

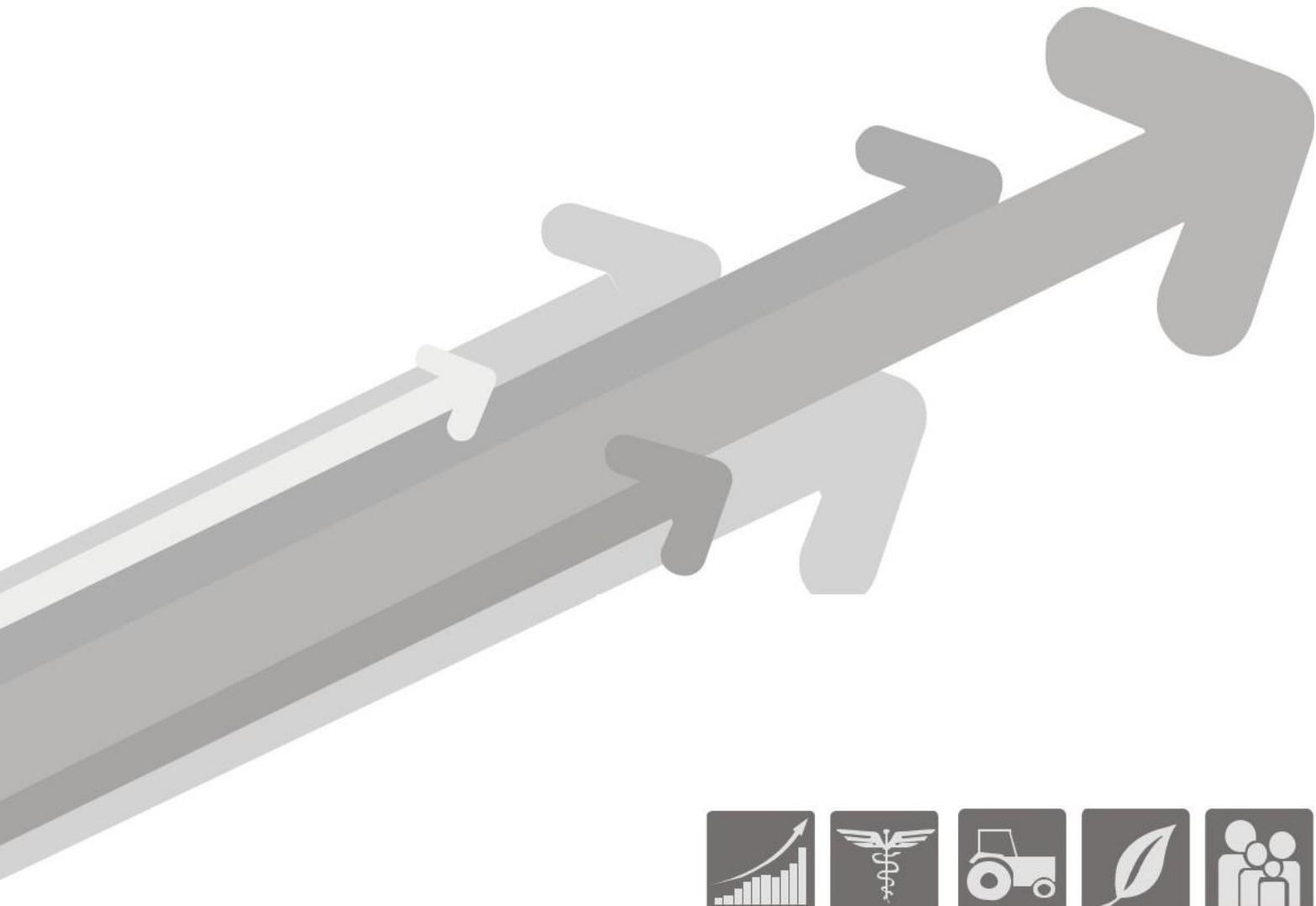
MODELOS DE CLASIFICACIÓN INTERPRETABLES PARA LA CONSTRUCCIÓN DE UN SISTEMA RECOMENDADOR DE LICITACIONES PÚBLICAS

INTERPRETABLE CLASSIFICATION MODELS FOR THE DEVELOPMENT OF A PUBLIC BID RECOMMENDER SYSTEM

LUIS ARTURO ESPIN PAZMIÑO

GABRIEL ELIAS REYES PALMA

GERARDO JUNIOR SANCHEZ CEVALLOS



Modelos de clasificación interpretables para la construcción de un sistema recomendador de licitaciones públicas

INTERPRETABLE CLASSIFICATION MODELS FOR THE DEVELOPMENT OF A PUBLIC BID RECOMMENDER SYSTEM

Luis Espín Pazmiño¹, Gabriel Reyes Palma² y Gerardo Sánchez Cevallos³

Como citar: Espín Pazmiño L. A., Reyes Palma G. E., Sánchez Cevallos G. J. (2025) Modelos de clasificación interpretables para la construcción de un sistema recomendador de licitaciones públicas. *Investigación, Tecnología e Innovación*. 17(23), 97-112. DOI: <https://doi.org/10.53591/iti.v17i23.2173>

RESUMEN

Contexto: El presente trabajo es un proyecto FCI de la Universidad de Guayaquil donde el principal objetivo es el análisis de modelos de clasificación interpretables para la construcción de un sistema recomendador de licitaciones públicas. **Objetivo:** Recopilar datos históricos de licitaciones públicas tomadas desde las páginas oficiales de datos públicos abiertos y procesar los datos obtenidos para desarrollar modelos analíticos. **Método:** Utilización y aplicación de técnicas como regresión logística, en Python como lenguaje de programación, y bibliotecas aptas especializadas para la fácil identificación de patrones y tendencias, **Resultados:** Se realizaron validaciones y comparaciones de métricas de desempeño, el coeficiente de determinación y la precisión en la clasificación, lo que ayudará a tomar la decisión de qué modelo es el más apto para implementarlo en un sistema recomendador, adicional si bien el proyecto se centra en solo dar un solo modelo, se darán conclusiones de qué otro modelo puede usarse en la construcción de este sistema. **Conclusiones:** Este proyecto contribuye al fortalecimiento de la gestión de los recursos públicos mediante herramientas tecnológicas avanzadas que promueven la equidad y la transparencia en la adjudicación de contratos. Las conclusiones y recomendaciones presentadas pueden servir como base para futuros desarrollos de sistemas recomendadores en licitaciones públicas.

Palabras clave: análisis de modelos, licitaciones públicas, modelos interpretables, metodologías.

ABSTRACT

Context: The present work is an FCI project developed at the University of Guayaquil, whose main objective is the analysis of interpretable classification models for the development of a recommendation system for public procurement processes. **Objective:** To collect historical data on public tenders from official open government data platforms and to process the retrieved information for the development of analytical models. **Methodology:** This involves the application of techniques such as logistic regression, utilizing Python as the primary programming language along with specialized libraries designed to facilitate the identification of patterns and trends. **Results:** Performance metrics, including the coefficient of determination and classification accuracy, were validated and compared. These evaluations contribute to determining the most suitable model for implementation in the recommendation system. Although the project is focused on selecting a single model for implementation, conclusions will also be drawn regarding alternative models that could be considered for the system's development. **Conclusions:** This project contributes to the improvement of public resource management through the use of advanced technological tools that promote fairness and transparency in contract awards. The conclusions and recommendations provided herein may serve as a foundation for future developments of recommender systems in the domain of public procurement.

