TITULO: “MODELACIÓN DE PROBLEMAS REALES”

Autor: Limas Pérez Luis

## RESUMEN

El propósito de este proyecto de investigación fue la creación de una página web, fue así como se implementaron las matemáticas y la computación para poder hacerle frente a un problema del área de la salud. Primeramente se tuvo que tomar una base de datos acerca de alguna enfermedad , donde los datos fueron tomados de una plataforma llamada UCI la cual está enfocada a la parte de computación como es el aprendizaje automático y la inteligencia artificial, siendo esta una fuente confiable y fidedigna; los datos proporcionados por dicha plataforma pueden ser utilizados para diversas finalidades, en nuestro caso se utilizaron como base para poder realizar un análisis detallado de la enfermedad del Parkinson y poder determinar si un individuo tiene o no dicha enfermedad, tomando en cuenta un gran número de variables como son: el nerviosismo de las personas, grabaciones de audio como vocales y frases pequeñas, por mencionar algunas variables; siendo 26 en total.

A partir de la base de datos especificada se crearon 2 nuevas bases donde se dividieron a las personas que padecían la enfermedad y otra base donde estaban las personas que no tenían la enfermedad, fue así como se obtuvieron 2 conjuntos de entrenamiento. Después de contar con los datos de entrenamiento se crearon datos de prueba que también fueron obtenidos de la plataforma UCI con la intención de poder comparar ambos datos ya que nuestra finalidad fue crear un algoritmo de clasificación dentro de nuestra página web que nos permitiera predecir si el paciente de prueba que se ingresara en la página web era una persona que padecía de Parkinson o no.

El primer algoritmo que se implementó en nuestra página fue K-NN (o también llamado el vecino más cercano), este algoritmo predice a cuál categoría ya establecida pertenece un nuevo dato de prueba.

## INTRODUCCIÓN.

“Ninguna investigación humana puede ser llamada ciencia real si no puede demostrarse matemáticamente”. Leonardo da Vinci

El proyecto fue el objetivo que se tuvo siempre en mente para aplicar todo el conocimiento aprendido durante el tiempo de investigación, este consistió en la elaboración de una página web que fue desarrollada en PHP el cual es un lenguaje de programación de uso general diseñado para el desarrollo web de contenido dinámico. Se tomó una base de datos de la enfermedad del Parkinson que contaba con un conjunto de entrenamiento el cual contenía 1040 resultados obtenidos de las mediciones realizadas, de donde 520 eran datos positivos con individuos que si tienen Parkinson y 520 negativos con individuos que no tienen Parkinson.

Se separaron los datos en esas dos categorías ya que fueron de mayor utilidad para la comparación de los resultados cuando se ingresó el conjunto de prueba; al tener las dos categorías (personas que padecen Parkinson y personas que no padecen la enfermedad) fue más fácil clasificar el conjunto de prueba ya que se obtenía como resultado la distancia que había entre el conjunto de prueba y los de entrenamiento. De igual manera se utilizó el algoritmo K-NN como primer clasificador, su funcionamiento se basa en la obtención de las distancias que existen entre el dato de prueba con los de entrenamiento como se mencionó anteriormente; entre las cuales se puede elegir la distancia euclidiana, la Manhattan y Chebyshev permitiendo así observar si los resultados se veían afectados al cambiar de distancia o se mantenían igual.

El usuario que utilice la página web puede decidir cuántos vecinos quiere considerar para llevar a cabo este algoritmo, esto ayuda a ver el comportamiento del algoritmo K-NN cuando se le asignan diferentes cantidades de vecinos. Cuando se obtienen los resultados el algoritmo K-NN cuenta con una función que ordena de forma ascendente las distancia, después cuenta cuantos vecinos de cada categoría hay y es así cuando decide a que categoría pertenece el nuevo dato ingresado. Finalmente se muestra un mensaje donde en la página web nos dice si el dato de prueba era una persona que tiene o no Parkinson.

