



ENFIN 761 BUSINESS INTELLIGENCE PARA LAS FINANZAS EXAMEN

Profesores: *David Díaz S. & Gabriel E. Cabrera*

Ayudante: *Andre Escobar*

1. Intro ML y Aprendizaje No Supervisado

Respecto de las siguientes afirmaciones, señale si son Verdaderas (V) o Falsas (F). Justifique **TODAS** sus respuestas.

- La principal diferencia entre el aprendizaje supervisado y el no supervisado, es que en este último el algoritmo no tiene acceso a las etiquetas de la data.
- El algoritmo de “Reglas de Asociación” es útil cuando queremos encontrar causalidad entre eventos de nuestra data.
- Un resultado de clustering con baja varianza inter cluster, y alta varianza intra cluster es un buen resultado.
- El reducir la dimensionalidad de la data, por ejemplo, a través del algoritmo PCA, nos permite detectar más fácilmente observaciones que se escapan de la normalidad.
- Un algoritmo no supervisado puede ser utilizado para generar data sólo si ha visto o están disponibles las etiquetas de la misma.

2. Aprendizaje Supervisado

Considere la siguiente matriz de confusión:

		Valores Reales	
		Clase 1	Clase 2
Predicción	Clase 1	1231	146
	Clase 2	250	629

- ¿Diría usted que es una muestra balanceada? Justifique su respuesta.
- Calcule el *accuracy* del modelo.
- Calcule el *recall* y *precisión* de ambas clases.
- ¿Qué puede concluir de este clasificador y sus resultados separados por clase? ¿podría mejorar el clasificador?
- Mencione y explique algunas posibles medidas de negocios en base a estos resultados ya calculados. Puede utilizar una situación hipotética para argumentar sus medidas.

3. Aplicación: Modelo de impago Bancario

Utilizando el dataset `créditos_bancarios.xlsx` cree un modelo de predicción de la variable “`credit_risk`”. Pruebe con al menos 3 algoritmos supervisados distintos, por ej, regresión logística, árboles de decisión y redes neuronales:

En sus resultados comente respecto de:

- a. Qué tipo de preprocesamientos fue necesario realizarle a los datos, especialmente respecto de si fue necesario balancear la data y qué método de balanceo ocupó y por qué
- b. Los resultados de su modelo tanto en muestra de training como de testing.
 - Cree una tabla donde se puedan comparar las métricas de recall, precisión y accuracy para ambas clases “good” y “bad” payer (pagador). Comente y justifique cual de los 3 algoritmos funcionó mejor y por qué.
- c. Realice una breve discusión de cómo se podría implementar dicho modelo en la práctica.
- d. (**Pregunta Bonus**) ¿Qué variables fueron más importantes para discriminar entre malos y buenos pagadores?