International Journal of Enhanced Research in Science Technology & Engineering, ISSN: 2319-7463
Vol. 3 Issue 2, February-2014, pp: (293-300), Impact Factor: 1.252, Available online at: www.erpublications.com

A survey: Data Mining System for Mobile Devices

Ashish Kumar[1], Dr. Kavita Choudhary[2]
[1]Ph.D. Student, Dept. of Computer Science and Engineering, Jagannath University, Rajasthan, India
[2]Department of Computer Science and Engineering, Jagannath University, Rajasthan, India

Abstract: Data Mining is a process of automatic searching large data set to discover patterns and trends; those are used for sophisticated analysis after applying statistical and mathematical algorithms to determine meaningful data and the prognostication of future events and trends. The existing trends of application and tools of data mining can fulfill the need of any enterprise; but with location dependency. Apart from the desktop machines knowledge based information can be fetched from hand held devices. Till now, a number of researchers have been implemented data mining tool for specific domain and purpose. This paper is a survey on the data mining tools implemented for hand held devices.

Keyword: data ming; data mining on hand held devices.

I. INTRODUCTION

The contemporary arena of internet is growing exponentially. The business and trading has become E-Trading or E-Commerce; and the trend of electronic data exchange has been accelerated. Almost in all the organization belong to any domain; a huge amount of data is being collected to scale own opportunities in market and define new policies. But by the passes of time, the significance of collected data will depend on how much right information can be fetched from it. Gigantic business organization and research laboratories has big bank of data; for the policy makers and scientists or research scholars have to track the right data to take right set of action respectively. Research, and development towards this area, referred to as knowledge discovery from gold mine of huge data; called data mining.

Data mining is used to filter knowledge based, novel information from a huge database. Data mining is the computer-assisted process of digging and sifting through data warehouse data to obtain meaning of the data. By fetching the information through any data mining tool we can prognosticate the behavior and future trends of a business. These data will helpful to define the policies; and take knowledge-driven decision. All the companies in the market belong to any domain - retail finance, health care, manufacturing, transportation, and aerospace - are already using tools and techniques of data mining to take vantage of historical data.

Fetching, filtering and mining of refined information in varies domain of business is the core issue. But instead of sitting on a desk and using desktop or laptop kind of devices is not only the way to collect the required information. Now a days we are surrounded with handheld mobile devices like smart phones, PDA and tablets; those are playing important roles in our life. No doubt, the contemporary mobile devices are bringing the era of advanced mobile application, through which you would like to get any information at any time.

Hand held devices play a significant role in ad hoc computing environment. These devices are highly computationally capable devices with a number of sensors. Mining of information at anytime and anywhere can be possible only using the hand held devices with several new smart applications. Such applications can monitor and fetch the data for finance, security surveillance, home appliances, healthcare, athlete, soldiers in battle, etc. These applications can play a vital role in area of data mining.

The paper is organized as follows. Section II discusses literature survey on related work in the area of mobile data mining, this section is divided in two sub-section, functionalities are discussed in first section ‘A’ and architecture is discussed in second section ‘B’. The classification of problem will discuss in section III. And, the paper is concluded in Section IV.

II. LITERATURE REVIEW

Although, there are several number of data mining tools are available in the market. These tools are available in form of applications and in form of data mining solutions software packages. Few of them are solutions for distributed environment and few belong to a dedicated and particular domain, e.g., health care, health club, automobile etc. But

majorly all these fall in two categories, mobile based (e.g., MobiMine[1], VEDAS[2], Genie of the Net[3], PDM Framework[4]) and desktop based(e.g., SAS Enterprise Miner[10], SPSS Clementine [11], Microsoft Analysis Services [12], Rapid Miner[13]).

A. Functionalities of Implemented Tools

Kargupta et al [5], [6], [7] have developed the first ubiquitous data stream mining system termed MobiMine[1]. It is distributed data mining application for financial data streams based on PDA-based client-server architecture. It allows "intelligent" method to monitor the stock market from mobile devices. Using this tool, a user can store her data related to stock portfolio, and monitor the specific and relevant area of stock market. It is not a stock market forecasting system but facilitate the user to identifying their behavior and casual relationship on the base of different features and characteristics of stocks. It doesn't provide any feature for stock selection and portfolio management.

