**PREDICTIVE ANALYSIS IN MongoDB/BlinkDB/(Tentative) DATABASE**

**PROBLEM STATEMENT:**

The problem with predictive analysis (in this case I am considering linear and logistic regression) on large scale data is it requires too much mathematical computations on data which is memory intensive process. Computations on such large scale data is a challenge leading to questions:

How to optimize our predictive analysis computation for big data when we are having l limited computational resource?

2. What could be done in order to process large data with limited memory?

**POSSIBLE EXISTING SOLUTIONS:**

Use of Parallel computations - Hadoop **Issues:**

Learning MapReduce adds to the learning curve

This solution does not work when we have limited resources

2. Use of statistical tools like R in conjunction with the database: **Issues:**

In order to perform computations, R loads all the data into the memory. To work with huge datasets, this is the limitation of R.

3. Computation of data in chunks **Issues:**

Gradient descent is one of the main concepts for performing linear regression. Using chunks of data for calculation may not give accurate results.Since calculation of gradient descent is an iterative process, the computation would become slow.

4. USING R and MySQL in conjunction **Issues:** R programming adds to the learning curve

**RECOMMENDATIONS OF MAKING CHANGES IN DATABASE ITSELF, SO THAT THE COMPUTATION IS FAST**

**RESEARCH to find if any solutions like this exist. The purpose is to remove the requirement of multiple softwares.**