**Transcripción corregida y estructurada de la retroalimentación**

**Profesor:**  
La predicción consiste en analizar qué tan cercano está el modelo al valor real. Por eso, cuando se repite el modelo varias veces, se puede medir su desempeño en términos de precisión.  
Si comparas los resultados de cada modelo, podrás identificar cuál presenta menor error y, por tanto, mayor exactitud predictiva.

**Profesor:**  
El objetivo general está bien planteado: *desarrollar y comparar un esquema predictivo desde el punto de vista de la calidad*, mostrando la evolución de tres modelos principales en la investigación.

**Profesor:**  
Sin embargo, los **objetivos específicos** deben estar alineados directamente con ese objetivo general.  
Por ejemplo:

* El **primero**, sobre el preprocesamiento de la data, está bien.
* El **tercero**, que trata de la evaluación de resultados, también está correcto.
* Pero el que dice *“identificar las variables con mayor fuerza de actividad”* no responde directamente al objetivo general.

En otras palabras, está bien querer identificar las variables más influyentes, pero ese propósito no se refleja en el problema principal.

**Estudiante:**  
Podríamos anotarlo porque también era parte de lo que queríamos, aunque quizá faltó incluirlo explícitamente en el objetivo general.

**Profesor:**  
No hace falta modificar el objetivo general; pueden incluir esa parte dentro de los específicos.  
También revisen el tercer objetivo: *“Evaluación comparativa del desempeño predictivo mediante métricas generalizadas”*.  
Ahí se puede incorporar la validación del modelo.

**Profesor:**  
Recuerden que la validación corresponde a la comparación del desempeño. Si dividen el dataset (por ejemplo, 70% para entrenamiento y 30% para validación), pueden especificar que se validará con los datos más recientes.

**Profesor:**  
Entonces, podrían simplificar esa parte. La evaluación comparativa y la validación del modelo son etapas del mismo proceso.  
No es necesario repetirlo como dos objetivos separados.

**Profesor:**  
Cuando ustedes validan, recuerden que imputan los datos por épocas (epochs). Podrían simplemente dejar una frase que indique que se hace la evaluación comparativa del desempeño predictivo, sin redundar.

**Profesor:**  
En su caso, la validación se hace con el *test set*, mientras que la evaluación de desempeño se hace sobre el *train/validation set*. Si lo explican así, queda más claro.

**Profesor:**  
Ahora, respecto al enfoque temporal: ustedes ingresan los datos hasta un tiempo (por ejemplo, de enero de 2020 a diciembre de 2024) y esperan predecir el valor en (por ejemplo, el primer trimestre de 2025).

Por eso deben dejar claro **qué variables de entrada usan (features)** y **qué variable están prediciendo (target)**.  
Si el modelo usa variables como tasas, índices o condiciones climáticas, deben precisar cuál de ellas corresponde al tiempo futuro.

**Profesor:**  
Por ejemplo, si su modelo aprende los patrones de natalidad desde 2020 hasta 2024, entonces la predicción sería la tasa de natalidad del primer trimestre de 2025.

**Profesor:**  
También podrían mostrarlo con un gráfico de línea temporal para visualizar la predicción del siguiente periodo.

**Profesor:**  
Cuando mencionan el **dataset**, asegúrense de aclarar cómo lo agrupan: por región, por trimestre, o por año.  
Así, podrán hacer que el modelo trabaje con datos estructurados y comparables entre regiones.

**Profesor:**  
Tengan en cuenta que el dataset tiene dependencias temporales; por eso el *cross-validation* tradicional no es aplicable.  
Deben usar una validación temporal (por ejemplo, dividir por años o trimestres).

**Profesor:**  
Ahora, sobre los modelos:  
Pueden usar **KNN**, pero expliquen **por qué su desempeño es bajo**.  
El KNN mide distancias entre puntos, y si los datos no están correctamente escalados o si hay pocas observaciones, el resultado puede no ser bueno.  
Lo importante es justificar teóricamente por qué cada modelo funciona o no.

**Profesor:**  
No basta con decir “el modelo no funcionó”. Deben explicar el porqué, relacionando la teoría con los resultados.

**Profesor:**  
En cuanto al **enfoque del trabajo**, deben decidir si es una **comparación de modelos** o el **desarrollo de un modelo óptimo**.  
Si buscan un RMSE alto (por ejemplo, mayor a 75 u 80%), eso sería desarrollo de modelo óptimo.  
Si su intención es comparar varios modelos y analizar cuál tiene mejor desempeño, deben reflejarlo en el **título y los objetivos**.

**Estudiante:**  
Sí, creo que nuestro trabajo es más de comparación.

**Profesor:**  
Entonces, el título podría ajustarse, por ejemplo:

“Predicción de tasas de natalidad mediante comparación de modelos de aprendizaje automático.”

**Profesor:**  
Otra recomendación: cuando presenten sus resultados, muestren también el **tiempo de entrenamiento**, la **velocidad de aprendizaje** y el **rendimiento final** de cada modelo.  
Podrían incluir un gráfico tipo *boxplot* comparando los indicadores (RMSE, MAE, tiempo de ejecución, etc.).

**Profesor:**  
Me gustó mucho la parte de la introducción bibliográfica; se nota que han extraído conceptos del material del curso.  
Solo procuren conectar la parte técnica con una **aplicación real**.

**Estudiante:**  
Sí, un profesor nos comentó que podríamos enfocar el trabajo en predecir tasas de natalidad anómalas (muy altas o muy bajas).

**Profesor:**  
Exacto. Si encuentran comportamientos anómalos o picos de natalidad, pueden relacionarlo con medidas de salud pública o factores socioeconómicos.

**Profesor:**  
Incluso podrían usar transformaciones (por ejemplo, logaritmos) para estabilizar distribuciones no normales y facilitar el modelado.

**Profesor:**  
En la sustentación final, expliquen cómo los modelos pueden **aplicarse en la realidad**, especialmente en contextos sociales o gubernamentales.  
Por ejemplo, si detectan un aumento en la natalidad en una provincia, el Estado podría anticipar medidas de infraestructura o salud.

**Profesor:**  
Recuerden: su proyecto no solo debe tener sustento técnico, sino también **impacto social**.  
Eso lo hace más completo.

**Profesor:**  
Cuando preparen la difusión o presentación final, conecten ambos enfoques —el técnico y el social—.  
Ese equilibrio es lo que diferencia un buen proyecto académico de uno sobresaliente.

**Profesor (cierre):**  
En general, su trabajo está bien encaminado. Los objetivos necesitan un poco de ajuste para evitar redundancias, y deben precisar mejor la validación temporal.  
Pero están avanzados y el proyecto tiene potencial.

El segundo JP nos acoto lo siguiente:

natalidad a nivel de departamento, porque de ser a nivel nacional nos vamos a rayar. Analizar en que ddepartamente hay mayor cantidad de data(mas de 3 mil registros) y tenga menos puntos críticos. Podemos hacer una análisis de todos los ubigeosf