**CONCLUSION**

Technologies related to data mainly consist of data warehousing, business intelligence using dashboards and data mining. We have implemented all the above mentioned technologies in this project.

We have implemented the data mart in star schema which consists of four dimension tables and a single fact table. The four dimensions are Time, Region, Customer and Product. During the implementation we have first loaded the data to the different staging areas tables and then loaded the warehouse table as defined by the data load scheduling process. We have implemented this using the Informatica PowerCenter (ETL toolset).

Based upon the data available in the data mart, business intelligence is implemented using the Microstrategy 9.0.1 toolset. First, the attributes, facts and the hierarchy are defined using the Microstrategy Architect. Then, various metrics, filters and reports are designed in the Microstrategy Desktop. Using the reports dashboards are generated using Microstrategy Web.

Data-mining activities can be divided into two categories:

1) Descriptive data mining, which produces new, nontrivial information based on the available data set, or

2) Predictive data mining, this produces the model of the system described by the given data set.

In descriptive data mining we have implemented mining of frequent itemsets and in predictive data mining we have implemented classification and prediction using Bayesian classification.

**FUTURE ENHANCEMENTS**

We have implemented the data mart which constitutes four dimensions. We would try to include more dimensions to the existing data mart to make it address complex scenarios. As more dimensions are added, the Business Intelligence module can also be improved by adding on different reports and dashboards.

In data mining, we would like to implement different classification and prediction algorithms and compare the efficiencies of these algorithms analyzing the same data set. For instance we are planning implement Apriori algorithm to compare the FP-Growth algorithm efficiency with it.

**BIBILIOGRAPHY**

[1] Inmon William H., “Snapshots in data warehouse”, 2000

[2] Rakesh Agrawal and Kyuseok Shim, “Developing Tightly-Coupled Data Mining Applications on aRelational Database System”, San Jose, CA 95120.

[3] JIAWEI HAN, JIAN PEI, YIWEN YIN and RUNYING MAO, “Mining Frequent Patterns without CandidateGeneration: A Frequent-Pattern TreeApproach”, 2004.

[4] CristianBolgert, “An Implementation of the FP-growth Algorithm”, 2005

[5] S. Makki, A. Mustapha, J. M. Kassim, E. H. Gharayebeh and M. Alhazmi, “Employing Neural Network and Naive Bayesian Classifier in Mining Data for Car Evaluation”, Dept. of Computer Science, Faculty of Computer Science and Information Technology, University Putra Malaysia, Malaysia, April 2011.

[6] Erik G. and Learned-Miller, “Supervised Learning and Bayesian Classification”, Department of Computer ScienceUniversity of Massachusetts, Amherst, MA 01003, February 7, 2011

[7] S. Sumathi and S.N. Sivanandam, “Data Mining in Customer Value and Customer RelationshipManagement*”*, Springer-Verlag Berlin Heidelberg, 2006

[8] Jiawei Han and Michael Kamber, “Data Mining Concepts and Techniques-Second Edition”, Morgan Kaufmann Publishers, 2006

[9] Kimball, Ralph and Ross, Margy, “The Data Warehouse Toolkit: The Complete Guide toDimensional Modeling”, 2nd Edition, John Wiley and Sons, Inc., Chichester, 2002.

[10] Inmon William H., “Building the Data Warehouse”, 4th Edition, Wiley Publishing, Indianapolis, 2005.

[11] “An Introduction to Data Mining”, <http://www.thearling.com/text/dmwhite/dmwhite.htm>.

[12] Microstrategy Guides and information retrieved from www.microstrategy.com

[13] Informatica Guide and information retrieved from www.informatica.com