# REGLAMENTO DE ESTUDIANTES de la PONTIFICIA UNIVERSIDAD JAVERIANA

*Faltas disciplinarias graves y gravísimas:*

Artículo 123 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 124 inciso (b): Todas las modalidades de plagio. (Falta disciplinaria gravísima)

Artículo 124 inciso (e): La suplantación en una evaluación académica, en exámenes preparatorios, en trabajos de grado y tesis. (Falta disciplinaria gravísima)

*Sanciones Disciplinarias:*

Artículo 127: 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 128: 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).

NO SE PERMITE INTERCAMBIO DE INFORMACIÓN CON PERSONA ALGUNA DURANTE EL DESARROLLO DEL PARCIAL. SOLO PUEDE INTERACTUAR CON EL PROFESOR. PARA COMUNICARSE CON EL PROFESOR, ENVÍELE UN MENSAJE PRIVADO POR TEAMS.

IMPORTANTE

* Debe generar un notebook para cada punto, responda las preguntas planteadas dentro de una celda de texto en el notebook. Nombre cada notebook así: “**Punto X – nombre del estudiante**”. Antes de que se cumpla el plazo límite, debe cargar los 3 notebooks en el buzón “Parcial 1” configurado en la plataforma de BS. Si se cargan después del plazo, el parcial se califica sobre 4.0 si es en los siguientes 30 minutos. Después de estos 30 minutos ya no se recibirán parciales.
* Para el entrenamiento y prueba de los modelos, **solo puede utilizar los comandos vistos en clase**.
* Para todos los comandos que involucren una selección aleatoria, por ejemplo, dividir el dataset entre entrenamiento y prueba, o seleccionar registros de un dataset, fije siempre la semilla con los **últimos 3 números de su cédula** de ciudadanía.
* Para evaluar el desempeño de sus modelos, sobre los datasets de prueba, utilice en problemas de clasificación**: accuracy, y F-Score** para cada clase; en problemas de regresión, **RMSE y R2**.
* Si hay datos faltantes simplemente elimine los registros donde se presenten estos casos.

**PUNTO 1 (20). ANÁLISIS DE COMPONENTES PRINCIPALES**

La base de datos HBPAPER.xlsx consta de 100 observaciones sobre 18 variables distintas, se basa en un estudio de segmentación del mercado de los clientes de la compañía HBPAPER. Esta compañía vende productos de papel a dos segmentos de mercado: la industria del papel para periódicos y la industria de las revistas. Además, los productos de papel se venden a estos segmentos de mercado directamente al cliente o indirectamente a través de un intermediario. En las encuestas se recolectan dos tipos de información:

* El primer tipo de información es la percepción del rendimiento de HBPAPER en 13 atributos. Estos atributos, desarrollados a través de grupos focales, una prueba previa y el uso en estudios anteriores, se consideran los más importantes para los clientes de HBAPER. Se considera que estos atributos son los que más influyen en la selección de proveedores en la industria papelera. Entre los encuestados había directores de compras de empresas que compraban a HBAPER, y evaluaron esta compañía en cada uno de los 13 atributos utilizando una escala de 0-10, siendo 10 «Excelente» y 0 «Deficiente».
* El segundo tipo de información se refiere a los resultados de la compra y las relaciones comerciales (por ejemplo, la satisfacción con HBAPER y si la empresa está satisfecha con el servicio y si la empresa consideraría una alianza o asociación estratégica con HBAPER).

Un tercer tipo de información procede del almacén de datos de HBAPER e incluye datos como el tamaño del cliente y la antigüedad de su relación de compras con HBAPER.