Keywords: model analysis, public bids, interpretable models, methodologies.

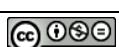
Fecha de recepción: Febrero 14, 2025.

Fecha de aceptación: Mayo 27, 2025

¹ Máster en Administración de Empresas mención Telecomunicaciones, UG, Ecuador, Correo electrónico: luis.espinp@ug.edu.ec

² Ingeniero en Tecnologías de la Información, UG, Ecuador, Correo electrónico: gabriel.reyesp@ug.edu.ec

³ Ingeniero en Tecnologías de la Información, UG, Ecuador, Correo electrónico: gerardo.sanchez@ug.edu.ec



Esta obra está bajo una licencia internacional Creative Commons Atribución-NoComercial-SinDerivadas 4.0. Los autores mantienen los derechos sobre los artículos y por tanto son libres de compartir, copiar, distribuir, ejecutar y comunicar públicamente la obra.

INTRODUCCIÓN

En Ecuador, la contratación pública se dedica a un papel muy importante en la administración eficiente de los recursos del estado, facilitando la adquisición de bienes y servicios esenciales para satisfacer necesidades sociales y promover el desarrollo nacional. No obstante, este proceso enfrenta desafíos importantes que limitan la participación de pequeñas y medianas empresas en las licitaciones públicas. A pesar de los avances logrados con la Ley Orgánica del Sistema Nacional de Contratación Pública (Ley LOSNCP), persisten obstáculos que dificultan la competencia justa y la distribución equitativa de oportunidades.

Este trabajo de investigación se centra en trabajar en el análisis y en la recomendación de configuraciones y modelos de clasificación, tales como árboles de decisión y regresiones logísticas, con el propósito de probar, comparar y seleccionar aquellos modelos que logren un equilibrio óptimo entre precisión y capacidad de interpretación. Se busca proporcionar información útil que permita a los usuarios comprender las recomendaciones y decisiones generadas por dichos modelos.

En el trabajo llamado “Análisis de datos sobre Licitaciones Públicas de Ecuador aplicando técnicas de explicabilidad de modelos basados en características para la mejora de toma de decisiones”, se toma en cuenta y comparan los resultados que obtuvieron al aplicar algoritmos de clasificación y varias técnicas de explicabilidad para sacar los mejores resultados y poder mostrarlos en una interfaz. Este proyecto enfatiza y ayuda a determinar que modelos pueden ser los más precisos para la toma de decisiones y determinar qué empresa puede ser la mejor opción para las licitaciones (Pincay & Velásquez, 2024).

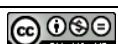
A nivel internacional, el uso de inteligencia artificial explicable ha ganado relevancia, especialmente en sistemas recomendadores. Smith y Jones (2020) demostraron que la implementación de modelos explicables en sistemas de contratación pública en Europa aumentó la transparencia y permitió que se mejore la toma de decisiones en la asignación de contratos gubernamentales. Estas investigaciones han proporcionado una base sólida para la utilización de técnicas como Local Interpretable Model-Agnostic Explanations (LIME) y SHapley Additive exPlanations (SHAP) en la evaluación de modelos complejos (Smith & Jones, 2020).

Los objetivos principales incluyen el análisis detallado de variables y datos relacionados con las licitaciones públicas, la selección de algoritmos de inteligencia artificial explicables, y la optimización de modelos en función de su precisión y transparencia. Este estudio ofrecerá una guía práctica que puede servir como base para la implementación futura de sistemas recomendadores de licitaciones. Dichos sistemas contribuirían a aumentar la transparencia y accesibilidad en los procesos de contratación pública en Ecuador. Finalmente, este proyecto proporcionará pautas y fundamentos técnicos dirigidos a desarrolladores y profesionales, beneficiando directamente a pequeñas, medianas y microempresas al promover un acceso más justo y equitativo a las oportunidades de licitación.

METODOLOGÍA

En el desarrollo del presente proyecto, se implementó una metodología que combina los elementos de CRISP-DM y PMI, ofreciendo un enfoque adaptable y ágil según los requerimientos del proyecto. Esta integración metodológica no solo garantiza un desarrollo estructurado y ordenado, sino que también aborda de manera integral los aspectos técnicos y organizativos del proyecto.

La metodología CRISP-DM, ampliamente reconocida en los campos de minería de datos y aprendizaje automático, destaca por su capacidad para estructurar proyectos basados en datos. En este caso, su implementación fue fundamental para analizar datos históricos de licitaciones públicas en Ecuador y evaluar modelos interpretables como las regresiones logísticas, proporcionando un marco lógico y eficiente para el tratamiento de datos.



Por su parte, la metodología PMI complementa este enfoque al proporcionar un marco estructurado para la gestión del proyecto. Esta metodología asegura la ejecución ordenada de las actividades y permite un control eficiente de los recursos, garantizando que cada etapa del proyecto se lleve a cabo con precisión y efectividad.

La combinación de ambas metodologías permitió abordar el problema de forma técnica y organizada, generando un marco de referencia replicable en proyectos similares relacionados con la contratación pública en Ecuador. Este enfoque integrado asegura que los resultados obtenidos sean tanto técnicamente sólidos como estratégicamente alineados con las necesidades del sector, promoviendo soluciones efectivas y sostenibles en el ámbito de las licitaciones públicas.

CRISP-DM: Comprensión y Preparación de los Datos / PMI: Ejecución

Objetivo. - Recolectar, analizar y preparar datos históricos de licitaciones públicas para su uso en el modelo.

Actividades para realizar. – Recolección de datos, análisis exploratorio, preparación de datos.

TABLA DE ACTIVIDADES

Actividades	Descripción
Recolección de datos	Extraer información de fuentes como el SOCE y RUP
Análisis exploratorio	Identificar variables clave (monto adjudicado, plazo de ejecución, tipo de contrato).
Preparación de datos	Limpieza, estructuración y normalización de los datos.

Nota: Tabla elaborada por autores sobre actividades realizadas en base al desarrollo del proyecto.

Como resultado tenemos el Dataset estructurado y listo para el modelado, reporte sobre la calidad de los datos recolectados.

Modelado: (CRISP-DM: Modelado / PMI: Ejecución)

Objetivo. – Desarrollar simulaciones de los modelos elegidos de clasificación interpretables para analizar el éxito o fracaso en licitaciones públicas.

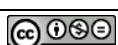
Actividades para realizar. – Simulaciones de **Árbol de Decisión** para clasificar licitaciones como también la Aplicación de Regresión Logística para evaluar probabilidades de éxito. Con lo indicado tenemos:

- Modelos entrenados y validados con métricas de rendimiento.
- Visualizaciones de los modelos, como árboles de decisión graficados.

