Dibujo de una persona

Descripción generada automáticamente con confianza bajaInstituto Politécnico Nacional

Escuela Superior de Cómputo

**Integrante:**

Torres Abonce Luis Miguel

Salazar Carreón Jeshua Jonatan

**Grupo:**

6CV1

**Materia:**

Machine Learning

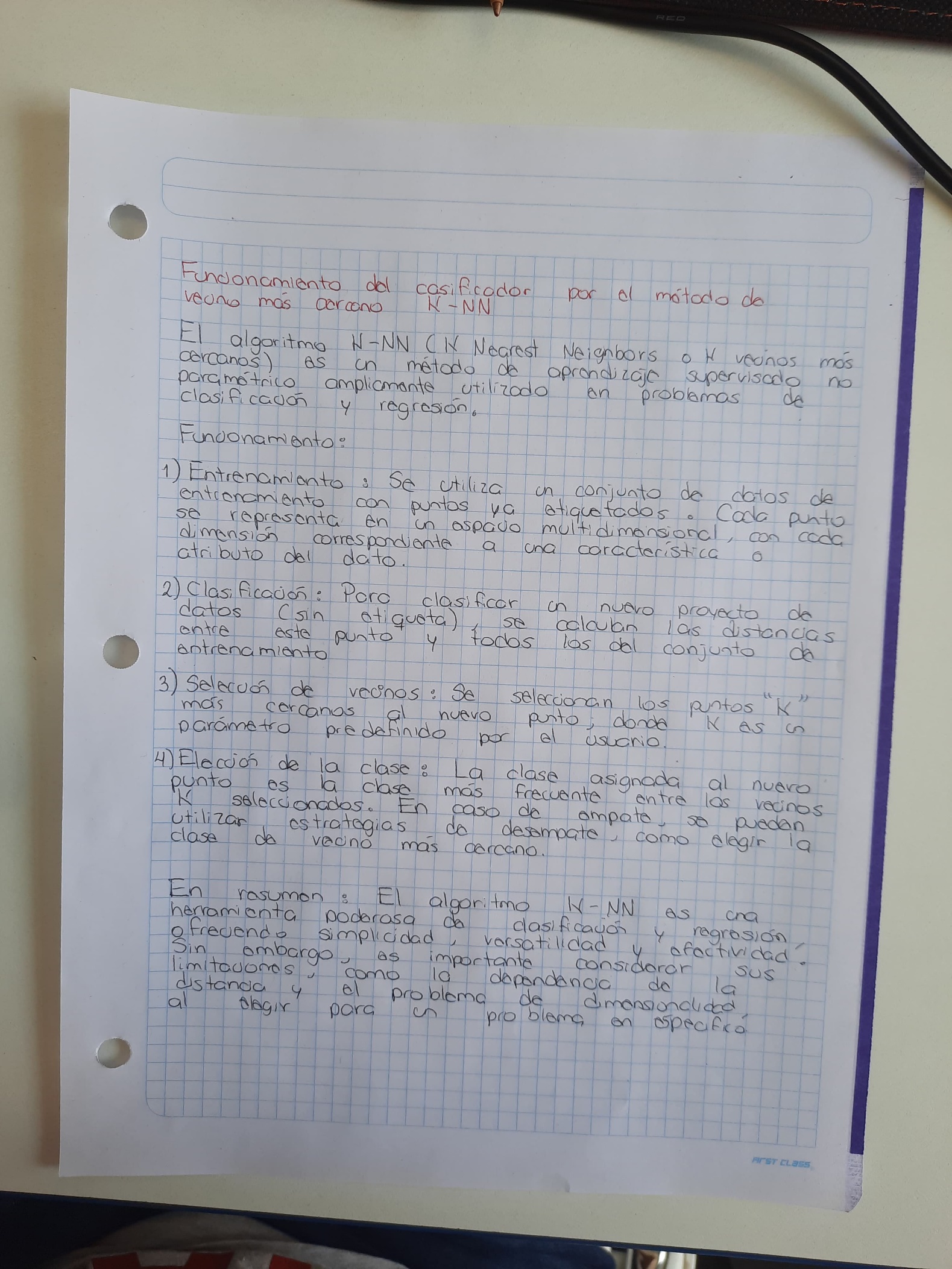
**Profesor:**

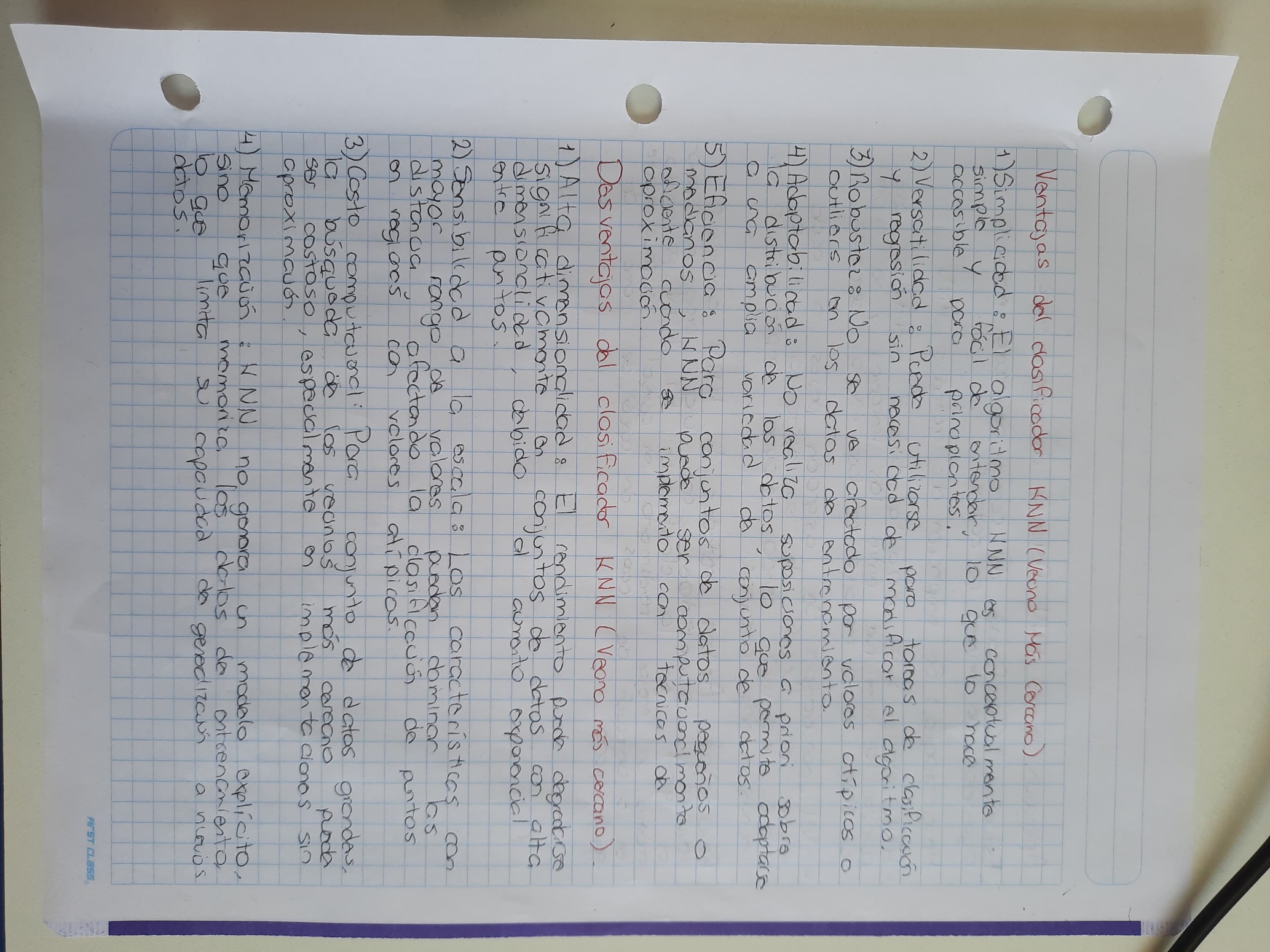
