

# **Predicción de la apertura de los ojos en base a datos de un Electroencefalograma**

Juan Miguel Torrez Manchego

jtorrezm@fcpn.edu.bo

Universidad Mayor de San Andres

## **Justificación del clasificador**

EL clasificador utilizado para la investigación del caso de estudio es el algoritmo de K-Nearest-Neighbor (KNN), conocido también por los k- vecinos mas cercanos.

El algoritmo KNN es un método que busca las observaciones más cercanas a que se está tratando de predecir y clasifica el punto de interés basado en la mayoría de datos que le rodean, es un algoritmo supervisado, además es basado en instancia, esto nos indica que nuestro algoritmo no aprende explícitamente un modelo (como por ejemplo Regresión Logística o arboles de Decisión). En cambio, memoriza las instancias de entrenamiento que son usadas como “base de conocimiento” para la fase de predicción.

Se denomina clasificación basada en la memoria como los ejemplos de entrenamiento deben estar en la memoria en tiempo de ejecución. Cuando se trata de atributos continuos, la diferencia entre los atributos se calcula utilizando la distancia euclidiana. Un gran problema cuando se trata de la fórmula de distancia euclidiana es que la frecuencia de los valores grandes inunda a la más pequeña. Por ejemplo, en pacientes que buscan registros de enfermedades cardíacas, la medida de colesterol oscila entre 100 y 190, mientras que la edad medida oscila entre 40 y 80. Así que la influencia de la medida de colesterol será mayor que la edad. Para superar este problema, los atributos continuos son normalizados para que tengan la misma influencia en la distancia medir entre instancias” Diagnosis of Diabetes Mellitus using K Nearest Neighbor Algorithm ISSN: 2347-8578”.

kNN generalmente trata con atributos continuos, sin embargo, puede también tratan con atributos discretos. Cuando se trata de atributos discretos, si los valores de atributo para las dos instancias  $a_2$ ,  $b_2$  son diferentes por lo que la diferencia entre ellos es igual a uno, en caso contrario es igual a cero. Un estudio muestra la sensibilidad, especificidad y precisión de los resultados de KNN en el diagnóstico de pacientes con enfermedades del corazón. El valor de K osciló entre uno y trece. La precisión alcanzada osciló entre el 94 % y el 97,4 % con diferentes valores de K. El valor de K igual a 7 logró la mayor precisión y especificidad (97,4% y 99% respectivamente). El estudio también indica que kNN es uno de los métodos de minería de datos más utilizados Técnicas en problemas de clasificación. Su sencillez y la velocidad de convergencia relativamente alta lo convierte en una opción popular. Sin embargo, una desventaja principal de los clasificadores KNN es el gran requisito de memoria necesario para almacenar toda la muestra. Cuando la muestra es grande, el tiempo de respuesta en una computadora secuencial es también grande.” Application of k- Nearest Neighbour Classification in Medical Data Mining ISSN: 2223-4985”.