

**TECNOLOGIA SUPERIOR**

**BIG DATA E INTELIGENCIA DE NEGOCIO**

**DATA MINING**

**WILLIAM ESTUARDO JIMÉNEZ MIGUEZ**

william.jimenez@cenestur.edu.ec

**Profesor: JOHANNA CRISTINA JARA BUSTILLOS**

johanna.jara@cenestur.edu.ec

**Quito, Ecuador**

**2025**

1. **Instrucciones.**

**ACTIVIDAD PRÁCTICA 4:**

Desarrollar un modelo de clasificación utilizando un dataset proporcionado.

**ACTIVIDAD PRÁCTICA 5:**

Comparar resultados de diferentes modelos de regresión sobre el mismo conjunto de datos.

**INSTRUCCIONES:**

* Trabaje con el dataset [Fraud.csv](https://drive.google.com/file/d/1kmm-JcJeFgX89mvrbUqIS_j_-u_3pXhn/view?usp=drive_link), utilice **dos** modelos para realizar la clasificación.
* Explique el resultado de cada modelo y justifique su selección.
* Redacte un párrafo (mínimo 10 líneas) comparativo de conclusión.
* La actividad puede realizarla en Python, Rapidminer o Weka.
* ***Al final de cada actividad coloque un enlace con los programas***.

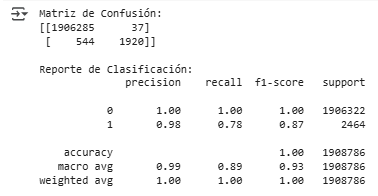
1. **Desarrollo**

**Herramienta:** Google Colab (Python)

**Modelo de Clasificación:** RandomForestClassifier

**Link:** <https://github.com/Willyejm/Deber-5-y-6.git>

***Resultado*:**

****

**Modelo de Regresión:** LinearRegression

***Resultado*:**

****

|  |  |  |
| --- | --- | --- |
| **ASPECTO** | **CLASIFICACIÓN (*RandomForest*)** | **REGRESIÓN**  **(*LinearRegression*)** |
| Variable objetivo | isFraud (binaria) | newbalanceOrig (continua) |
| Tipo de modelo | Clasificación supervisada | Regresión supervisada |
| Precisión | Muy alta (F1: 0.87) | Muy alta (R²: 0.9982) |
| Interoperabilidad | Media | Alta (modelo lineal) |
| Resistencia a desbalance | Alta | No aplica |
| Aplicación práctica | Detección de fraudes | Estimar saldo final |
| Mejor uso | Prevención de delitos financieros | Planificación financiera, validación de cálculos |

**Comparación de los modelos:**

En el análisis comparativo realizado, se aplicaron modelos de clasificación y regresión sobre el dataset Fraud.csv. El modelo de clasificación Random Forest mostró alta efectividad en la detección de fraudes, con un F1-score de 0.87 y un recall de 0.78, siendo adecuado para sistemas de seguridad financiera. Por su parte, la regresión lineal logró un excelente ajuste (R² = 0.9982) en la predicción del saldo final de las cuentas, útil para fines contables. En conclusión, ambos modelos son altamente precisos dentro de sus objetivos específicos, y su elección dependerá del propósito del análisis, prevención de fraude o estimación financiera. Integrar ambos enfoques puede aportar soluciones más completas e inteligentes en el ámbito bancario o empresarial.