Abstract

Many applications deal with huge amount of data and that scattered data needs to be transformed into something relevant and meaningful. To make sense of such data is the need of many applications and areas of technology. The data that is already present is very huge, noisy and has a complex structure. We are working on the idea of integrating data mining with data deduplication. Data Dashboard is a tool which can take complex data involving various dimensions and simultaneously uses data deduplication algorithms that help in removing redundancy in the data up to 95%. Thus it provides high reliability, low disk space and high throughput. The tool then uses this data mined information and deploys it on dynamic platform such as web which provides ease to user to access huge database.

References

- Ming-Syan Chen, “Data Mining: An Overview from a Database Perspective”, IEEE Transactions on Knowledge and Data Engineering, VOL 8, NO. 6.
- Michael Steinbach, Levent Ertöz, and Vipin Kumar “The Challenges of Clustering High Dimensional Data”
- Mamatha Nadikota, Satya P Kumar Somayajula, Dr. C. P. V. N. J. Mohan Rao, CSE Department, Avanthi College of Engg & Tech, Tamaram, Visakhapatnam, A.P., India
- Ying-Hsiang Wen, Jen-Wei Huang, and Ming-Syan Chen, Hardware-Enhanced Association Rule Mining with Hashing and Pipelining; IEEE Transactions on Knowledge and Data Engineering, Vol. 20, No. 6, June 2008.
- Alissar NASSER, Denis HAMAD, Chaiban NASR, KAIST Ky Ho Park CORE Lab, KAIST K-means Clustering Algorithm in Projected Spaces;

Index Terms

Computer Science

Artificial Intelligence

Keywords

Data mining Data Deduplication Clustering Hashing API