In addition to different standard portfolio management operations, the MobiMine server and client apply several advanced data mining techniques of offer the user a variety of different tools to monitor the stock market at anytime from anywhere[1]. The lists of core functionalities of the MobiMine are as following:

1) Portfolio Management and Stock Tickers: Standard book-keeping operations on stock portfolios includig stock tickers to keep an eye on the performance of the portfolio[1].

2) Focus Area: It is not an easy job for the mobile user and event for the professional to trace all the developments in the market. It gives a unique way to monitor changes by selecting a subset of user interesting events. This features is designed to support the following functionalities:

a) WatchList : The highest scored and the most interesting stocks of user will goes through a personalization module and stored on client device as WatchList. Change number of columns: Select the “Columns” icon from the MS Word Standard toolbar and then select “1 Column” from the selection palette.

b) Context Module: This is the collection of two services StockConnection Module and StockNuggets Module for understanding the market's dynamics. StockConnnection Module gives the information currently active and influent stock on the user's portfolio in graphical format. StockConnection shows the effect of market acctivity on user's portfolio[1]. StockNuggets Module applies clustering algorithms on the active stocks in order to get and identify similar behaving stocks in same or specific sector.

c) Reporting Module: It's an advance reporting system based on multimedia, i.e., it can record audio clips too, and can be invoked from any interface of the system.

In summary MobiMine perform the following functionalities on server:

- Data collection from different financial web sites and storage.
- Selection of active stocks using common statistics methods.
- Applying online data mining techniques to the stock data
- Computes the most active stocks in the market
- Server performs online data mining and then transforms the results using Fourier transformation and finally sends this to the client.

The client functionalities are:

- Portfolio management about user’s preferences
- Select a subset of this list (interesting events to the user) to construct the personalized WatchList according to an optimization module.

Kargupta et al [2] have developed Vehicle Data Stream Mining System (VEDAS)[4]. It is also a ubiquitous data stream mining system, using it user can continuously monitor the pattern of data streams generated on-board a moving vehicle. The mining component is located on the PDA[4]. A commercial version of VEDAS termed as MineFleet has been successfully deployed [8], [14] [2].

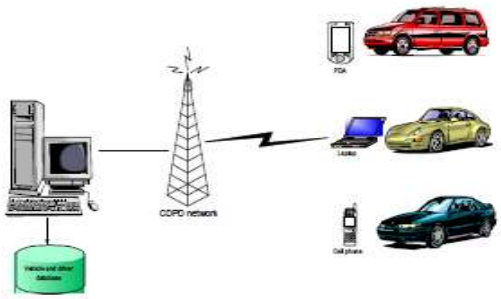

Figure 1. Conceptual overview of the VEDAS system. [2, Figure 1]

VEDAS uses to real time vehicle-health and driver characterization monitoring. The conceptual diagram is given in Figure 1. VEDAS is specially designed for vehicle monitoring through a PDA using wireless networks. It gives real-time technical solution to some sort of real-life problem, such as a drunk-driver driving a truck. No doubt, there is a need of a efficient on-board monitoring and mining system. Following are the unique characteristics of VEDAS:

- Using the PDAs or others hand-held mobile computing devices, it supports data mining and on-board data stream management.
- It supports distributed mining with less inclusion of centralization of data.
- In context of design, there are following some important constraints: (i) Reduce data communication over wireless network, (ii) minimize power consumption, (iii) reducing on-board storage; (iv) reducing time and space complexity.
- Privacy of data is also a major constraints

Pirttikangas et al [3] have implemented a mobile agent based ubiquitous data mining for a health club for cyclists. The system is called Genie of the Net[4]. The main development for the health club system is that the user has a chart for an exercise. All the required information about the health such as heart rate is recorded throughout the exercise. This information is analyzed using data mining techniques to advise the user after each exercise. It is a context-aware application.

The application Health Club implementation is on Genie architecture. This system provides the services related to exercises, planning and scheduling, instruction creation for exercising, and also instructing to cyclist during exercise. If we talk about the any health club then Genie system is helpful for the person before, during and after the exercise. Before the exercise user can use Genie for the task : exercise schedule at own terminal, summarize the instructions for every exercise, get the calendar having exercise schedule that is used to let her know the forthcoming exercise; and the planned exercises. She can monitor her heart rate, cadence and speed of bike during the exercise. And the most important, analysis after the exercise, she/he can go to own terminal, what was right or wrong during exercise, and especially advices or suggestions regarding improvement and changes.

