

# A BIG DATA APPROACH FOR CLASSIFICATION AND PREDICTION OF STUDENT RESULT USING MAPREDUCE

Md Anoosh, Abin, Naweed  
B.Tech student, B.Tech student, B.Tech student  
Department of Computer Science and Engineering,  
PRIST Deemed University, Puducherry, India.

## **Abstract:**

In recent years the amount of data stored in an educational database is growing rapidly. The stored database contains hidden information which if used aids improvement of student's performance and behavior. In this paper predictive modeling approach is used for extracting this hidden information. Data is collected, a predictive model is formulated, predictions are made, and the model is validated as additional data becomes available. The predictive models will help the instructor to understand how well or how poorly the students in his/her class will perform, and hence the instructor can choose proper pedagogical and instructional interventions to enhance student learning outcomes. The implementation is done in Hadoop framework with Map Reduce and Revolutionary R Enterprise RRE.

**IndexTerms** – Data collection, prediction, Hadoop, Map Reduce.

## **1.INTRODUCTION**

The main goal of this paper is to identify academically at-risk students and to develop a predictive model to predict student academic performance in educational institutions, which helps to predict their future results. Student academic performance is affected by numerous factors. The scope of this research is limited to the investigation of learning progression on their academic performance. The proposed system consists of two functionalities: 1. Identifying academically at-risk students 2. Prediction of student result.

The data collected from different applications require a proper method of extracting knowledge from large repositories for better decision making. This makes an extreme challenge for institutions using traditional data management mechanism to store and process huge datasets. So it is required to define a new paradigm called "Big Data Analytics" to re-evaluate the current system and to manage and process huge data. We implement a component of Big Data Analytics known as "Learning Analytics". Learning analytics( LA) refers to the interpretation of a wide range of data produced by and gathered on behalf of students in order to assess academic progress, predict future performance, and spot potential issues.

- The multiple regression algorithms are used for predicting student results. The map function performs the procedure of assigning each sample to the closest center while the reduce function performs the procedure of updating the new centers.
- In order to decrease the cost of network communication, a combiner function is developed to deal with a partial combination of the intermediate values with the same key within the same map task. The multiple linear regression Map-Reduce algorithms are applied to student previous result data to build a predictive model.
- This model can then predict the student final result in the form of grade or marks. This will be helpful for teacher, student and their parents to know in advance about student final predicted result and will enable them to take preventive measure.

## 2. SYSTEM ARCHITECTURE

The system architecture is a conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system.

“Big data” is a term used to describe a collection of data sets with the following three characteristics:

- Volume- Large amounts of data generated.
- Velocity-Frequency and speed of which data are generated, captured and shared
- Variety-Diversity of data types and formats from various sources.

Big data analytical approaches can be employed to recognize inherent patterns, correlations, and anomalies which can be discovered as a result of integrating vast amounts of data from different data sets. Analytics, when applied in the context of big data, is the process of examining large amounts of data, from a variety of data sources and in different formats, to deliver insights that can enable decisions in real or near real time.

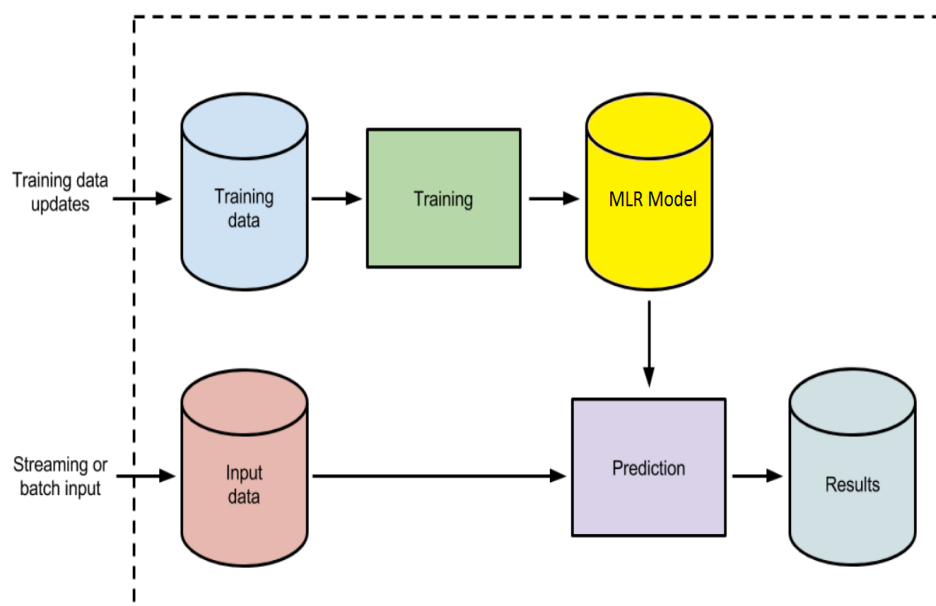


Fig. 1 : System Architecture

### 3. DATA FLOW DIAGRAM

The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system. DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output. DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.