Análisis de factibilidad

Para realizar el análisis de factibilidad de este proyecto desde un punto de vista técnico, se ha determinado que los modelos interpretables seleccionados, como Árboles de Decisión, Regresión Logística y k-Veinos Más Cercanos (k-NN), son ampliamente reconocidos, robustos y utilizados en diversos contextos. Estos modelos cuentan con implementaciones optimizadas en bibliotecas de Python, como scikit-learn, lo que facilita su adopción en este proyecto. Además, su eficiencia los hace adecuados para manejar datasets de tamaños pequeños, medianos y grandes, lo cual es común en el ámbito de las licitaciones públicas, donde los datos suelen estar ordenados y estructurados. Este nivel de comprensión asegura que los usuarios no técnicos puedan interpretar fácilmente los resultados, promoviendo la aceptación del sistema y los modelos propuestos.

En cuanto a la infraestructura tecnológica, se considera que es adecuada para el desarrollo del proyecto. La tecnología necesaria para entrenar y desplegar los modelos resulta accesible y hacedero tanto a nivel de hardware como a nivel de software en este trabajo. Por el lado del hardware, el entrenamiento de modelos interpretables no requiere el uso de GPU ni equipos de alto rendimiento; una computadora con especificaciones



Esta obra está bajo una licencia internacional Creative Commons Atribución-NoComercial-SinDerivadas 4.0. Los autores mantienen los derechos sobre los artículos y por tanto son libres de compartir, copiar, distribuir, ejecutar y comunicar públicamente la obra.

estándar es suficiente para realizar el procesamiento necesario. Por el lado del software, herramientas como Python, en combinación con editores de código como Visual Studio o Google Colab, y bibliotecas como *pandas*, *numpy* y *matplotlib*, son gratuitas, de código abierto y accesibles, lo que permite la manipulación y edición de datos, la implementación de modelos y su visualización sin costo adicional.

Respecto a la visualización y la interpretabilidad de los resultados, se dispone de bibliotecas como *matplotlib*, *seaborn* y *plotly*, que facilitan la representación gráfica de los resultados y mejoran su comprensión. La regresión logística, por ejemplo, proporciona coeficientes que reflejan una alta importancia de cada variable en la predicción que se está trabajando, mientras que los árboles de decisión ofrecen diagramas claros y detallados que muestran el proceso de clasificación. Estas características hacen que los modelos sean fáciles de interpretar y comprender, lo que fortalece su utilidad práctica y su capacidad para ser adoptados por los actores involucrados en los procesos de licitaciones públicas.

Factibilidad operacional

Este análisis de factibilidad operacional aborda aspectos clave que garantizan la viabilidad del proyecto en términos de adopción y funcionamiento práctico. A continuación, se plantean las principales interrogantes que guían este análisis y se describen los factores que sustentan su factibilidad:

¿Cuál es la disposición de los usuarios para adoptar el sistema, y cómo esto podría fomentar el apoyo administrativo?

Los usuarios finales, como las empresas interesadas en participar en licitaciones, muestran una alta probabilidad de adoptar el sistema propuesto, siempre que puedan comprender fácilmente por qué una licitación específica es recomendada para su perfil. Los modelos interpretables empleados permiten ofrecer explicaciones claras y comprensibles, lo que fomenta la confianza de los usuarios en las recomendaciones generadas.

El proyecto tiene el potencial de recibir un mayor apoyo administrativo al demostrar que estos modelos:

- Mejoran la precisión de las recomendaciones, ajustándolas a las necesidades específicas de las empresas.
- Ofrecen transparencia y explicabilidad, elementos fundamentales en la gestión de licitaciones públicas.

El enfoque en la transparencia asegura que las recomendaciones sean justificables, aumentando la confianza en el sistema. Este beneficio práctico resulta atractivo para los usuarios, ya que reduce la carga manual y los tiempos extensos asociados con la administración, análisis y ordenamiento de grandes volúmenes de datos.

¿Se puede lograr la aceptación de los métodos empleados, y cómo mejora el sistema la experiencia de los usuarios?

Los métodos tradicionales empleados en los portales de licitaciones suelen ser funcionales pero ineficientes, ya que dependen de búsquedas manuales y extensivas. El sistema propuesto, basado en modelos interpretables, mejora significativamente la experiencia del usuario al:

- Reducir el tiempo requerido para analizar oportunidades de licitación.
- Proporcionar recomendaciones basadas en datos relevantes, ajustadas a las características específicas de cada empresa.
- Garantizar la explicabilidad de los resultados, facilitando la comprensión de las recomendaciones generadas.

Este enfoque no solo mejora ya agiliza el proceso, sino que a su vez mejora la confianza de los usuarios al ofrecer soluciones claras y personalizadas.

¿Cómo asegura el involucramiento de los usuarios en el proyecto su aceptación y compromiso? Se debe de tomar en cuenta que el poder involucrar a los usuarios desde las primeras etapas del proyecto es lo más esencial en este trabajo, lo que garantiza su aceptación y compromiso con el sistema propuesto. En este sentido, se han considerado los siguientes aspectos:



- Análisis de tiempos manuales: Se evaluaron los tiempos requeridos para la clasificación y separación de datos en procesos manuales, resaltando la eficiencia que ofrece el sistema automatizado.
- Pruebas y validación de modelos: Los usuarios participaron en la verificación de cómo los modelos seleccionados, como la regresión logística o los árboles de decisión, generan recomendaciones y predicciones de licitaciones.
- Pruebas piloto con datos reales: Se realizaron pruebas utilizando datos reales del portal de compras públicas, con retroalimentación continua para ajustar y mejorar el sistema según las necesidades de los usuarios.

Estas acciones aseguran que el sistema no solo cumpla con las expectativas técnicas, sino que también se alinee con las necesidades prácticas de los usuarios, promoviendo su aceptación y compromiso con el proyecto.

Datos Abiertos

El proyecto considera el uso de datos abiertos disponibles en el portal de contratación pública de Ecuador como un elemento clave para garantizar la viabilidad operativa de la investigación. Estos datos son fundamentales para entrenar, evaluar y validar modelos de clasificación interpretables, los cuales servirán como base para el desarrollo de un sistema recomendador de licitaciones.

El portal proporciona acceso a un conjunto amplio y estructurado de datos relacionados con los procesos de licitación. Su clasificación organizada permite seleccionar información de manera puntual, según las necesidades del análisis, y facilita la identificación de datos relevantes. Entre la información disponible se incluyen:

- Contratos adjudicados: Detalles sobre las empresas ganadoras, montos y condiciones contractuales.
- Pliegos y términos de referencia: Especificaciones técnicas y requisitos establecidos en los procesos de licitación.
- Historial de licitaciones: Registros completos de procesos anteriores, que incluyen categorías, ubicaciones y resultados.

Estos datos, accesibles de forma gratuita en formatos estándar como CSV, JSON y XLSX, simplifican su integración en el análisis de los modelos de clasificación interpretables. Su disponibilidad en formatos compatibles con herramientas tecnológicas asegura que puedan ser procesados de manera eficiente, maximizando su utilidad en la generación de modelos predictivos y fortaleciendo la capacidad de los análisis realizados en este trabajo.

