

# Proyecto final análisis de datos para la toma de decisiones

## *Tercer entregable*

*Iván A. Trujillo A.*

BIT

CODE

---

## 1 Temas a evaluar

- Metodologías de análisis (Multivariado y aprendizaje de máquina).
- Manejo de base de datos (pandas).

## 2 Criterios de evaluación

El ejercicio solo tiene un punto, pero se evaluarán los siguientes aspectos:

- Empalme de la base de datos.
- Análisis exploratorio de los datos.
- Construcción de indicadores de acuerdo a la literatura.
- Comparación de modelos y especificaciones de los mismos.
- Métricas de evaluación (performance) como estrategias de validación del modelo.

## 3 A tener en cuenta

### 3.1 Métrica

Recuerde que el uso de las métricas de desempeño también depende del fenómeno a modelar, en algunas ocasiones la sensibilidad es más importante que la especificidad, mientras que en otras no.

### 3.2 Validación

Los problemas deben tener criterios de validación. Se podría sobreestimar el desempeño de un modelo si se evalúa con el mismo conjunto de datos de entrenamiento.

### 3.3 Ejercicio (El modelo de Altman)

De acuerdo a el modelo de Altman (Puede consultar el documento Aquí).

Construir un modelo para predecir la quiebra empresarial de las firmas en Colombia usando las bases de datos disponibles en el siguiente link que son de acceso al público y que también están disponibles en el sitio oficial de la Superintendencia de Sociedades.

### 3.4 Guía

- Empalme los estados financieros de acuerdo a una identificación única.
- Defina el evento de interés de acuerdo a la aparición de la palabra liquidación en la razón social de la firma.

- Seleccione un vector de características exógenas  $X_t$  para el periodo  $t$  con el objetivo de predecir el periodo  $t + 1$ .
- Compare al menos dos modelos y defina las estrategias de validación del rendimiento de los modelos.