1. Exploración del problema
2. Obtención de datos
    a. Obtener el train
3. Análisis exploratorio de los datos
    a. Entender el dataset (columnas, significado)
    b. Tipos de datos (categóricos? Float? Texto?)
    c. % Nulos (>10%? Eliminar)
    d. Análisis univariado
        i. Distribución de cada variable (está centrada? Se asemeja a una normal? Simetría? Curtosis?
        ii. Estadística de la variable
        iii. Relación de esa variable vs Pronóstico
    e. Análisis Multivariado
        i. Una variable vs otra variable
        ii. Correlaciones contra la variable pronóstico-→ las variables con más correlación, son las que se llevan al modelo a utilizar.
        iii. Correlaciones entre todas las variables.
    f. Preseleccionar o idea, de cuáles son las variables más importantes del dataset. (las más relacionadas con el pronóstico) y aquellas que sean “únicas”. Ojo, todo este análisis se llama “feature selecction” y “feature engineering”.
4. Limpieza de los datos
    a. Eliminar duplicados
    b. Corregir outliers (opcional)
    c. Completar los valores faltantes (nulos reemplazar por la media/mediana/moda.
    d. Eliminar filas duplicadas
5. Selección de modelos
    a. Entrenar muchos modelos rápidas usando parámetros estándar (Random forest, regresión logística, knn, etc)
    b. Medir los resultados (con la métrica correspondiente. Ej, clasificación)
    c. Para cada modelo usamos CV (cross validation) y analizamos la media y el desvío estándar de la medida de rendimiento en cada evaluación. (opcional).
    d. Realizar una selección de atributos más importantes
    e. Realizar una o dos iteraciones de los pasos anteriores. (opcional)
    f. Hacer una lista de los 2 o 3 modelos más “prometedores”. Lo ideal es elegir el modelo que cometan diversos tipos de errores (diversidad de errores).
6. Afinar los modelos
    a. Ajustar hiperparámetros (Grid Search/CV)
    b. Probar los métodos de ensamble (si combino diferentes modelos, por lo general se obtiene un mejor desempeño que si los utilizo individualmente).
    c. Una vez que estamos seguros del modelo final, medimos el rendimiento en el TEST y miden el error.
7. Interpretación del modelo
    a. Cuáles son las características más importantes?
    b. Cuánto contribuye cada una? Pesos.
    c. Mirar el error. ¿por qué se puede dar? Outliers?, clase esté desbalanceada? Errores en los datos?
    d. Ajusto los que haga falta y vuelvo a medir.