Tecnológico de Costa Rica, Campus local Cartago

Profesor: Jose Enrique Araya Monge

Estudiante: Yordan Danilo Salazar Villalobos

Carnet: 2018102170

I Semestre 2021

**Escuela de computación   
Bachillerato de ingeniería en computación**

**Recuperación de información textual**Tarea Programada # 3

**Introducción**

|  |  |  |
| --- | --- | --- |
| **Etapa** | **% de complet.** | **Comentario o aclaración** |
| **Lectura y conversion** |  |  |
| Puede leer y convertir datos de test\_set | 100 |  |
| Puede leer y convertir datos de train\_set | 100 |  |
| **Calcular similitudes** |  |  |
| Puede calcular las similitudes entre vectores | 100 |  |
| **Escalafón** |  |  |
| Produce un escalafón para cada documento del test\_set | 100 |  |
| **Asignación de clases** |  |  |
| Saca el promedio por clase para los primeros 10 documentos del escalafón. | 100 |  |
| Asigna la clase al documento | 100 |  |
| **Evaluación** (puede ser manual) |  |  |
| Elabora la tabla de contingencia para cada clase | 100 |  |
| Calcula las métricas: acierto, error, precisión y exhaustividad | 100 |  |

Este proyecto consiste en almacenar los resultados del algoritmo de clasificación K-NN a dos vectores(***training-set.csv***, ***test-set.csv)***  el cual extrae información noticiosa, en cada documento habrán cuatro campos separados por tabulaciones después de dicho algoritmo se harán cálculos de cada acierto, error, precisión y exhaustividad.

**Concepto**  
El algoritmo clasifica cada dato nuevo en el grupo que corresponda, según tenga k vecinos más cerca de un grupo o de otro, Es decir, calcula la distancia del elemento nuevo a cada uno de los existentes, y ordena dichas distancias de menor a mayor para ir seleccionando el grupo al que pertenecer.