



Reporte 02

Nombre: Martínez López Andrés

Fecha: 31/03/2021

| | |
|--|--|
| Referencia bibliográfica | APA Xindong W., Xingquan Z., Gong-Qing W., y Wei D., (2014), Data Mining with Big Data, IEEE Computer Society Vol. 26. IEEE W. Xindong, Z. Xingquan, W. Gong-Qing y D. Wei, "Data Mining with Big Data", IEEE Computer Society Vol. 26. 2014 |
| Autor (es) | Xindong Wu, Xingquan Zhu, Gong-Qing Wu, y Wei Ding |
| Título | Data Mining with Big Data |
| Año | 2014 |
| Tipo de publicación | Artículo |
| Nombre de la revista, conferencia, Editorial u otro | IEEE Transactions on Knowledge and Data Engineering |
| Número de páginas | 11 páginas |
| Problema abordado | Teorema HACE para el análisis de Big Data. |
| Objetivo | Revisar y estudiar el teorema HACE, el cual plantea un análisis de Big data, desde un enfoque de la minería de datos. Esto se realiza tomando en consideración los problemas y vertientes que pudiera tener el teorema antes mencionado, además de las implicaciones de éste con el gran flujo de datos. |
| Justificación | Una de las áreas que ha tenido mayor expansión y crecimiento en los últimos años debido a la evolución de las redes de datos, almacenamiento y procesamiento, es big data, por esto el entender todas las implicaciones que tienen los flujos de datos masivos es importante para descubrir nuevas aplicaciones. |



| | |
|---|--|
| Marco teórico | <p>Big Data: Son mecanismos que pueden procesar y administrar cantidades masivas de datos de diversas fuentes y pueden usarse para encontrar patrones, modelos información útil de entre millones de datos. El objetivo es procesar estos datos para convertirlos en información que se pueda interpretar y ayude para la toma decisiones.</p> <p>Minería de datos: es el proceso computacional para la explotación y análisis inteligente de datos como apoyo para el proceso de la toma decisiones.</p> <p>Heterogéneo: Está compuesto por partes de diferente naturaleza.</p> <p>Asociación: Es la unión entidades que comparten un fin común u objetivo.</p> <p>Fuentes de información: Se refiere a la información que esta disponible para ser recabada con y para un objetivo dado. Las cuales son una de características principales en las aplicaciones de Big Data</p> |
| Método utilizado | <p>A partir del entendimiento de toda la cantidad de información que puede llegar a ofrecer Big data, se plantea el uso del teorema HACE, lo que permite la extracción de información valiosa y útil utilizando técnicas de minería de datos.</p> <p>Análisis en el área de Big data, por el alto impacto que tiene en las sociedades modernas en los últimos años.</p> <p>Estudio de la minería de datos para extraer información útil.</p> <p>Se proponen escenarios ficticios o ideales para la mezcla de las áreas antes mencionadas se tiene como resultado la propuesta del teorema HACE.</p> <p>Con el teorema que se plantea, en función se espera el poder obtener información de gran valor a partir de un flujo masivo de datos (Big Data).</p> |
| Fuentes de investigación utilizada | <p>Estadísticas y datos referentes al consumo y movimiento de información en el internet.</p> |
| Herramientas utilizadas | <p>Análisis de algunas plataformas tales como Facebook, Twitter, Walmart, Google, y Flickr, por su enorme flujo de datos, éstos sitios son tomados como ejemplo para realizar escenarios ficticios y realizar algunas estimaciones o resultados esperados.</p> <p>Algoritmos de minería de datos para la extracción búsqueda y análisis en los datos para realizar diferentes pruebas para comprobar el funcionamiento del teorema HACE.</p> |



| | |
|------------------------------|---|
| Resultados alcanzados | <p>Con el análisis propuesto a lo largo del artículo, se concluye que a pesar de conllevar un costo alto, la explotación de Big data, ésta resulta ser una tarea muy rentable, por todo lo que ofrece.</p> <p>A partir del teorema HACE, se establece que las principales características de Big data son:</p> <ol style="list-style-type: none">1. Grandes cantidades de datos diversas y heterogéneas.2. El control de las fuentes de información es descentralizado y distribuido.3. Las asociaciones entre los datos y el conocimiento están en desarrollo y son complejas. <p>Una de las principales características en el muestreo de los datos es que estos tienen que estar libres de ruido, errores o incumplimientos de privacidad ya que, de no serlo, se producen copias de datos alteradas que afectan el análisis. Lo que da paso a protocolos de extracción de datos para garantizar una operación exitosa.</p> <p>Se requiere de algoritmos diseñados para analizar las correlaciones entre modelos y fusionar decisiones para obtener el mejor resultado.</p> <p>El objetivo principal en el teorema HACE, se debe considerar relaciones complejas entre muestras, modelos y fuentes de datos, además de sus cambios en evolución con el tiempo.</p> |
| Aspectos de interés | <p>Con la evolución que ha tenido la computación en las sociedades modernas, se estima que todos los días se crean 2.5 trillones de bytes de datos y el 90 por ciento de los datos del mundo actual se produjeron en los dos últimos años.</p> <p>En la actualidad, el procesamiento de big data se basa principalmente en modelos de programación paralela como MapReduce.</p> <p>Según el plan de big data en la administración del ex presidente de los Estados Unidos Obama, surgió la aplicación BIGDATA en 2012, lo que derivó en varios y diferentes proyectos para seguir desarrollando y fortaleciendo esta área.</p> <p>La configuración de World Wide Web (WWW), consiste en que cada servidor web proporciona una cantidad de información y cada servidor puede ejecutarse completamente independiente de otros servidores.</p> |