

Propuesta de Proyecto de Minería de Big Data

Ana Barbosa Vargas¹ and Kiara Lucas Herrera¹

Carrera de Ingeniería en Sistemas Computacionales
Universidad de Guayaquil

PROYECTO: AMAZON

1. Definición del problema de investigación

Actualmente las redes mundiales de información están transformando al mundo y acercando más a la gente a través de la innovación de las comunicaciones, lo cual posibilita cambios en todos los ámbitos de la actividad humana dando como resultado al nuevo mundo llamado Comercio Electrónico”.

El comercio electrónico, también conocido como e-commerce (electronic commerce), consiste en la compra y venta de productos o de servicios a través de medios electrónicos, tales como Internet y otras redes informáticas.

Por tal motivo como propuesta de proyecto hemos escogido el Sitio Web de **Amazon.com**, ya que ésta compañía se dedica al comercio electrónico y servicios de computación en la nube.

Este tema de proyecto es importante e interesante porque realizaremos el siguiente análisis:

- ¿Cómo encontrar artículos similares?
- ¿Cuál es el conjunto de elementos frecuentes?
- ¿Cuál es la clasificación jerárquica de productos?

2. Datos

Los datos que utilizaremos para resolver la problemática detallada en el punto anterior lo realizaremos accediendo al sitio web de dicha compañía; por medio del cual implementaremos los métodos para recolectar información **crawling/scraping**. Los mismos que nos ayudarán a obtener los datos que se encuentran dentro del código HTML.



Este paso puede tomar una cantidad considerable de tiempo si no se es cuidadoso; por tal motivo simplificaremos la tarea enfocándonos en un subconjunto de datos o usando un conjunto de datos ya existente.

3. Metodología

¿Cómo se resolverá el problema?

- Resolveremos el problema identificando el área de investigación con las herramientas de aprendizaje que han sido y seguirán siendo proporcionadas en clases.
- Buscando información en sitios de fuentes confiables como lo es el sitio web de dicha compañía en estudio <https://aws.amazon.com/es/datasets/>.
- Investigar información de artículos relacionados al comercio electrónico en scholar.google.com.

4. Evaluación de metodología

De acuerdo a la metodología establecida lograremos saber si fué exitoso nuestro experimento ya que nos concentraremos en un conjunto de datos pequeños para poder determinar la categorización de productos más vendidos; por lo cual la métrica de evaluación será "Determinar.^{el} conjunto de elementos **frecuentes** que tienen mayor acogida en el comercio de ésta compañía.

5. Resultados esperados

Los resultados esperados se encuentran detallados en todo el contenido de la presente propuesta de proyecto, adicional los entregables de todo el proceso que conlleve el experimento de la extracción de los datos del sitio web de ".AMAZON" serán expuestos al momento de la calificación del mismo.