Camacho Vazquez Vanessa Alejandra

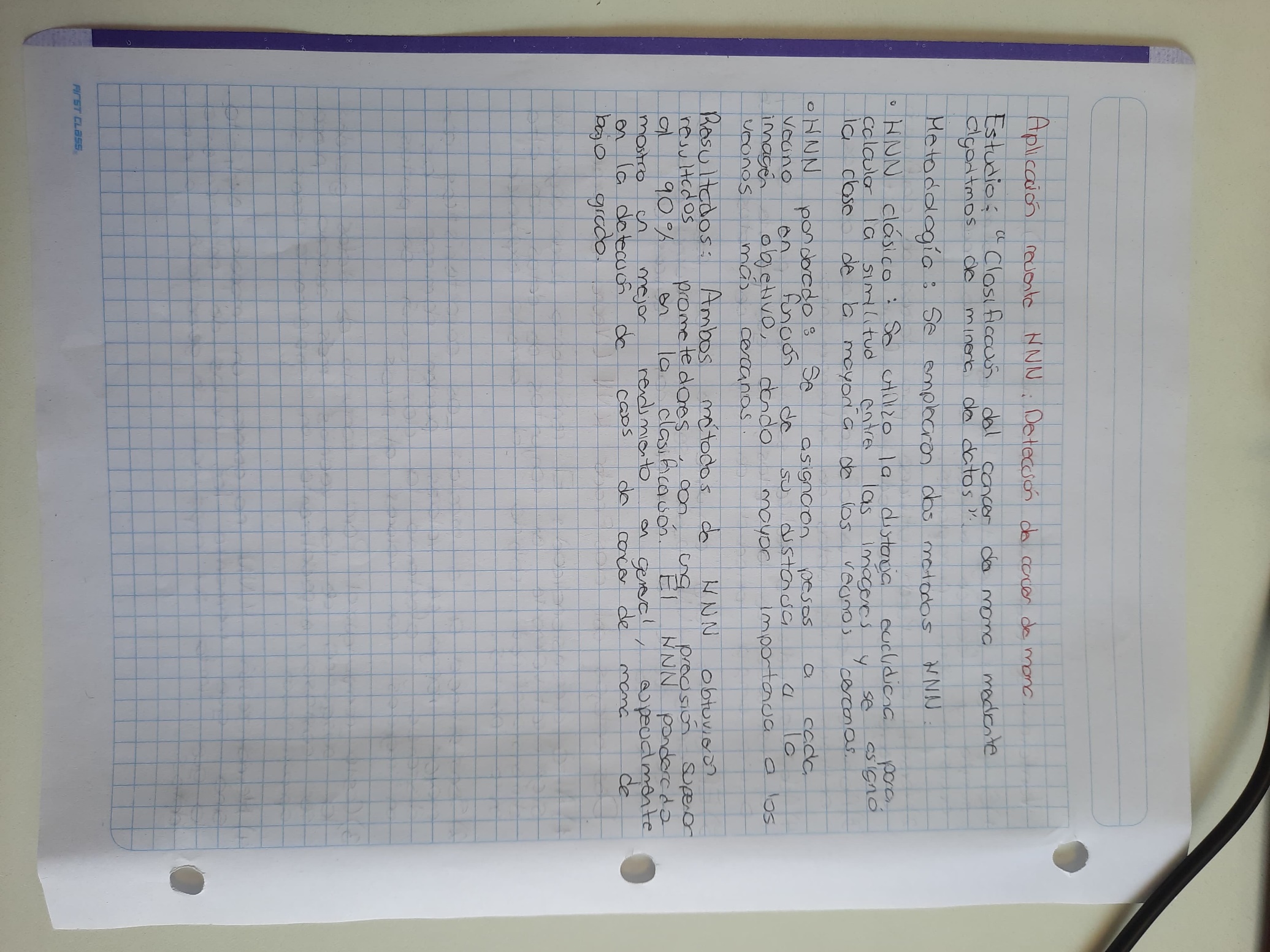
**Practica 3**

Clasificador por el método del vecino más cercano KNN

Marco teórico.







