# Pontificia Universidad Javeriana - Wikipedia, la enciclopedia libre

**APRENDIZAJE DE MÁQUINA** - **PARCIAL 1**

**Fecha: marzo 7 de 2024. Duración: 2.5 horas**

# 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 EL USO DE HERRAMIENTAS DE CORREO, CHAT, DRIVES. SOLO PUEDE USAR LA PLATAFORMA DE BS ASÍ COMO GOOGLE PARA CONSULTA DE INFORMACIÓN GENERAL.

NO SE PERMITE USO DE CELULARES O TABLETS. PUEDE USER APUNTES DE CLASE.

IMPORTANTE

* Debe generar un notebook para los puntos 1 y 2, responda las preguntas planteadas dentro de una celda de texto en el notebook. Nombre cada notebook así: “**Punto X – nombre del estudiante**”. Uno de los 2 notebooks se deberá actualizar y cargar nuevamente según las instrucciones del punto 3. **Plazo para cargar todos los notebooks: 10 am.**
* Para el entrenamiento y prueba de los modelos, **solo puede utilizar los comandos vistos en clase en R**.
* 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.
* Debe preparar los datos adecuadamente para ser usados con el modelo solicitado.
* Si hay datos faltantes simplemente elimine los registros donde se presenten estos casos.

**PUNTO 1 (20). K-NEAREST NEIGHBORS**

Con la temporada navideña acercándose rápidamente, la organización liderada por Santa Claus está en proceso de preparar su equipo de manufactura para determinar qué niños recibirán regalos este año. El departamento encargado de tomar esta decisión, Delivery Gifts, ha buscado la ayuda de expertos y expertas de la Universidad Javeriana para abordar este desafío para la temporada de 2025, utilizando datos históricos recopilados en los años 2023 y 2024. Con un conjunto de datos compuesto por 15 variables de predicción, 1 variable de respuesta y 250 registros históricos (**gifts\_historico.xslx**) se le ha encomendado la tarea de ayudar a Delivery Gifts a crear un modelo de analítica para clasificar nuevos niños para la próxima Navidad. Desarrolle un modelo KNN para predecir quienes deben recibir regalo e indicar y evaluar las métricas obtenidas. La variable a predecir es **reciberegalo.**

Preguntas:

1. (6) Encuentre el mejor valor de K desde el punto de vista de ACCURACY. Evalúe el desempeño del modelo con dicho valor.
2. (7) Plantee un valor de K que lleve a un *overfitting*. Explique porqué se da el *overfitting*.
3. (7) Plantee un valor de K que lleve a un *underfitting*. Explique porqué se da el *underfitting*.

**PUNTO 2 (12). MODELO NAIVE BAYES**

La base de datos *“caso1a.xlsx”* contiene el registro de 7261 clientes actuales de un supermercado con las siguientes variables:

* **Genero**: género declarado por la persona
* **Estadocivil**: estado civil declarado por la persona
* **Edad**: Edad declarada por la persona
* **VIviendaPropia**: información traída de una base externa sobre si la persona posee o no vivienda propia
* **Compra**: el resultado de compra a una oferta de seguros de hogar (contenidos de la casa y reparaciones).

Desarrolle un modelo Naive Bayes para predecir la posible compra de futuros clientes de los que se posee la misma información.

Preguntas:

1. (12) Dentro de su modelo, ensaye 3 valores del umbral de probabilidad, para decidir si una persona va a realizar una compra de un seguro de hogar. ¿Cuál valor ofrece el mejor desempeño?

**PUNTO 3 (18). CURVA ROC – AUC**

Seleccione uno de los modelos desarrollados anteriormente. Varíe el umbral decisorio, desde 0 hasta 1, en incrementos de 0.1. En la hoja provista, va a escribir la matriz de confusión para cada umbral, y a partir de esta, muestre los cálculos requeridos para el punto que irá en la curva ROC, la cual deberá ser graficada en otra parte de la hoja provista. El entregable es la hoja con la curva ROC graficada y las memorias de cálculo correspondiente (12)

En el notebook usado para desarrollar el modelo, genere la curva ROC y compárela con la que graficó en la hoja (6). Vuelva a cargar el notebook en la plataforma (6).