la matriz, cada columna significa una palabra diferente, obteniendo de esta forma 19.535, columnas.  **Reducción de Dimensionalidad**  Al contar con múltiples palabras se hizo necesario aplicar una reducción de estas mediante PCA, que, al iterar con todas las variables, se pudo notar que con 10.000 componentes se explicaba casi el 100% de las variables. (Ver figura 3)  **Balanceo de Muestras**  En una segunda iteración del primer ciclo, para corregir el desbalance en las muestras, aplicamos la técnica “Smote” sobre el Dataframe procesado con 10.000 componentes, se escogió una estrategia de oversampling de la muestra “1” de un 0,30 de la muestra mayoritaria “0” y como resultado se pudo presenciar el aumento de casos sintéticos desde una muestra inicial de 5.000 a casi 14.000 casos.   MODELAMIENTO En la primera iteración del primer ciclo se aplicaron 3 modelos diferentes arrojando los resultados que se muestran a continuación:   |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | **Modelo** | **ROC** | **Accuracy** | **Precision 0** | **Precision 1** | **Recall**  **0** | **Recall**  **1** | **F1-score**  **0** | **F1-score**  **1** | | Regresión Logística | 0,97 | 0,95 | 0,96 | 0.84 | 0.98 | 0.67 | 0.97 | 0,74 | | XGBoost | 0,96 | 0,95 | 0,95 | 0.87 | 0.99 | 0.57 | 0.97 | 0,69 | | Naive Bayes | 0,83 | 0,70 | 1 | 0.26 | 0.67 | 0.99 | 0.80 | 0,41 |   En la segunda iteración luego de aplicar el balanceo de muestras, obtenemos los siguientes resultados:   |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | **Modelo** | **ROC** | **Accuracy** | **Precision 0** | **Precision 1** | **Recall**  **0** | **Recall**  **1** | **F1-score**  **0** | **F1-score**  **1** | | Regresión Logística | 0,988 | 0,95 | 0,95 | 0.93 | 0.98 | 0.83 | 0.97 | 0,88 | | XGBoost | 0,96 | 0,90 | 0,90 | 0.92 | 0.98 | 0.62 | 0.94 | 0,74 | | |



# EVALUACIÓN DEL MODELO

**CONTEXTO CRÍTICO**

ANÁLISIS DE LOS RESULTADOS

**LÍNEAS DE CONTINUIDAD**

**Si el proyecto continuase, ¿Cuál sería el foco a futuro?**

**CONCLUSIONES Y RECOMENDACIONES**

**Anexos**

Figura 1

Interfaz de usuario gráfica

Descripción generada automáticamente

Figura 2

Texto

Descripción generada automáticamente

Figura 3

Gráfico

Descripción generada automáticamente con confianza media