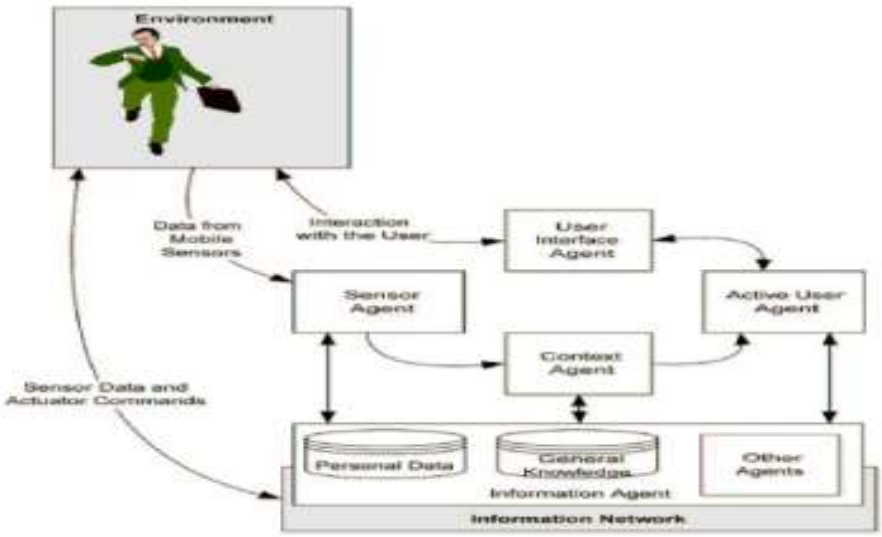

Figure 2. Genie of the Net architecture [3, Figure 1]

In case of distributed computing environments, Pocket Data Mining (PDM) is the best option for collaborative mining using hand held devices for streaming data. The new existing and advanced communication technology like Wi-Fi and Bluetooth has given opportunity to hand held devices to implement applications for data mining and analysis task in ad-hoc computing environment. Fredric et al proposed an agent based PDM framework[4]. This framework use three types of agents : AM (mobile agent miner) are spread over the whole network, MRD (mobile agent resource discovers) are used for exploring the data sources, processing techniques and computational resources, and MADM (mobile agent decision makers) roam through the network and communicate with AM to reaching the ultimate decision.

B. Architecture of Implemented Tools

The MobiMine is a PDA-based application for managing stock portfolios and monitoring the continuous stream of stock market data[1]. It is a client-server based, distributed application, designed for low-bandwidth wireless connections for financial data streams based on PDA-based client-server architecture. A client machine like mobile devices, hand-held PDAs and cell-phones, having a client application that monitor financial data stream coming through the MobiMine server. In the system prototypes single data source interact with multi mobile clients; however it is being designed to handle multiple data sources.

The complete architecture of MobiMine is divided in the two following segments:

1) The MobiMine Server Architecture: The MobiMine server regularly collects and processes the stock data from different web sources. It interacts with mobile clients to provide services supported to sift data by advanced data mining services. Figure 3 shows the server architecture. Different conceptual aspects of the server functionalities are discussed in the following[1].

a) Data Collection and Storage: This module includes the mobile database server and financial data source. The functionality of this module is to collect stock market data from different financial web sites, and it does it in iterative manner after a certain interval of time to collect a new block of data.

b) Selection of Active Stocks: This module detects the currently "active stocks" on user's portfolio. It perform some operations e.g., percentage change, volume of transactions; relevant breaking news. This list is sent to the client in order to create it personalized version – the WatchList[1].

c) Data Mining: The MobiMine makes uses of a collection of online mining techniques including several statistical algorithms, clustering, Bayesian nets, and decision trees[1]. The StockConnection module uses online techniques (e.g., Fourier spectrum-based decision trees and Bayesian learning) for detecting communication between active stock and portfolio. The StockNuggets module applies clustering algorithms on the active stocks in order to get and identify similar behaving stocks in same or specific sector.

2) The MobiMine Client Architecture : This section describes end user interface and client architecture. Client architecture is shown in Figure 4. The main components of the client system are described in the following[1].

a) Portfolio Management : This module offer a utility to the client to manage her stock portfolio and monitor the trends. Portfolio data may have such information like, stocks, quantity of purchased stocks, purchased price, buying date, review, etc. Using this interface you can edit portfolio. User can create a number of portfolios.