RESULTADOS OBTENIDOS

El presente trabajo cuenta con los siguientes resultados indicados en etapas las cuales se muestran a continuación

Comprensión del negocio y planificación del proyecto

Las actividades que se realizaron en esta etapa constaron en identificar el problema y definir cuáles son los objetivos. Como primer punto se analizaron las deficiencias en los sistemas actuales de licitaciones públicas, como la escasez de interpretabilidad y transparencia. Continuando se evaluaron modelos interpretables como árboles de decisión y regresión logística para la optimización en la consultoría y elección de decisiones.

Comprensión y preparación de datos

Se recolectaron y analizaron datos históricos de licitaciones públicas para utilizar en cada uno de los modelos seleccionados, depurando datos en blancos e irrelevantes para el entrenamiento de los modelos, con lo cual se garantiza que el entrenamiento sea el más adecuado y así poder elegir e que mejores resultados otorgue.

Modelado



Esta obra está bajo una licencia internacional Creative Commons Atribución-NoComercial-SinDerivadas 4.0. Los autores mantienen los derechos sobre los artículos y por tanto son libres de compartir, copiar, distribuir, ejecutar y comunicar públicamente la obra.

El modelo se entrena con un conjunto de datos estructurado, donde se aplican técnicas de preprocesamiento y selección de características. Específicamente, se han utilizado variables clave como el monto de la licitación (*value_amount*), la duración del proceso (*tenderPeriod_durationInDays*), la categoría del procedimiento (*mainProcurementCategory*) y características textuales como el título y la descripción de la licitación (*title*, *description*). Estas transformaciones permiten que el modelo capture patrones relevantes en la toma de decisiones de adjudicación.

En la implementación del modelo, se ha realizado una validación cruzada utilizando Stratified K-Fold, con el fin de evaluar diferentes valores del hiperparámetro de regularización (C). Posteriormente, se selecciona el modelo con el mejor desempeño y se evalúa sobre un conjunto de prueba, midiendo su precisión, matriz de confusión y área bajo la curva ROC (AUC).

REGRESIÓN LOGÍSTICA

El modelo se entrena con un conjunto de datos estructurado, donde se aplican técnicas de preprocesamiento y selección de características. Específicamente, se han utilizado variables clave como el monto de la licitación (*value_amount*), la duración del proceso (*tenderPeriod_durationInDays*), la categoría del procedimiento (*mainProcurementCategory*) y características textuales como el título y la descripción de la licitación (*title*, *description*). Estas transformaciones permiten que el modelo capture patrones relevantes en la toma de decisiones de adjudicación.

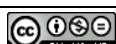
En la implementación del modelo, se ha realizado una validación cruzada utilizando **Stratified K-Fold**, con el fin de evaluar diferentes valores del hiperparámetro de regularización (C). Posteriormente, se selecciona el modelo con el mejor desempeño y se evalúa sobre un conjunto de prueba, midiendo su **precisión, matriz de confusión y área bajo la curva ROC (AUC)**.

Imagen #1: Algoritmo

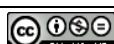
```

1 import pandas as pd
2 import numpy as np
3 from sklearn.feature_extraction.text import TfidfVectorizer
4 from sklearn.preprocessing import OneHotEncoder, StandardScaler
5 from sklearn.model_selection import train_test_split
6 from sklearn.linear_model import LogisticRegression
7 from sklearn.metrics import accuracy_score, confusion_matrix, classification_report, roc_auc_score
8 from sklearn.model_selection import cross_val_score, StratifiedKFold
9
10 # Ruta del archivo con la carga del dataset con encoding adecuado
11 file_path = "tender_2022_licitacion.csv"
12 df = pd.read_csv(file_path, encoding="latin1", delimiter=",")
13
14 # Muestra las primeras filas y la información general del dataset
15 df.head(), df.info()
16
17
18 # AJUSTE DE LOS DATOS DEL DATASET
19 # Elimina columnas innecesarias o vacías
20 columns_to_drop = ["ocid", "release_id", "id", "awardPeriod_startDate", "awardPeriod_durationInDays"]
21 df_cleaned = df.drop(columns=columns_to_drop)
22
23 # Corrige codificación de caracteres en textos
24 text_columns = ["title", "description", "eligibilityCriteria", "procuringEntity_name"]
25 for col in text_columns:
26     df_cleaned[col] = df_cleaned[col].apply(lambda x: x.encode("latin1").decode("utf-8") if isinstance(x, str) else x)
27

```



```
28 # Convierte fechas a formato datetime
29 date_columns = [
30     "tenderPeriod_startDate", "tenderPeriod_endDate", "tenderPeriod_maxExtentDate",
31     "enquiryPeriod_startDate", "enquiryPeriod_endDate", "enquiryPeriod_maxExtentDate",
32     "awardPeriod_endDate", "awardPeriod_maxExtentDate"
33 ]
34 for col in date_columns:
35     df_cleaned[col] = pd.to_datetime(df_cleaned[col], errors="coerce")
36
37 # Rellena valores nulos en 'hasEnquiries' con False
38 df_cleaned["hasEnquiries"] = df_cleaned["hasEnquiries"].fillna(False).astype(bool)
39
40 df_cleaned.info(), df_cleaned.head() # muestra un resumen de los cambios
41
42
43 # Codificación One-Hot para variables categóricas
44 categorical_columns = ["status", "procuringEntity_id", "procuringEntity_name",
45                         "procurementMethod", "mainProcurementCategory", "awardCriteria"]
46
47 encoder = OneHotEncoder(sparse_output=False, drop="first", handle_unknown="ignore")
48 categorical_encoded = encoder.fit_transform(df_cleaned[categorical_columns])
49
50 # Obtener nombres de las columnas generadas por One-Hot Encoding
51 categorical_feature_names = encoder.get_feature_names_out(categorical_columns)
52
53 # Vectorización TF-IDF para variables de texto
54 text_columns = ["title", "description", "eligibilityCriteria"]
55 vectorizer = TfidfVectorizer(max_features=500) # Limitar a 500 características para evitar sobreajuste
56 text_features = vectorizer.fit_transform(df_cleaned[text_columns].apply(lambda x: ' '.join(x), axis=1)).toarray()
57
58 # Escalado de variables numéricas
59 numeric_columns = ["value_amount", "tenderPeriod_durationInDays", "enquiryPeriod_durationInDays", "numberOfTenderers"]
60 scaler = StandardScaler()
61 numeric_scaled = scaler.fit_transform(df_cleaned[numeric_columns])
62
63 # Consolidar todos los datos transformados en un nuevo DataFrame
64 processed_data = np.hstack((categorical_encoded, text_features, numeric_scaled))
65 processed_columns = list(categorical_feature_names) + [f"tfidf_{i}" for i in range(text_features.shape[1])] + numeric_columns
66
67 df_processed = pd.DataFrame(processed_data, columns=processed_columns)
68
69 # Mostrar las primeras filas del dataset transformado
70 df_processed.head()
71
72
73 # Creación, entrenamiento y ajuste del modelo de regresión logística
74
75 y = df_cleaned['status'] # variable objetivo
76 X = df_processed # Estas son las características transformadas que has generado
77
78 # Divide en entrenamiento y prueba
79 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
80
81 # Lista de valores de C para probar con la regularización
82 C_values = [0.001, 0.01, 0.1, 1, 10, 100]
83
84 # Inicializa el clasificador de regresión logística
85 log_reg = LogisticRegression(solver='liblinear', class_weight='balanced', C=1)
```



```

87 # Validación cruzada y cálculo de AUC (Area bajo la curva) para diferentes valores de C
88 for C in C_values:
89     log_reg.C = C
90     auc_scores = cross_val_score(log_reg, X_train, y_train, cv=StratifiedKFold(n_splits=5), scoring='roc_auc')
91     print(f"AUC para C={C}: {np.mean(auc_scores)}")
92
93 log_reg.C = 0.001 #Ajusta el valor de C según el resultado de la validación cruzada
94 log_reg.fit(X_train, y_train)
95
96 # Predecir sobre el conjunto de prueba
97 y_pred = log_reg.predict(X_test)
98 y_prob = log_reg.predict_proba(X_test)[:, 1] # Probabilidades para AUC
99
100 # Imprimir el reporte de clasificación
101 print(classification_report(y_test, y_pred))
102
103 # Matriz de confusión
104 print(confusion_matrix(y_test, y_pred))
105
106

