**Tarea: Reforzamiento de Estadística en Machine Learning**

**Introducción**

En el mundo del Machine Learning, el manejo adecuado de conceptos estadísticos es esencial para la construcción, validación e interpretación de modelos predictivos. Esta tarea tiene como objetivo repasar y profundizar en varios de estos conceptos, que van desde fundamentos teóricos, como el teorema del límite central, hasta aspectos prácticos en la evaluación de modelos, como la regresión lineal y sus métricas. Además, se abordarán temas críticos relacionados con la inferencia estadística y la integridad de los datos, tales como los errores de tipo I y II, la selección de bias y la probabilidad binomial.

A través de esta actividad, se espera que desarrolles un entendimiento sólido de cada uno de los temas y que seas capaz de articular cómo se aplican en el análisis de datos y en la implementación de algoritmos de Machine Learning.

**Temas a Desarrollar**

**1. El Teorema del Límite Central (Central Limit Theorem)**

* **Definición:** Explica en tus propias palabras qué es el teorema del límite central.
* **Importancia:** Describe por qué este teorema es fundamental en la inferencia estadística y en el análisis de muestras.
* **Aplicación Práctica:** Menciona ejemplos o situaciones en Machine Learning donde se aprovecha el hecho de que, bajo ciertas condiciones, la distribución muestral se aproxima a una distribución normal.

**2. Sampling (Muestreo)**

* **Definición:** Define qué se entiende por muestreo en el contexto de la estadística.
* **Tipos de Muestreo:** Describe brevemente algunos métodos de muestreo (por ejemplo, aleatorio simple, estratificado, sistemático, etc.).
* **Relevancia en Machine Learning:** Explica la importancia de realizar un muestreo adecuado para entrenar y validar modelos de Machine Learning, considerando aspectos como la representatividad y la variabilidad de los datos.

**3. Diferencia entre Error Tipo I y Error Tipo II**

* **Error Tipo I:** Define qué es un error tipo I (falso positivo) y proporciona un ejemplo.
* **Error Tipo II:** Define qué es un error tipo II (falso negativo) y proporciona un ejemplo.
* **Comparación:** Discute las implicaciones de cometer cada uno de estos errores en la toma de decisiones y en la implementación de modelos predictivos.

**4. Regresión Lineal y sus Métricas**

* **Definición de Regresión Lineal:** Explica qué es la regresión lineal y cómo se utiliza en Machine Learning.
* **Métricas de Evaluación:**
  + **p-value:** Define qué es y cómo se interpreta para determinar la significancia de las variables en el modelo.
  + **Coeficientes:** Explica el significado de los coeficientes (pendientes e intercepto) y su interpretación en el contexto del modelo.
  + **R-squared (R²):** Describe qué representa el R² en términos de la variabilidad explicada por el modelo.
* **Ejemplo Práctico:** (Opcional) Incluye un ejemplo en el que se muestre la aplicación de la regresión lineal y el análisis de sus métricas.

**5. Estadística por Iteración**

* **Definición:** Describe qué se entiende por estadística por iteración. Puedes referirte a métodos como el bootstrapping o la aplicación de algoritmos iterativos para la estimación de parámetros.
* **Aplicación en Machine Learning:** Explica cómo estos métodos iterativos pueden ayudar a mejorar la robustez de las inferencias estadísticas y la estimación de parámetros en modelos complejos.
* **Ventajas y Desventajas:** Discute brevemente los beneficios y las limitaciones de utilizar enfoques iterativos en la práctica.

**6. Selección de Bias (Sesgo de Selección)**

* **Definición:** Define qué es el sesgo de selección y cómo puede afectar la validez de los estudios y modelos de Machine Learning.
* **Impacto:** Explica las consecuencias de trabajar con datos sesgados en el entrenamiento y evaluación de modelos.
* **Estrategias de Mitigación:** Propón al menos dos métodos o estrategias para minimizar el sesgo de selección en proyectos de análisis de datos.

**7. Probabilidad Binomial**

* **Definición:** Describe qué es la distribución binomial y en qué consiste la probabilidad binomial.
* **Fórmula y Componentes:** Explica la fórmula de la distribución binomial y el significado de cada uno de sus componentes.
* **Aplicación Práctica:** Ofrece un ejemplo donde la probabilidad binomial se aplique en un contexto relacionado con Machine Learning, por ejemplo, en la evaluación de un modelo de clasificación.

**Parámetros de Calificación**

La evaluación de esta tarea se realizará en base a los siguientes criterios:

* **Claridad y Precisión en las Definiciones (30%):**

Cada concepto debe estar explicado de forma clara y precisa, utilizando la terminología adecuada y demostrando un buen entendimiento teórico.

* **Profundidad del Análisis (30%):**

Se valorará la capacidad para conectar los conceptos teóricos con aplicaciones prácticas en Machine Learning, así como la inclusión de ejemplos o casos de uso relevantes.

* **Aplicación Práctica y Ejemplos (20%):**

Se evaluará el uso de ejemplos, gráficos o ejercicios prácticos que demuestren la correcta aplicación de los conceptos presentados.

* **Estructura y Redacción (10%):**

La tarea debe estar organizada de manera lógica y coherente, con una introducción, desarrollo y (opcional) conclusión. Se considerará la ortografía y la calidad de la redacción.

* **Originalidad y Pensamiento Crítico (10%):**

Se valorará la capacidad para aportar ideas propias y críticas sobre la aplicación de los conceptos en escenarios reales, demostrando un enfoque reflexivo y analítico.

**Entrega**

* **Formato:** La tarea deberá entregarse en formato PDF o como documento de Word.
* **Fecha Límite:** 19 de febrero 2025
* **Medio de Entrega:** Carpeta del alumno