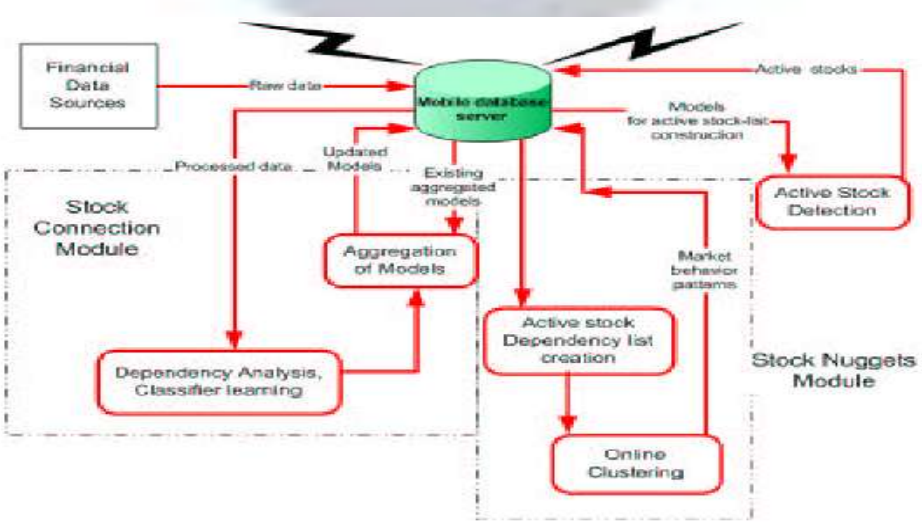

Figure 3. The architecture of the MobiMine Server [1, Figure 1]

b) Construction of the WatchList: It is constructed based on two levels of processing[1]. First level is the process to identify active stocks in the market by performing some computation using several stock features. The second level is the process of addressing the new selected active stocks into WatchList. There are two kinds of personalization, one set of stocks that user desire to watch and other user's acceptable level of risk.

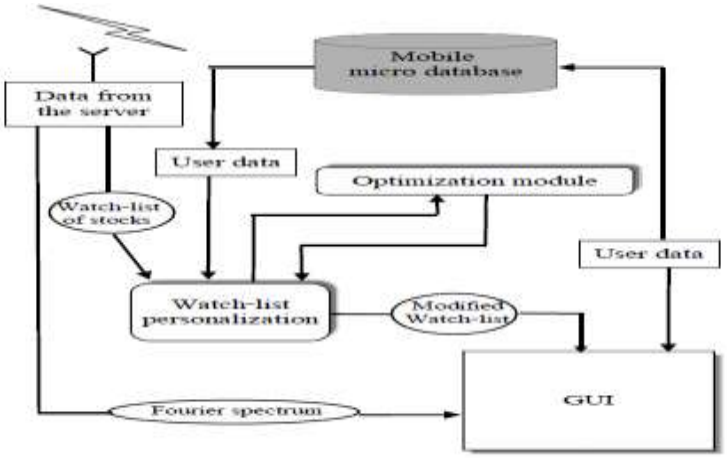

Figure 4. Data flow in the client system [1, Figure 2]

VEDAS is an experimental mobile data stream mining environment where the mobile devices perform various non-trivial data mining tasks on-board a vehicle in real-time[2]. It perform the analysis on the data that is produced by the various sensors attached with the modern vehicles. It continuously monitors, collect; and report the data-stream patterns to the remote control centre through wireless network connection. This system is basically designed and implemented for the mobile devices. Essentially, VEDAS is the collection of four components:

- A in-house developed hardware interface with a GPS module that couple with software for the on-board diagnostic (OBD-II) data bus.
- An on-board data mining and management module for data stream that connects to the a remote control station.
- A remote desktop-based control station that supports the following operations (i) interaction with on-board module, (ii) data analysis, (iii) vehicle health monitoring, (iv) driving characteristics visualization
- A module for privacy management.

The section below is given a brief about the hardware and software component of VEDAS.

Interfacing with the On-board Data Bus[2]: VEDAS is having an interfacing with OBD-II (Figure 5) for collecting the data generated by different sensors attached with a vehicle, those continuously generate large amount of data regarding different process parameters in various sub-system (e.g. engine system, fuel system, and transmission system). VEDAS access these data using the PDAs connected with OBD-II. Basically OBD-II was designed by SAE to collect and access a set of generic and manufacturer related information from the various sensors integrated with the sub-system of the car. OBD-II is GPS enabled system and collects the data with at regular intervals and can be controlled by the software on the remotely connected desktop control station or PDAs.