Este algoritmo fue la herramienta que se tomó en cuenta para la clasificación en la página web, permitiendo así clasificar cada nuevo dato que se ingresó a la página y para una mejor comprensión se mostraban los resultados de forma gráfica.

## MATERIALES Y MÉTODOS (METODOLOGÍA)

Desde el principio se trabajó con una base de datos la cual contenía diversas mediciones, estas estaban dadas de forma cuantitativa, es decir, la información era puramente numérica.

Las formas en las cuales se tomaron las mediciones fueron muy diferentes ya que existen diversas variables que se tomaron en cuenta y no todas podían ser medidas de la misma manera, es por ello que se utilizaron diferentes maneras para la obtención de resultados (cifras numéricas).

Al trabajar con la base de datos tanto de entrenamiento como de prueba se tuvo que aplicar fórmulas matemáticas para poder llegar a un resultado, tal fue el caso del algoritmo K-NN y de la obtención de la distancia del conjunto de prueba con los de entrenamiento. La operación consistió en observar que tan cercas se encontraban los datos de entrenamiento de los de prueba y así hacer una comparación para poder decidir a qué categoría pertenecía el nuevo dato ingresado, cabe señalar que para poder llegar a esta conclusión solo se necesitó el trabajo de herramientas de cómputo y matemáticas.

Todo lo relacionado al proyecto se resolvió de forma cuantitativa, desde la base de datos hasta la obtención de resultados pasando por la implementación de fórmulas y análisis de las situaciones que se presentaban.

RESULTADOS ESPERADOS

Se esperaba un buen resultado al crear un clasificador web cuya finalidad fuera el predecir o saber si un paciente padecía la enfermedad del Parkinson a partir de un conjunto de mediciones que se le hicieran a dicho paciente, gracias a la implementación de un algoritmo matemático llamado K-NN el cual sería la base para la obtención de los resultados esperados.

DISCUSIÓN

Cabe destacar que el proyecto fue terminado con éxito, gracias al conocimiento creado en conjunto por las áreas de estudio Matemáticas y Sistemas, las cuales en unión pudieron aplicar el potencial y obtener un resultado satisfactorio.

En general se pueden utilizar otros algoritmos de clasificación, pero el KNN en específico está dando resultados favorables ya que las pruebas que se realizaron tuvieron una efectividad en la clasificación de un 90%, es decir solo cuenta con un 10% de error, el cual en el transcurso de mediciones y pruebas no presento inconveniente.

## CONCLUSIONES

En general el proyecto concluyo satisfactoriamente, y cabe destacar que cuando se juntan dos áreas del conocimiento que tienen tanto que complementarse una de la otra se pueden realizar grandes cosas, en este claro ejemplo está el proyecto realizado, teniendo como resultado una conjunción de áreas.

## Se realizó un clasificador web donde se puede destacar que es una herramienta que no suple en ningún momento la observación y diagnóstico de algún médico, sino que todo lo contrario, puede ser vista como una herramienta más y puede ser utilizada por especialistas para realizar de una manera eficaz su trabajo, ya que podrá ser introducido el conjunto de datos y mediciones que ha realizado al paciente en dicho clasificador y así podrá tener otro diagnostico que podrá tomar en cuenta.

LITERATURA CITADA

David McG. (2004). CSE5230 Tutorial: The ID3 Decision Tree Algorithm. Retrieved from https://users.info.uvt.ro/~dzaharie/dm2016/projects/DecisionTrees/DecisionTrees\_ID3Tutorial.pdf

Larranaga, P., Inza, I., & Moujahid, A. (n.d.). Tema 8. Redes Neuronales. Retrieved from http://www.sc.ehu.es/ccwbayes/docencia/mmcc/docs/t8neuronales.pdf

The MathWorks. (2017). Features - MATLAB - MATLAB. Retrieved July 27, 2017, from https://ch.mathworks.com/products/matlab/features.html