A continuación, se presentan las variables, junto con sus descripciones:

**Variables de clasificación del almacén de datos**

* X1 Tipo de cliente
* X2 Tipo de industria
* X3 Tamaño de la empresa
* X4 Región
* X5 Sistema de distribución

**Variables de percepción del rendimiento**

* X6 Métrica de calidad del producto
* X7 Métrica de actividades de comercio electrónico/página web
* X8 Medida de la asistencia técnica
* X9 Métrica de resolución de reclamaciones
* X10 Medida de la publicidad
* X11 Métrica de la línea de productos
* X12 Métrica de imagen de la fuerza de ventas
* X13 Métrica de precios competitivos
* X14 Métrica de garantías y reclamaciones
* X15 Métrica de nuevos productos
* X16 Métrica de pedidos y facturación
* X17 Métrica de flexibilidad de precios
* X18 Métrica de la velocidad de entrega

**Medidas de resultados/relaciones**

* X19 Medición de la satisfacción
* X20 Medida de la probabilidad de recomendación
* X21 Medida de la probabilidad de futuras compras
* X22 Medida del nivel de compra/uso actual
* X23 Considera la posibilidad de una alianza o asociación estratégica en el futuro, 1 = SI, 0 = NO

Se desea llevar a cabo un análisis factorial – con base en Análisis de Componentes Principales (ACP) – de las variables **X6 a X18**, que son las variables que explican las percepciones del rendimiento. Aplique la metodología vista en clase para resolver las siguientes preguntas:

1. (3) Aplique el test apropiado para determinar un apropiado MSA (Medida de adecuación de la muestra). Busque que todas las variables tengan un MSA mínimo de 0.5. ¿Cuál es el MSA final para el dataset final?
2. (3) Aplique el test apropiado para comprobar que las correlaciones entre variables son suficientes para aplicar ACP. Explique los resultados que arroja este test.
3. (4) Aplique los 3 criterios para determinar el número de componentes principales. ¿Cuál de los 3 escogería? Justifique su respuesta.
4. (3) Analice la solución con 4 componentes principales. ¿Cree que con esta solución se podría identificar claramente los factores agrupando variables originales? Justifique su respuesta.
5. (7) Aplique una rotación VARIMAX. Identifique los factores que agrupen las respectivas variables. Plantee un nombre apropiado para cada factor, desde el punto de vista del negocio.

**PUNTO 2 (30). REGRESIÓN LOGÍSTICA**

Utilizaremos un conjunto de datos de Recursos Humanos que consta de 10 variables de entrada y 1 de salida para determinar si un empleado va a renunciar o no.

Las variables de entrada son las siguientes:

* Porcentaje de retraso
* Porcentaje de iniciativa en proyectos
* Porcentaje de entrega de proyectos a tiempo
* Porcentaje de correos electrónicos intercambiados
* Porcentaje de receptividad
* Porcentaje de respuesta profesional a correos electrónicos
* Porcentaje de compartir ideas
* Porcentaje de ayudar a compañeros
* Porcentaje de publicaciones empresariales en LinkedIn
* Porcentaje de comentarios en Facebook

La variable de salida se denomina «Renuncia» y es una clase binaria en la que 0 representa que un empleado se queda y 1 que renuncia.

Se desea desarrollar un modelo de regresión logística para predecir la variable de salida con base en las variables de entrada. Desarrolle un modelo stepwise usando un corte de 0.5 (vale 5 puntos) y responda las siguientes preguntas:

1. (3) ¿De las variables presentes en el modelo final, si tuviera que eliminar una variable de las que quedaron en el modelo final, cuál eliminaría y porqué?
2. (6) Tomando como probabilidad base la probabilidad de renuncia, determine la probabilidad final que se obtiene al mostrar un aumento de 3 puntos porcentuales (por ejemplo, subir de 10% a 15%) en la probabilidad de retraso, suponiendo que las demás variables no tienen ningún cambio.
3. (3) Se selecciona un grupo de 2000 personas para evaluar la implementación de la estrategia. Se ha estimado que por cada persona categorizada correctamente se estima un ahorro promedio de $800, mientras que por cada persona mal categorizada se estima un costo de $350. Basado en corte de 0.5, ¿cuál sería la ganancia esperada en términos monetarios para este modelo?
4. (5) Esta pregunta es conceptual: Si hipotéticamente el cutoff se disminuyera a 0.3, ¿cómo afectaría el *precision* y el *recall*? ¿Afectaría este cambio el valor de AUC? Justifique sus respuestas.
5. (8) Con el fin de mejorar el desempeño del modelo actual, desarrolle un modelo ELASTIC NET, escogiendo adecuadamente los parámetros alfa y lambda que conduzcan al mejor desempeño posible. ¿Cómo lo compra con el modelo sin regularización?