Figure 5. VEDAS system connected to a car. The scantool is connected to the car through the OBD-II connector. A GPS device is connected for location information and the iPAQ downloads the data from the car for analysis [2, Figure 2].

On-board Data Stream Management and Mining[2]: After collecting the data from OBD-II interface, on-board PDA analyze the data by a data mining software. Major part of the data mining will be done on-board to reduce the overheads of communication cost. On-board system is designed in such a way that it uses low bandwidth for the communication with central control site, using wireless network. DSMS (Data Stream Management System) (Figure 6) is use to buffer and systematic control the collected data through the OBD-II. The on-board module keeps track of different statistical aggregates like the mean, variance, and covariance[2].

It reduces the dimensionality and construct new representation of data. The current implementation supports the following techniques for constructing the new representation[2] : (i) incremental principal component analysis (PCA)[2]; (ii) incremental Fourier Transformation[2];(iii) online linear segmentation. All these functionalities can be turned on and off remotely.

Central Control Station Module: Central Control Station Module: It is the combination of a server with a database, both runs on a desktop computer. Figure 7 shows it's architecture, and following are the main features of the control station systems:

- It has a visualization module used to view local and global models

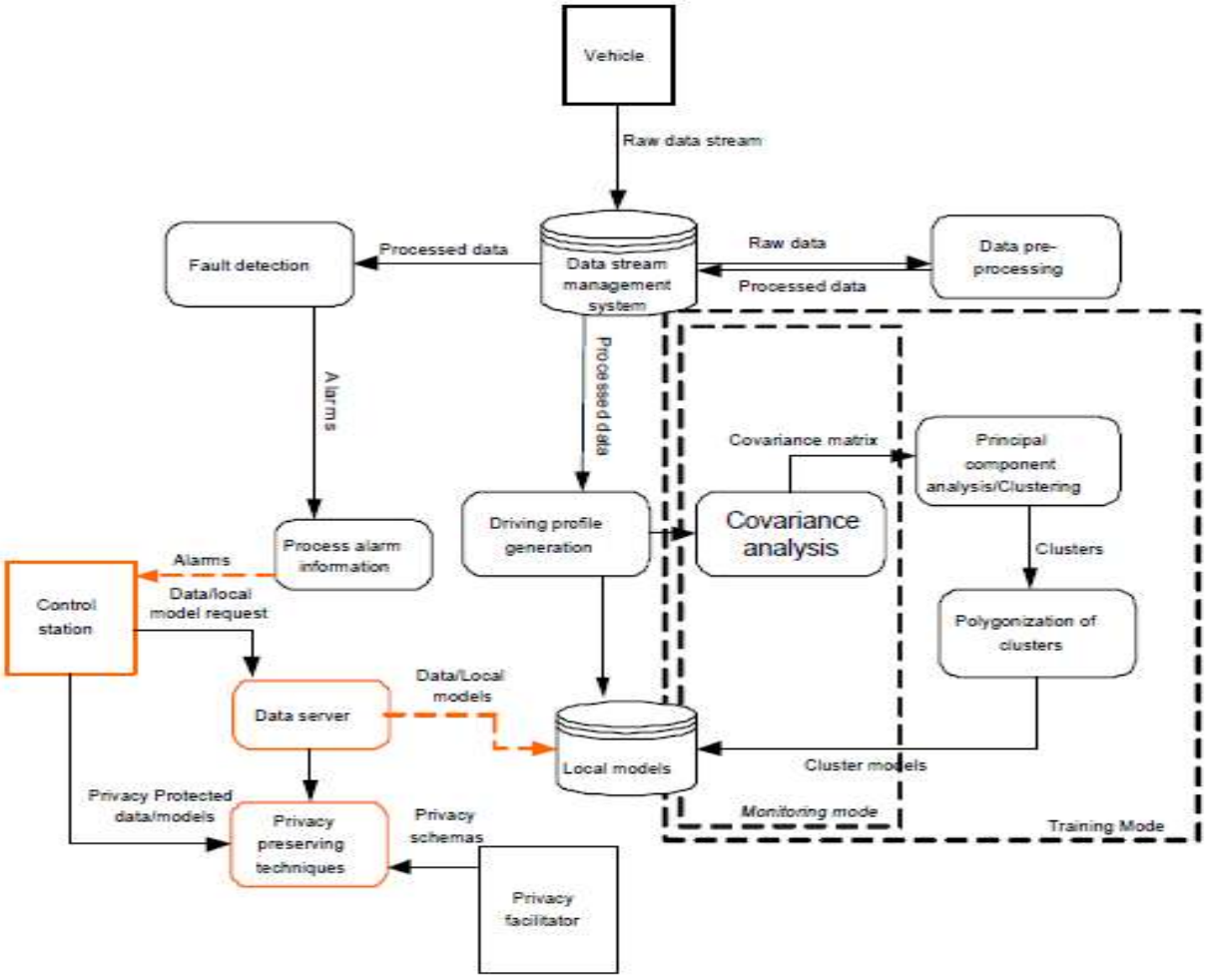

Figure 6. The system architecture for the on-board data management and mining module. It also shows part of the control station module[2, Figure 3].

