Mohammad rahil quazi
Mtech(CSE)

Title : Diabetes Data analysis using distributed data mining

Project Description:
Recent years have shown impressive growth in the development of ubiquitous healthcare (u-Healthcare) systems which aim for the next generation in e-Health services and associated research.
The vast amount of data collected from the distributed users of u-Health services (different clinics and hospital under same banner) results in a growing need for analyzing them across geographical lines using distributed and parallel systems.Data mining is an indispensable aspect of such systems and represents the process of analyzing databases to extract hidden knowledge and relationships.
This project aims at analyzing the data produced at different geographical locations(clinics and hospitals) at the site itself and produce a global result by combining the local results of different geographical sites.
I used the 'pima-indians diabetes' dataset in order to train and test the distributed model.

The Tasks involved in this project includes:

- To study how to distribute a given data set into multiple clusters in order to create the illusion that we are getting distributed data from different nodes(participants in a distributed network).
- After creating a different data cluster for each node, we need to create the ML learning model for each node.
- Using the Vote sharing scheme combine local analytical(ML Learning) models into a global one and finally show the pattern formation in the data.

Additional tasks:
I would like to add following points in the project:

- Using different machine learning models at different geographical sites to produce local results, as right now only decision tree is being used for learning at all sites.

References:

1. Viswanathan, Murlikrishna & Whangbo, Taeg & Lee, Ki-Jung & Yang, Young. (2007). A Distributed Data Mining System for a Novel Ubiquitous Healthcare Framework. 4489. 701-708. 10.1007/978-3-540-72588-6_117.
2. Viet Tran, Ondrej Habala, Branislav Simo, and Ladislav Hluchy. 2011. Distributed data integration and mining. In Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services (iiWAS '11). Association for Computing Machinery, New York, NY, USA, 435–438. DOI:https://doi.org/10.1145/2095536.2095624
3. Park, Byung-hoon & Kargupta, Hillol. (2002). Distributed Data Mining: Algorithms, Systems, and Applications. Data Mining Handbook. 341-358.
4. Zaki, Mohammed. (1999). Parallel and Distributed Data Mining: An Introduction. 1759. 1-23. 10.1007/3-540-46502-2_1.