

---

---

# Detección de Fraude en Logística Combinando Aprendizajes Supervisados y No Supervisados

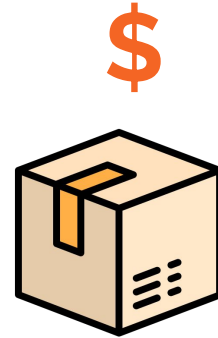
Short Paper - Comunicaciones Científicas UTN  
2022

---

# Introducción

En las plataformas de comercio electrónico, se estima el costo de los envíos de cara al destinatario (posible comprador) utilizando la información proporcionada por el remitente sobre el producto que está vendiendo.

Luego de que el envío sucede, se conoce el peso real de ese producto y por lo tanto, su costo asociado...



¿Qué pasa si el valor estimado y el real difieren?

**Se incurre en pérdidas monetarias**



¿Y qué pasaría si el remitente abusara de la confianza depositada en su declaración de contenido para obtener un beneficio monetario?

**Aparece el perfil de fraude**

# Desafío: ¿cómo separar una diferencia en pesajes normal de una fraudulenta?

¿Es lo mismo 500 gr de diferencia en un paquete de 1 kg que en uno de 5 kg? ¿Cuán frecuente es que sucedan estas diferencias?

# Método

## → Objetivo

Identificar envíos anómalos en la muestra

## → Dataset

206.000 envíos random

## → Librería python

Sklearn

## → Modelo

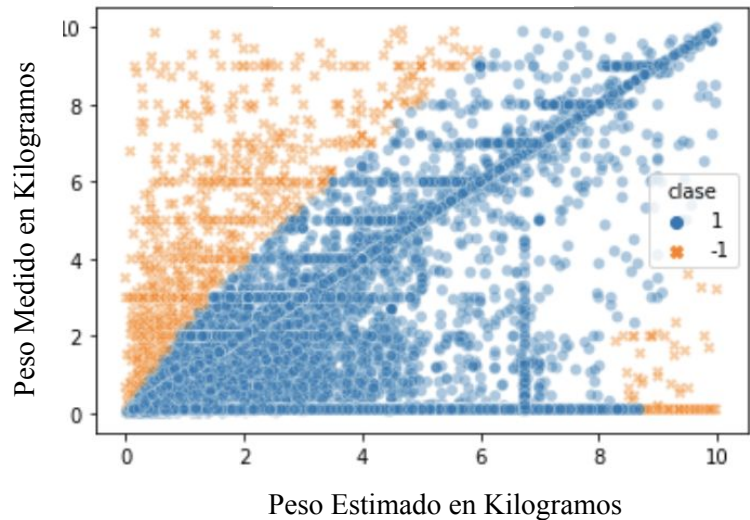
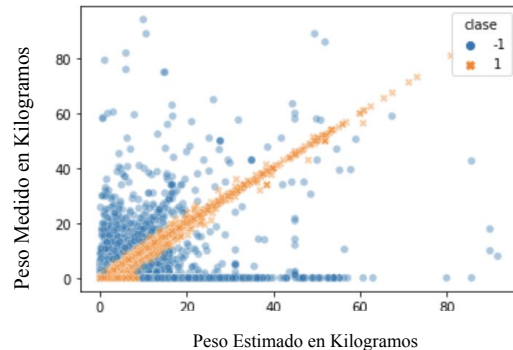
No supervisado: Isolation forest, 300 árboles, 100 muestras por árbol

## → Variables

Diferencia Absoluta=  $\text{Peso Medido} - \text{Peso Estimado}$

Ratio Pesos =  $\text{Peso Medido} / \text{Peso Estimado}$

# Resultados



# Método

## → Objetivo

Identificar envíos anómalos en la muestra

## → Dataset

206.000 envíos random

## → Librería python

Sklearn

## → Modelo

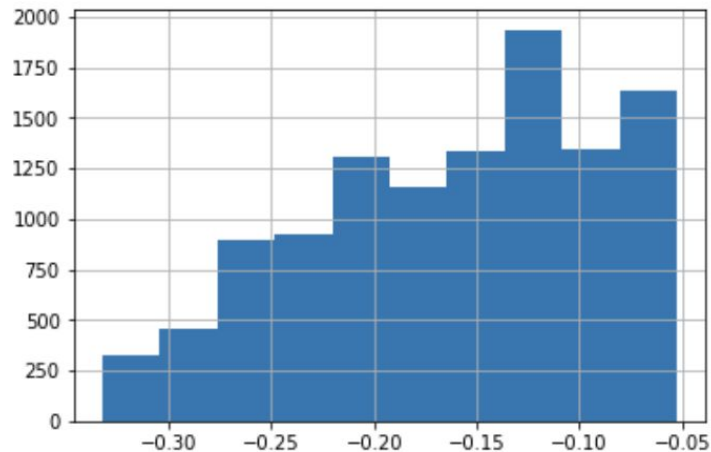
No supervisado: Isolation forest, 300 árboles, 100 muestras por árbol

## → Variables

Diferencia Absoluta= Peso Medido-Peso Estimado

Ratio Pesos =  $\text{Peso Medido} / \text{Peso Estimado}$

# Resultados



**Score Isolation Forest**

Corte: -0.15

Total de casos analizados: 206.000

Casos anómalos finales: 5.700

% casos anómalos: 2.76%

TP análisis= 100%



# Método

## → Objetivo

$$Marca_{peso} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 * V_1 + \beta_2 * V_2 + \beta_3 * V_3)}}$$

## → Dataset

Output Isolation forest

## → Librería python

statsmodel

## → Modelo

Supervisado: Regresión Logística

## → Variables

Diferencia Absoluta= Peso Medido-Peso Estimado

$$\text{Diferencia Relativa} = \frac{\text{Peso Medido} - \text{Peso Estimado}}{\text{Peso Estimado}}$$



# Resultados

## Regresión Logística:

score  $\geq$  0.99

Volumen relativo  $\approx$  3% de los envíos

Precisión  $\approx$  98%

F1  $\approx$  88%

$$Marca_{pesoV1} = \frac{1}{1 + e^{-(0.134 * dif.peso.absoluta + 1.778 * dif.peso.relativa)}}$$



# Discusión

- Sirve para analizar el falso negativo, pero no para prevenirlo



# Conclusiones

- Funcionaron los modelos para el objetivo propuesto
- Puede replicarse
- Primer trabajo en la temática