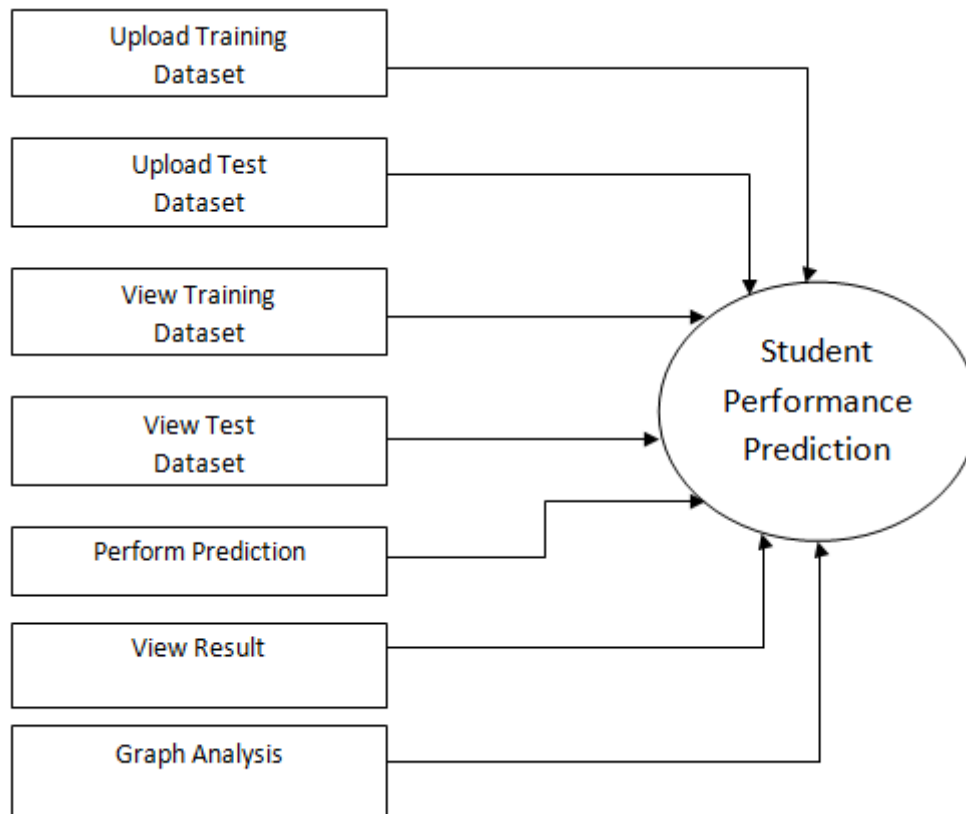


Fig. 2 : Data Flow Diagram

### 4. IMPLEMENTATION

- Data Set- Illustrate the 100 input training dataset, using which the predictive model is built. The training dataset contains information like Unit test Marks of current Year(Test 1, Test 2), Previous Result(FE, SE, TE), Attendance count in Percentage
- The clustered output of the training dataset is given. It consists of 6 clusters which contain all the dataset and partitioned them into different clusters.
- In the test dataset, a new dataset contains all the columns in training dataset except one column, i.e.; mark for one test (gp4) in training dataset. The gp4 for test dataset is predicted from the statistical relationship found in training dataset.
- Generally, the classification of data has two step processor learning and a classification step which is used to predict class labels for training data. In Training Data used Flat Files. Flat files are actually the most common data source for data mining algorithms. Flat files are simple data files in text or binary format with a structure known by the data mining algorithm to be used. In fig. represents the methodology of the research process can be learned for training data which are analyzing by a classification algorithm. In the classification step, test data are used to estimate the accuracy of classification rules.
- The final predicted result after applying test dataset in the predictive model developed using training dataset.

## 5. CONCLUSION

In this paper, we have presented a new approach called Learning Analytics and Predictive analytics to identify academically at-risk students and to predict students learning outcomes in educational institutions. The predictive models will help the instructor to understand how well or how poorly the students in his/her class will perform, and hence the instructor can choose proper pedagogical and instructional interventions to enhance student learning outcomes. It also helps instructors to predict students success and failure in the examination and also they can give proper advice to prevent failure in the examinations. Information's like Attendance, Class test, Seminar and Assignment marks were collected from the student's previous database, to predict the performance at the end of the semester. This study will help the students and the teachers to improve the division of the student. This study will also work to identify those students who needed special attention to reducing fail ration and taking appropriate action for the next semester examination.

## 6. REFERENCES

1. Alcalá, J., Sánchez, L., García, S., Del Jesus, M. ct. (2007). KEEL: A software tool to assess Evolutionary Algorithms to Data Mining problems. *Soft Comput*, 10.1007/s00500-008-0323y.
2. Clint McElroy, The Online Student Profile Learning System: a Learner-Centered Approach to Learning Analytics, April 2011.
3. Dyckhoff, A. L., Zielke, D., Bültmann, M., Chatti, M. A., and Schroeder, U. 2012 Design and Implementation of a Learning Analytics Toolkit for Teachers. *Educational Technology & Society* 15, 3, 58– 76.
4. George Lepouras, Akrivi Katifori, Costas Vassilakis, Angeliki Antoniou, Nikos Platis, Towards a Learning Analytics Platform for Supporting the Educational Process.
5. M.A. Chatti, A.L. Dyckhoff, U. Schroeder, and H. Thiis, A Reference Model for Learning Analytics, *International Journal of Technology Enhanced Learning (IJTEL) – Special Issue on “State-of-the-Art in TEL”*.