```

Nota: Código fuente elaborado por autores y ejecutado en Visual Studio

Imagen #2: resultado de ejecución de código

```

=> ÁREA BAJO LA CURVA:
AUC para C=0.001: 0.9830227743271222
AUC para C=0.01: 0.9964285714285713
AUC para C=0.1: 0.9995341614906831
AUC para C=1: 1.0
AUC para C=10: 1.0
AUC para C=100: 1.0

=> REPORTE DE LA CLASIFICACIÓN:
      precision    recall   f1-score   support
active          1.00      0.72      0.84      118
complete        0.33      1.00      0.49       16

accuracy          0.75      0.75      0.75      134
macro avg        0.66      0.86      0.66      134
weighted avg     0.92      0.75      0.80      134

=> MATRIZ DE CONFUSIÓN:
[[85 33]
 [ 0 16]]

//MODELO EJECUTADO CON ÉXITO//
```

Nota: Gráfico de Código fuente

Al ejecutar este algoritmo, los resultados obtenidos muestran dos áreas importantes para definir el valor de este modelo. Lo primero a destacar es que se realizó el AUC (área bajo la curva) para evaluar la capacidad discriminativa del modelo, es decir, qué tan bien puede separar las clases. El modelo está clasificando entre dos posibles estados (ejemplo; "active" y "complete"), el AUC es una medida clave para ver qué tan efectivo es el modelo para distinguir entre estas clases.

El AUC mide el área bajo la curva ROC (Receiver Operating Characteristic), que grafica la tasa de verdaderos positivos frente a la tasa de falsos positivos. El valor de AUC varía entre 0 y 1:

AUC = 0.5: El modelo no tiene poder discriminativo, es decir, clasifica aleatoriamente.

AUC = 1: El modelo es perfecto y puede distinguir entre las clases sin errores.



Esta obra está bajo una licencia internacional Creative Commons Atribución-NoComercial-SinDerivadas 4.0. Los autores mantienen los derechos sobre los artículos y por tanto son libres de compartir, copiar, distribuir, ejecutar y comunicar públicamente la obra.

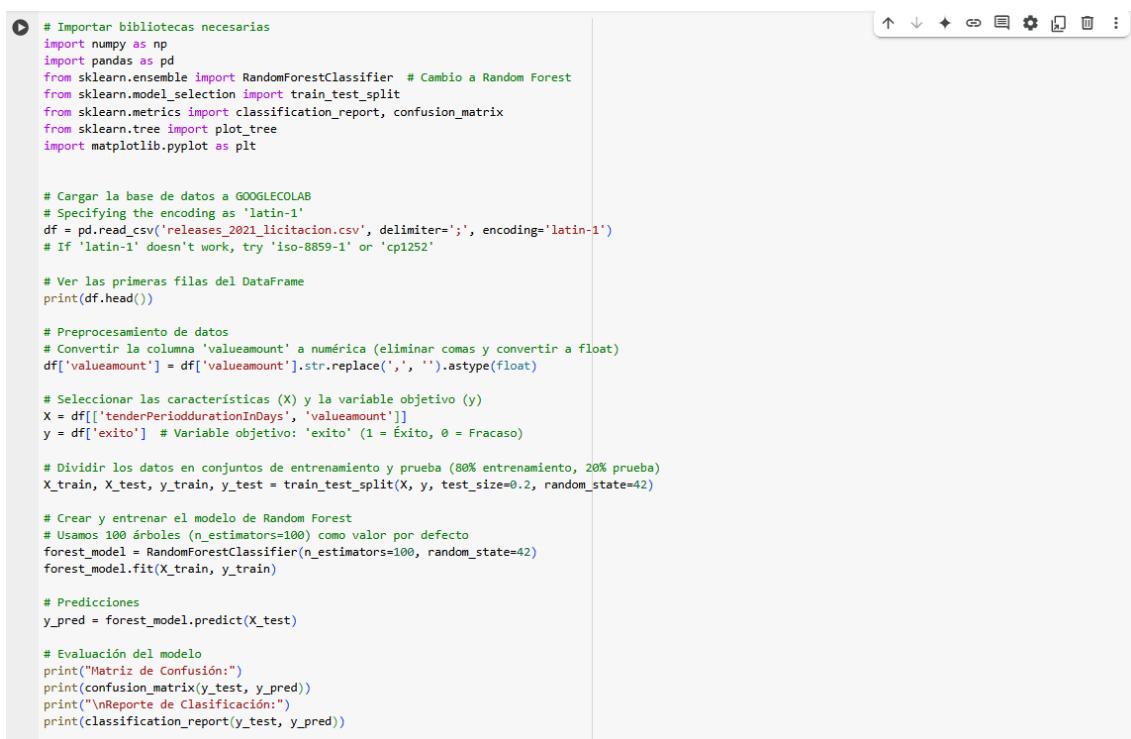
AUC < 0.5: El modelo está clasificando de forma errónea y está peor que el azar.

Por lo que se puede observar, al realizar un ajuste en el valor de C, que para este caso se ajustó en 0.001, para la regularización de los datos que el modelo va a interpretar, en el reporte de la clasificación da valores muy buenos de aprendizaje, lo que indica que ha mejorado considerablemente con el ajuste de **class_weight='balanced'**, logrando buenos resultados en términos de **AUC** y **recall** para la clase "complete". Sin embargo, sigue teniendo dificultades con la **precisión** de "complete", lo cual es típico en situaciones de clases desbalanceadas. Es importante seguir experimentando con otros enfoques y evaluar qué métrica es más relevante según se requiera evaluar.

