

Faltas disciplinarias graves y gravísimas:

**Artículo 113 inciso (d):** El fraude en actividades, trabajos y evaluaciones académicos y la posesión o utilización de material no autorizado en los mismos. (Falta disciplinaria grave). **Artículo 114 inciso (b):** Todas las modalidades de plagio. (Falta disciplinaria gravísima). **Artículo 114 inciso (e):** La suplantación en una evaluación académica, en exámenes preparatorios, en trabajos de grado y tesis. (Falta disciplinaria gravísima).

Sanciones Disciplinarias:

**Artículo 117:** Las faltas graves serán sancionadas con amonestación escrita con cargo a la hoja de vida del estudiante y la imposición de matrícula condicional durante el tiempo necesario para cumplir la condición. **Artículo 118:** Adicional a la sanción disciplinaria, el fraude en actividades, trabajos y evaluaciones académicos se sancionará académicamente con la pérdida de la asignatura, la cual será calificada con nota definitiva de cero punto cero (0.0).

## PARCIAL 1 – MÉTODOS Y APLICACIONES DE ANALÍTICA II Septiembre 9 de 2022, 2 horas

La Facultad de Ingeniería de la Universidad Iberoamericana (Universidad privada) está preocupada por los altos índices de deserción estudiantil (desertar es abandonar y no volver a la U) en sus programas de pregrado. Con el fin de diseñar estrategias para disminuir y/o evitar la deserción, es de vital importancia poder identificar con anticipación si un alumno que está matriculado en alguno de los programas de pregrado ofrecidos por la facultad será potencial desertor o no. La facultad ha recopilado información histórica sobre un buen número de estudiantes y graduados de la facultad. Los expertos de la Universidad han seleccionado varias variables que podrían afectar la deserción y se han tabulado los datos:

**SEX:** 1 para femenino, 2 para masculino.

**EDAD:** Edad a la que ingresó el estudiante a su primer semestre.

**EST:** Estrato socioeconómico.

**PROM:** Promedio académico acumulado del estudiante.

**MAT:** Número de materias que ha tenido que repetir el estudiante.

**PRO:** Programa académico de pregrado que cursa el estudiante.

**SEM:** Semestre del año en que ingresó a la universidad. 1 para primer semestre del año, 2 para segundo semestre del año.

**MATP:** número promedio de materias cursadas por semestre.

**TRA:** 1 si el estudiante trabaja mientras estudia, 2 si el estudiante no trabaja.

**FIN:** 1 si el estudiante paga su matrícula por medio de un crédito educativo directo con la universidad. 2 en cualquier otro caso.

**COL:** 1 si el colegio del que egresó es público o 2 si este es privado.

**ING:** La universidad aplica una prueba diagnóstica de ingreso cuyo puntaje puede ir entre 0 y 70 puntos (variable continua).

**DES:** 0 si el estudiante desertó o 1 si el estudiante se graduó.

1. (0.2 puntos) Convierta cada variable al tipo adecuado, solo se permite “int64”, “float64” y “category”.
2. (0.2 puntos) Genere las variables dummie que requiera.
3. (0.1 puntos) Defina el vector **y** (objetivo) y el vector **X** (características).
4. (0.2 puntos) Use el 70% de los datos para entrenamiento y el 30% de los datos para prueba (test). Haga la división de forma estratificada (stratify=y). **para random\_state use los 5 últimos dígitos de su ID.**
5. (0.2 puntos) Escale sus X de entrenamiento y sus X de prueba (no es necesario escalar **y** de entrenamiento o de prueba). Escalar es diferente a estandarizar.
6. Ajuste un primer modelo SVM con Kernel lineal, con C=1000 y **para random\_state use los 5 últimos dígitos de su ID. Para este modelo identifique:**
  - a) (0.3 puntos) El número de vectores de soporte del modelo
  - b) (0.5 puntos) La ecuación completa del hiperplano de separación resultante
  - c) (0.6 puntos) Las **tres variables** más influyentes a la hora de clasificar a un estudiante como DESERTOR. **Justifique su elección.**
  - d) (0.3 puntos) La anchura del margen
  - e) (0.5 puntos) Calcule la matriz de confusión para los datos de prueba e **interprete en contexto** dos de las métricas de la matriz.

f) (1 punto) Se tiene una observación nueva de un estudiante:

| SEX | EDAD | EST | PROM | MAT | PRO                 | SEM | MATP | TRA | FIN | COL | ING  |
|-----|------|-----|------|-----|---------------------|-----|------|-----|-----|-----|------|
| 2   | 20   | 2   | 3,21 | 4   | INGENIERIA<br>CIVIL | 2   | 5,5  | 1   | 2   | 1   | 35,8 |

Clasifique a este estudiante a partir de su modelo ajustado, **sin ayuda de Python**. Trabaje con los **máximos y mínimos** de cada variable dentro de la base de datos original (**aunque no necesariamente sean los mismos máximos y mínimos de los datos de entrenamiento del modelo**). Es posible que deba recordar la formula con la cual se escalan datos entre 0 y 1. Muestre el procedimiento realizado.

7. (0.6 puntos) Para la Facultad es muy importante identificar correctamente **a la mayor cantidad de estudiantes que SERÁN REALMENTE DESERTORES**, para poder diseñar las estrategias que sean necesarias. Ajuste un nuevo modelo e indique claramente todos los parámetros que utilizó (diferentes a los que están por defecto) que mejore esta medida de desempeño con respecto al modelo del punto 6. **para random\_state use los 5 últimos dígitos de su ID.**
8. (0.3 puntos) **Este punto es totalmente independiente de los anteriores:** ¿Siendo usted un experto en educación que variables de la base de datos eliminaría desde el principio y porqué (dos variables)? ¿Qué variables agregaría a la base de datos y porqué (mínimo dos)?