Procedimiento

El conjunto de datos de entrada que se creó está relacionado con la música y consta de cuatro características: 'Duración de la Canción', 'Género Musical', 'Número de Artistas' y 'Número de Reproducciones'. Estos datos se generaron de manera sintética y se guardaron en un archivo CSV llamado "datos\_musica.csv". El 'Género Musical' se representa como un número entero de 0 a 3, cada uno representando un género musical diferente. La 'Duración de la Canción' se mide en segundos, el 'Número de Artistas' es un conteo de cuántos artistas contribuyeron a la canción y el 'Número de Reproducciones' es un conteo de cuántas veces se ha reproducido la canción.

El clasificador K-Nearest Neighbors (KNN) que se utilizará tiene la tarea de clasificar una canción en uno de los cuatro géneros musicales basándose en sus características. El clasificador considerará la 'Duración de la Canción' y el 'Número de Reproducciones' para hacer esta clasificación. Se entrenará al clasificador con el conjunto de datos y luego se utilizará la validación cruzada para evaluar su rendimiento. Se calcularán varias métricas, incluyendo la precisión, el recall y el F-score, para tener una comprensión completa de cómo el clasificador está funcionando.

Interfaz de usuario gráfica

Descripción generada automáticamente con confianza media

Estos son los datos que se generan:

Pantalla de computadora

Descripción generada automáticamente

Después se generó una gráfica de dispersión de 2 dimensiones, también contamos la cantidad de objetos en cada clase:

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Un reloj digital en la pantalla

Descripción generada automáticamente con confianza baja

Texto

Descripción generada automáticamente

Posteriormente realizamos el escalado de datos, los datos escalados:  
Texto

Descripción generada automáticamente

Posteriormente realizamos la construcción de un clasificador por el método del vecino más cercano K-NN con valor K=3 introducimos los datos escalados, y lo evaluamos utilizando la técnica de validación cruzada técnica de validación cruzada técnica de validación cruzada técnica de validación cruzada de 10 iteraciones y las medidas accuracy, recall y F1/F-score.

Código:

Texto

Descripción generada automáticamente

Consola:

Texto

Descripción generada automáticamente

Posteriormente construimos otro clasificador por el método del vecino más cercano K-NN con valor K=15 (raíz cuadra de 200 objetos/instancias) introduciendo los datos escalados, y lo evaluamos utilizando la técnica de validación cruzada de 10 iteraciones y las medidas accuracy, recall y F1/F-score.

Texto

Descripción generada automáticamente

Texto

Descripción generada automáticamente

Posteriormente para cada clasificador, realizamos una predicción de un nuevo objeto escalando los datos e imprimiendo tanto la Clase asignada como las probabilidades por Clase obtenidas. También graficamos sus 200 objetos y los nuevos objetos usados para las predicciones (remarcados con color verde\*\*\*)

Para el primer clasificador:

Texto

Descripción generada automáticamente

Gráfico, Gráfico de dispersión

Descripción generada automáticamente



Para el segundo clasificador:

Texto

Descripción generada automáticamente

Gráfico, Gráfico de dispersión

Descripción generada automáticamente



Por último, se llevó a cabo, una gráfica por regiones de las clases (en total, 4 clases) entonces usamos 4 colores diferentes para iluminar cada región. Para visualizar las regiones de decisión de un clasificador en un espacio bidimensional, usamos una malla de puntos para cubrir el espacio de entrada y luego colorear cada punto de la malla según la clase que el clasificador asigna a ese punto.

Texto

Descripción generada automáticamente

Gráfico, Gráfico de dispersión

Descripción generada automáticamente