ÁRBOLES DE DECISIÓN

A continuación, se proporcionará un ejemplo simulando el trabajo del autor para darse cuenta como está estructurado el modelo y su funcionamiento:

Imagen #3: código de árbol de decisión



```

# Importar bibliotecas necesarias
import numpy as np
import pandas as pd
from sklearn.ensemble import RandomForestClassifier # Cambio a Random Forest
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, confusion_matrix
from sklearn.tree import plot_tree
import matplotlib.pyplot as plt

# Cargar la base de datos a GOOGLECOLAB
# Specifying the encoding as 'latin-1'
df = pd.read_csv('releases_2021_licitacion.csv', delimiter=';', encoding='latin-1')
# If 'latin-1' doesn't work, try 'iso-8859-1' or 'cp1252'

# Ver las primeras filas del DataFrame
print(df.head())

# Preprocesamiento de datos
# Convertir la columna 'valueamount' a numérica (eliminar comas y convertir a float)
df['valueamount'] = df['valueamount'].str.replace(',', '').astype(float)

# Seleccionar las características (X) y la variable objetivo (y)
X = df[['tenderPeriodDurationInDays', 'valueamount']]
y = df['exito'] # Variable objetivo: 'exito' (1 = Éxito, 0 = Fracaso)

# Dividir los datos en conjuntos de entrenamiento y prueba (80% entrenamiento, 20% prueba)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Crear y entrenar el modelo de Random Forest
# Usamos 100 árboles (n_estimators=100) como valor por defecto
forest_model = RandomForestClassifier(n_estimators=100, random_state=42)
forest_model.fit(X_train, y_train)

# Predicciones
y_pred = forest_model.predict(X_test)

# Evaluación del modelo
print("Matriz de Confusión:")
print(confusion_matrix(y_test, y_pred))
print("\nReporte de Clasificación:")
print(classification_report(y_test, y_pred))

```



```

# Importancia de las características
importances = forest_model.feature_importances_
feature_names = X.columns
print("Importancia de las características:")
for feature, importance in zip(feature_names, importances):
    print(f"{feature}: {importance:.4f}")

# Gráfico de importancia de las características
plt.figure(figsize=(10, 6))
plt.barh(feature_names, importances, color='skyblue')
plt.xlabel('Importancia')
plt.title('Importancia de las características en el modelo Random Forest')
plt.show()

# Visualización del Árbol de Decisión
plt.figure(figsize=(20, 10))
plot_tree(tree_model, feature_names=['tenderPerioddurationInDays', 'valueamount'],
          class_names=['Fracaso', 'Éxito'], filled=True, rounded=True, fontsize=10)
plt.title("Árbol de Decisión: Clasificación de Licitaciones", fontsize=16)
plt.show()

# Evaluación de la precisión del modelo
accuracy = tree_model.score(X_test, y_test)
print(f"Precisión del modelo de Árbol de Decisión: {accuracy * 100:.2f}%")

# Determinar si el modelo es favorable o moderado
if accuracy >= 0.7:
    print("El modelo tiene una precisión FAVORABLE.")
elif accuracy >= 0.5:
    print("El modelo tiene una precisión MODERADA.")
else:
    print("El modelo tiene una precisión BAJA.")

```

Nota: Código fuente en Google Colab sobre un ejemplo de árbol de decisión

EJEMPLO PRÁCTICO

Problema. - Dado un conjunto de datos históricos sobre licitaciones públicas, se requiere clasificar si una licitación será exitosa o fallida en función de variables como:

Monto Adjudicado: Valor del contrato (\$).

Plazo de Ejecución: Duración estimada del proyecto (días).

Tipo de Contrato: Clasificación (obras, bienes, servicios).

Solución. – Se usará un Árbol de Decisión para entrenar un modelo y clasificar licitaciones como exitosas (1) o fallidas (0). La explicación del algoritmo está detallada en la TABLA #4 (ver en la pág 24).

4. Evaluación

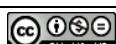
El propósito de esta etapa es validar la eficacia y aplicabilidad de los modelos desarrollados en el proyecto. Las actividades realizadas incluyeron la comparación de los resultados generados por los modelos con datos históricos reales de licitaciones públicas, el análisis del impacto de los modelos en términos de precisión y el registro de limitaciones identificadas, así como oportunidades de mejora.

Como resultado, se elaboró un informe detallado de evaluación que incluye comparativas de rendimiento entre los modelos utilizados y conclusiones clave sobre la efectividad de los modelos seleccionados. Este análisis permitió identificar tanto fortalezas como áreas donde los modelos podrían optimizarse para mejorar su desempeño en aplicaciones futuras.

Etapa 5 Despliegue y generación de pautas

El objetivo de esta etapa fue proporcionar recomendaciones y guías prácticas basadas en los hallazgos obtenidos durante el desarrollo del proyecto, con el fin de facilitar futuros desarrollos de sistemas recomendadores en el ámbito de las licitaciones públicas.

Las actividades realizadas incluyeron la elaboración de pautas técnicas que detallan los pasos necesarios para la implementación efectiva de modelos interpretables. Asimismo, se presentaron los resultados del proyecto y los beneficios asociados a las partes interesadas, enfatizando las mejoras alcanzadas en términos de transparencia y eficiencia. Por último, se formularon propuestas de mejoras y se identificaron áreas de investigación futura, estableciendo un marco de referencia que puede ser replicado en proyectos similares.



Esta obra está bajo una licencia internacional Creative Commons Atribución-NoComercial-SinDerivadas 4.0. Los autores mantienen los derechos sobre los artículos y por tanto son libres de compartir, copiar, distribuir, ejecutar y comunicar públicamente la obra.

Ambas etapas garantizaron que los modelos desarrollados no solo cumplieran con los objetivos planteados, sino que también proporcionaron bases sólidas para la continuidad y expansión de los avances logrados en este proyecto.

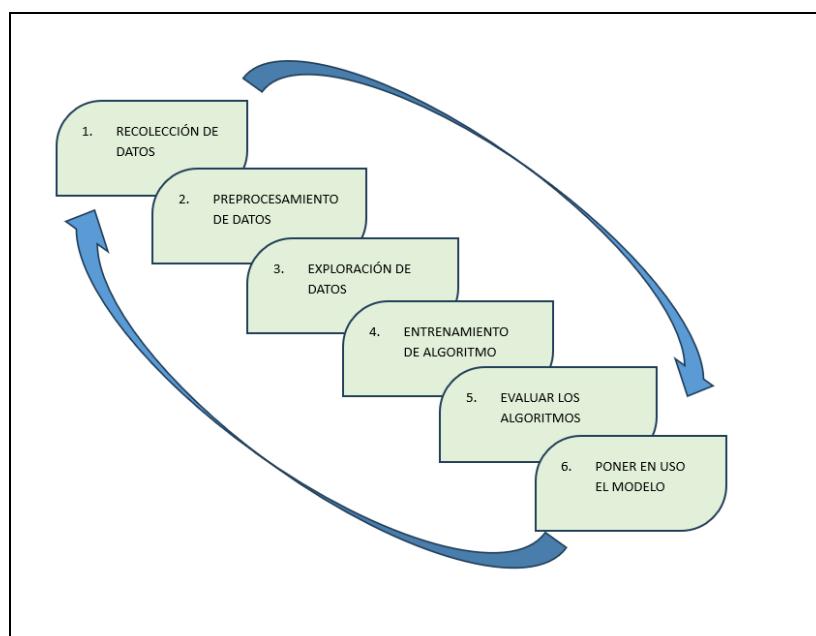


Imagen #4: Gráfico elaborado por autores con pasos para la creación de un modelo en machine learning.