- It provides an interface to control the data mining operations on-board the vehicles[2].
- It has an event management service which is used to detect unusual event, and this module runs on-board.
- It can be linked with GIS (Geographical Information System) used to provide mapping tool.

Genie collects information from sensors and databases, recognizes context based on this information, chooses relevant actions to serve the user on the basis of the recognized context, and performs the chosen actions[3]. Figure 2 shows the Genie of the Net Architecture. The application for health club started with the collection of a lot of data for about the exercise, person, scheduling, etc. The Sensor Agent perform pre-processing on the data gathered from sensors. User's context will be recognized by the Context Agent from the data collected by the sensors. Information Agent

communicates with other agents to give information regarding to store the data of sensor and recognized context. To take the appropriate course of action for serving the user, recognized context is utilized by the Active User Agent.

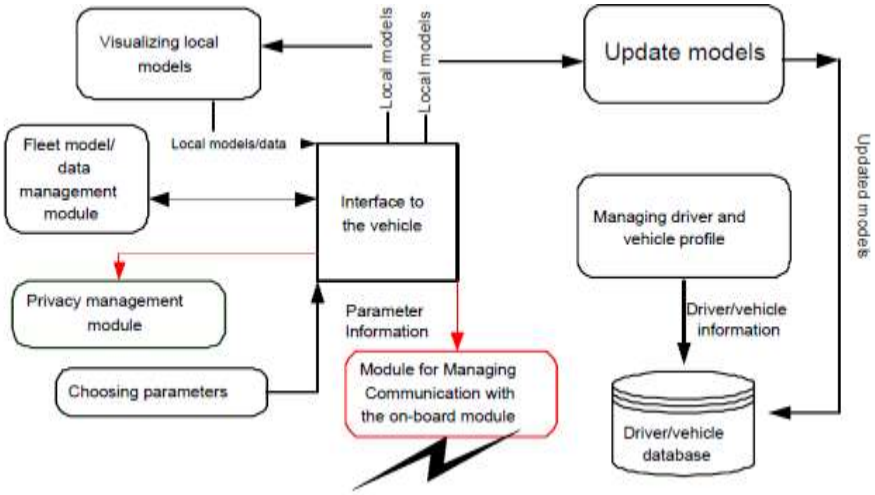

Figure 7. The functional architecture of the central control station [2, Figure 4]

The architecture of PDM Architecture is in Figure 8. As per the architecture the user's smart mobile phones contain mining module for data stream. Agent Miner is responsible for this job, and AMs spread before the initiating the mining. Stationary agents get the instruction from the task initiator for mining. Mobile Agents can travel from mobile to mobile (node to node) to perform mining related task.