DISCUSIÓN

Criterios de la propuesta

Para garantizar la efectividad y precisión del proyecto, fue esencial definir criterios específicos que permitan medir el desempeño de los modelos y verificar que los objetivos planteados sean alcanzados. Los criterios establecidos se detallan a continuación:

A. Precisión de la clasificación

Este criterio evalúa la cantidad de aciertos en la clasificación de las licitaciones como exitosas o fallidas. Se considera aceptable un nivel de precisión superior al 85%, asegurando que los modelos cumplan con los estándares de rendimiento requeridos.

B. Concordancia con los datos reales

El modelo fue validado comparando sus resultados con datos históricos de licitaciones públicas. Este criterio asegura la aplicabilidad práctica del modelo en escenarios reales y su capacidad para reflejar la realidad del proceso de contratación pública.

C. Robustez del modelo

Se evaluó la estabilidad del modelo al aplicarse en diferentes conjuntos de datos, verificando que mantuviera un rendimiento consistente. Este criterio es crucial para garantizar que el modelo no esté sobreajustado y pueda generalizar sus resultados.

D. Impacto de las variables

Este criterio analiza la importancia y relevancia teórica de las variables seleccionadas en los modelos, como el monto adjudicado y el plazo de ejecución. La confirmación de la relevancia de estas variables respalda la validez de los modelos y su capacidad para explicar los resultados.

E. Facilidad de interpretación

Se priorizó que los resultados y decisiones generados por los modelos sean comprensibles para los actores involucrados, como agencias de contratación y proveedores. Esto facilita su adopción y uso en la toma de decisiones.



F. Evaluación de beneficios

Se determinó cómo los resultados de los modelos contribuyen a mejorar la eficiencia y la transparencia en los procesos de licitación. Este criterio valida que los modelos no solo cumplen objetivos técnicos, sino que también generan beneficios prácticos y medibles para optimizar los procesos y promover una gestión más eficiente de los recursos públicos.

La definición y aplicación de estos criterios permiten una evaluación integral del proyecto, garantizando su alineación con las necesidades del sector y su impacto positivo en el ámbito de las licitaciones públicas.

Capacidad técnica

El diseño del sistema propuesto requiere capacidades técnicas suficientes para procesar grandes volúmenes de datos y realizar cálculos complejos asociados con los modelos de clasificación interpretables planteados en este trabajo. Esto incluye la capacidad de procesamiento necesaria para entrenar modelos mediante técnicas de Machine Learning, almacenamiento adecuado para gestionar datos históricos, y capacidad de red para descargar o actualizar información desde el portal de compras públicas. Estas necesidades técnicas pueden ser satisfechas con un equipo informático estándar. Además, la infraestructura técnica ofrece opciones de escalabilidad, lo que permite adaptarse a un mayor nivel de complejidad o tamaño del sistema en el futuro, facilitando el desarrollo de un sistema recomendador de licitaciones.

La capacidad técnica también considera la importancia de los datos en los modelos de clasificación interpretables. La selección y validación de variables relevantes permite identificar las características clave que influyen en las decisiones empresariales relacionadas con las licitaciones. Este trabajo utiliza estas características como base para construir conjuntos de datos estructurados que alimentan los modelos interpretables.

Ejemplos de variables incluyen la categoría del contrato, nombre y tipo de empresa, monto estimado, y plazo adjudicado. En el caso de la regresión lineal, se trabaja con dos variables: el plazo en días y el monto adjudicado, las cuales permiten analizar la relación entre el tiempo y los valores económicos de las licitaciones.

Para modelos como árboles de decisión, reglas de asociación y regresión logística, es fundamental contar con datos estructurados que permitan aprender patrones y generar predicciones significativas. La riqueza y diversidad de los datos disponibles en el portal de compras públicas aseguran que los modelos sean representativos del contexto ecuatoriano y de las licitaciones publicadas en dicho portal.

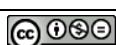
El enfoque en datos abiertos también facilita la evaluación de cómo los modelos justifican sus decisiones. Por ejemplo, un árbol de decisión puede explicar la recomendación de una licitación considerando variables como el presupuesto, el sector al que pertenece y la experiencia previa de una empresa, promoviendo así la transparencia y la comprensión de los resultados.

CONCLUSIONES

Impacto de los modelos aplicados, los árboles de decisión demostraron ser altamente efectivos para clasificar las licitaciones en exitosas o fallidas, permitiendo una interpretación visual y clara de los factores determinantes como el monto adjudicado, el plazo de ejecución y el tipo de contrato. Por otro lado, la regresión logística mostró una gran capacidad para prever probabilidades de éxito en función de las características de las ofertas, proporcionando una herramienta útil para la toma de decisiones fundamentadas, prosiguiendo con los árboles de regresión y la regresión lineal fueron fundamentales para analizar tendencias en los montos adjudicados y predecir resultados financieros, aportando una base sólida para optimizar el manejo de recursos.

Los modelos seleccionados fueron comparados en términos de precisión, interpretabilidad y capacidad predictiva, los árboles de decisión destacaron por su facilidad de interpretación mientras que los modelos de regresión permitieron una predicción más ajustada en variables continuas.

Elección del modelo principal



Esta obra está bajo una licencia internacional Creative Commons Atribución-NoComercial-SinDerivadas 4.0. Los autores mantienen los derechos sobre los artículos y por tanto son libres de compartir, copiar, distribuir, ejecutar y comunicar públicamente la obra.

Este proyecto concluye que el modelo más adecuado para el aprendizaje y la creación del sistema recomendador de licitaciones es el *Modelo de Regresión Logística*. Se tomó en cuenta los resultados que este modelo dio de presión, ya que es más alto que los otros modelos mencionados en este trabajo. El desarrollo de este modelo contribuye a la futura implementación de un sistema recomendador de licitaciones al proporcionar una base analítica para predecir el éxito de una licitación. Dicho sistema podría beneficiarse de este modelo tomando en consideración estos aspectos que se han podido establecer:

1. Filtrado de oportunidades: Permite a los postulantes identificar licitaciones con mayor probabilidad de éxito, optimizando su participación en procesos de contratación pública.
2. Optimización de estrategias: Proporciona información valiosa sobre qué factores influyen en la adjudicación, permitiendo a las empresas ajustar sus propuestas de acuerdo con tendencias detectadas en los datos.
3. Interfaz explicativa: Dado que la regresión logística es un modelo interpretable, los usuarios pueden comprender fácilmente qué factores aumentan o disminuyen la probabilidad de éxito de una licitación.

El código desarrollado, refleja este enfoque al implementar la regresión logística sobre un conjunto de datos estructurado y transformado, garantizando la integridad y calidad de la información utilizada en la predicción. Su correcta aplicación y optimización permiten que este modelo sirva como un punto de partida sólido para futuros desarrollos en sistemas de recomendación basados en inteligencia artificial aplicada a la contratación pública.

El desarrollo, investigaciones y recomendación de los modelos y el mejor propuesto para la creación del sistema recomendador, cuenta con una producción científica que sintetiza las recomendaciones clave derivadas de las pruebas realizadas, en cada uno de los ya mencionados modelos. Este artículo contribuye al campo de las licitaciones públicas. Proporcionando pautas prácticas para el desarrollo de sistemas recomendadores basados en modelos de clasificación interpretables.

Elección de otros modelos que pueden usarse

Si bien en el presente estudio se ha optado por la regresión logística como modelo principal para la clasificación del estado de las licitaciones, es importante reconocer que los árboles de decisión también representan una alternativa viable para la creación de un sistema recomendador de licitaciones públicas, como lo demuestra el presente estudio.

Los árboles de decisión destacan por su capacidad de manejar relaciones no lineales y capturar interacciones complejas entre variables, lo que los hace especialmente útiles en entornos donde los patrones de adjudicación pueden depender de múltiples factores con interacciones no triviales. En la parte de su estructura jerárquica, este permite realizar clasificaciones de manera intuitiva, dividiendo el espacio de decisiones en regiones basadas en reglas interpretables que se le pueden asignar.

Uno de los principales beneficios de este enfoque en este modelo, es su facilidad de interpretación, ya que permite visualizar de manera gráfica cómo cada variable influye en el resultado final. Se toma en cuenta que en un sistema recomendador, esto podría traducirse en explicaciones más claras para los usuarios sobre por qué una licitación es más favorable que otra. Además, los árboles pueden manejar valores faltantes y datos categóricos de manera más flexible en comparación con modelos lineales, sin requerir una transformación exhaustiva de los datos.

Sin embargo, los árboles de decisión pueden ser más sensibles al sobreajuste, especialmente en conjuntos de datos con alta dimensionalidad como el utilizado en este estudio. A pesar de su capacidad de clasificación, pueden perder generalización si no se aplican técnicas como la poda o el uso de árboles ensamblados (Random Forest, Gradient Boosting).



En comparación, la regresión logística ofrece ventajas en términos de estabilidad y simplicidad, asegurando interpretabilidad y la robustez en la predicción del estado de las licitaciones como se lo tenga planteado. La naturaleza de probabilidad que maneja la regresión logística permite obtener estimaciones cuantificables de la certeza de cada predicción, lo que es esencial en un sistema recomendador que busca fundamentar sus decisiones en análisis estadísticos sólidos.



Esta obra está bajo una licencia internacional Creative Commons Atribución-NoComercial-SinDerivadas 4.0. Los autores mantienen los derechos sobre los artículos y por tanto son libres de compartir, copiar, distribuir, ejecutar y comunicar públicamente la obra.

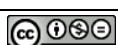
REFERENCIAS BIBLIOGRÁFICAS

- (BID), & (ILPES). (2022). *Material docente sobre gestión y control de proyectos: Programa de capacitación. Comisión Económica para América Latina y el Caribe (CEPAL)*. Santiago de Chile: CEPAL. Obtenido de <https://repositorio.cepal.org/server/api/core/bitstreams/71d847bf-2137-4a1b-8776-017aa056a0ca/content>
- Aguirre, J. (2 de Mayo de 2023). *Ecuavisa*. Obtenido de <https://www.ecuavisa.com/noticias/ecuador/los-5-defectos-de-la-contratacion-publica-en-ecuador-EI5039732>
- Bravo, J., & Brauner, D. (2022). *Aplicación para licitaciones públicas de código abierto*. Bogotá: Universidad del Bosque, revista de Innovación Pública y Tecnología. Obtenido de <https://repositorio.unbosque.edu.co/server/api/core/bitstreams/131a9585-6ec7-4251-b578-4be5a837315e/content>
- Cevallos, J., & Macias, D. (2022). *Desarrollo de un sistema recomendador de licitaciones basado en técnica de aprendizaje automático*. Guayaquil: Universidad de Guayaquil.
- Faz, W., Fuentes, L., Hidalgo, M., & Guerrero, K. (2023). *Government procurement in Ecuador: analysis and perspective*. Universidad Ciencia Y Tecnología. Obtenido de <https://doi.org/10.47460/uct.v27i119.714>
- García, M. (2022). *Las licitaciones públicas: análisis de datos y sistemas predictores utilizando métodos de 'machine learning'*. Obtenido de Repositorio Institucional de la Universidad de Oviedo: <http://hdl.handle.net/10651/65768>
- Hernández, J. (2018). *Principios rectores en la contratación pública*. Revista de Derecho Público.
- Hernández, N. (2020). *Transparencia y equidad en las recomendaciones basadas en IA en procesos de licitación pública*.
- Jiménez, Á., & Luna, D. (2024). *Análisis de inteligencia artificial explicable (XAI) y su aplicación en sistemas de recomendación*. Guayaquil: Universidad de Guayaquil.
- LOPDP, A. N. (2021). *Ley Orgánica de Protección de datos Personales* (Registro Oficial No. 459 ed.).
- LOTAIP, A. N. (2004). *Ley Orgánica de Transparencia y Acceso a la Información Pública* (Registro Oficial No. 337. ed.).
- Maydana Huanca, A. R. (2021). *Elección del mejor modelo entre regresión lineal múltiple y árboles de regresión para predecir el precio máximo de las acciones de Intel en función al precio de apertura y volumen de ventas de acciones por día - 2019*. Obtenido de Universidad Nacional del Altiplano: <http://repositorio.unap.edu.pe/handle/20.500.14082/15333>
- Pincay, G., & Velásquez, N. (2024). *Análisis de datos sobre licitaciones públicas de Ecuador apicando técnicas de explicabilidad de modelos*. Universidad de Guayaquil. Obtenido de <https://repositorio.ug.edu.ec/handle/redug/76137>
- Ponce, C., & Joel, V. d. (2024). *ANÁLISIS DE DATOS MEDIANTE GRÁFICOS EXPLICATIVOS PARA LA MEJORA EN LA TOMA DE DECISIONES EN LICITACIONES PÚBLICAS*. Guayaquil: Universidad de Guayaquil. Obtenido de <https://repositorio.ug.edu.ec/handle/redug/76650>
- Ramírez, S. (2023). *Implicaciones de la inteligencia artificial en la transparencia de los contratos administrativos*. Boyacá: Universidad Santo Tomás.



San Milan, J. (2020). *Predicción de precios en mercados financieros usando la información de las noticias a través de métodos estadísticos*. Universidad Pontificia de Comillas. Obtenido de <https://repositorio.comillas.edu/jspui/bitstream/11531/43711/1/TFG-SanMillanMontero%2CJavierArturo.pdf>

Smith, J., & Jones, M. (2020). *Explainable AI in public procurement systems*. European Journal of Artificial Intelligence.



Esta obra está bajo una licencia internacional Creative Commons Atribución-NoComercial-SinDerivadas 4.0. Los autores mantienen los derechos sobre los artículos y por tanto son libres de compartir, copiar, distribuir, ejecutar y comunicar públicamente la obra.