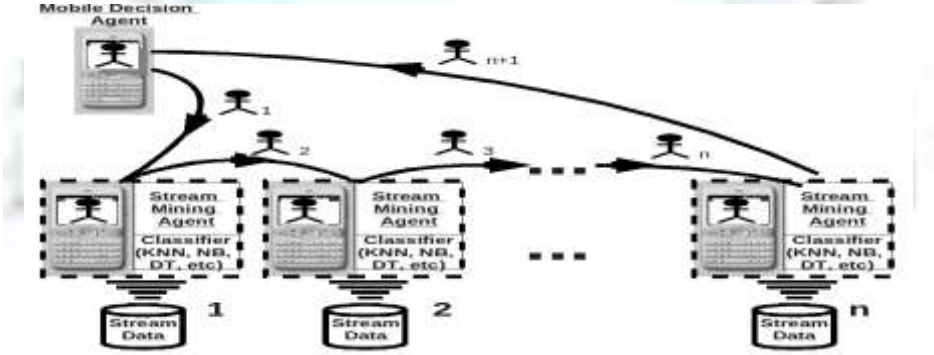

Figure. 8. PDM Architecture [4, Figure 1]

CONCLUSION

As now after the study of a number of research papers, I reached on a conclusion that all the discussed application and software were implemented for a specific domain and area without having the consideration of platform independence. Although all are useful in the proposed area of work but with the help of current technology something more can be implemented using hand held device that can be used anywhere at any time.

References

[1]. H. Kargupta, B. Park, S. Pittie, L. Liu, D. Kushraj, and K. Sarkar., MobiMine: Monitoring the Stock Market from a PDA. ACM SIGKDD Explorations Volume 3, Issue 2. January 2002, Pp. 37–46. ACM Press.

[2]. H. Kargupta, R. Bhargava, K. Liu, M. Powers, P. Blair, S. Bushra, J.Dull, K. Sarkar, M. Klein, M. Vasa, and D. Handy., VEDAS: A Mobile and Distributed Data Stream Mining System for Real-Time Vehicle Monitoring. Proceedings of the Fourth SIAM International Conference on Data Mining, Lake Buena Vista, Florida, USA, April 22-24, 2004, pp 300-311,

[3]. S. Pirttikangas, J. Riekki, J. Kaartinen, J. Miettinen, S. Nissila, and J. Roning. Genie of The Net: A New Approach For A Context-Aware Health Club. In Proceedings of Joint 12th ECML'01 and 5th European Conference on PKDD'01. September 3-7, 2001, Freiburg, Germany

[4]. Frederic Stahl, Mohamed Medhat Gaber and Max Bramer, "Pocket Data Mining: Towards Collaborative Data Mining in Mobile Computing Environments", Tools with Artificial Intelligence (ICTAI), 2010 22nd IEEE International Conference on (Volume:2), pp. 323 - 330, ISSN 1082-3409, Print ISBN: 978-1-4244-8817-9

[5]. H. Kargupta, B. Park, S. Pittie, L. Liu, D. Kushraj, and K. Sarkar (2002). MobiMine: Monitoring the Stock Market from a PDA. ACM SIGKDD Explorations. January 2002. Volume 3, Issue 2. Pp. 37–46. ACM Press.

[6]. H. Kargupta, K. Sivakumar, and S. Ghosh. (2002). Dependency Detection in MobiMine and Random Matrices. Proceedings of the 6th European Conference on Principles and Practice of Knowledge Discovery in Databases, Pp. 250–262. Helsinki, Finland
[7]. S. Pittie, H. Kargupta, and B. Park. (2003). Dependency Detection in MobiMine: A Systems Perspective. Information Sciences Journal. Volume 155, Issues 3-4, pp. 227-243, Elsevier
[8]. H. Kargupta, V. Puttagunta, M. Klein, K. Sarkar, On-board Vehicle Data Stream Monitoring using MineFleet and Fast Resource Constrained Monitoring of Correlation Matrices. Next Generation Computing. Invited submission for special issue on learning from data streams, volume 25, no. 1, pp. 5–32, 2007
[9]. J. Kaartinen, E. Mierisova, J.M.e. Oja, J-P. Usenius, R.A. Kauppinen, and Y. Hiltunen. Automated quantification of human brain metabolites by artificial neural network analysis from in vivo single-voxel 1h NMR spectra. Journal of Magnetic Resonance, 134:176-179, 1998
[10]. SAS Enterprise Miner. Available from: http://www.sas.com/technologies/analytics/datamining/miner/
[11]. SPSS Clementine. Available from: http://www.spss.com/software/modeling/modeler/
[12]. Microsoft Analysis Service. Available from: http://www.microsoft.com/sqlserver/2008/en/us/Analysis-Services.aspx
[13]. Rapid Miner, Available from: http://rapid-i.com/content/view/181/190/
[14]. Agnik, MineFleet Description, http://www.agnik.com/minefleet.html