

**APPLYING NoSQL DATABASE IN DATA MINING MODELS FOR  
EDUCATIONAL BIG DATA**

**BY**

**SUMON DEB  
ID: 183-25-714**

This Report Presented in Partial Fulfillment of the Requirements for the Degree  
of Master of Science in Computer Science and Engineering

Supervised By

**Abdus Sattar**  
Assistant Professor  
Department of CSE  
Daffodil International University

Co-Supervised By

**Dr. Md. Ismail Jabiullah**  
Professor  
Department of CSE  
Daffodil International University



**DAFFODIL INTERNATIONAL UNIVERSITY  
DHAKA, BANGLADESH  
DECEMBER 2019**

## **APPROVAL**

This Thesis titled “**Applying NoSQL Database in Data Mining Models for Educational Big Data**”, submitted by Sumon Deb, ID No. 183-25-714 to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of M.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 6<sup>th</sup> December 2019.

### **BOARD OF EXAMINERS**



**Dr. Syed Akhter Hossain**  
**Professor and Head**

Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

**Chairman**



**Dr. Md. Ismail Jabiullah**  
**Professor**

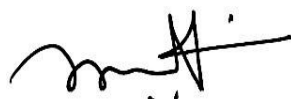
Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

**Internal Examiner**



**Dr. Sheak Rashed Haider Noori**  
**Associate Professor and Associate Head**  
Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

**Internal Examiner**



**Dr. Mohammad Shorif Uddin**  
**Professor**

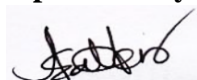
Department of Computer Science and Engineering  
Jahangirnagar University

**External Examiner**

## DECLARATION

I hereby declare that, this thesis has been done by me under the supervision of **Abdus Sattar, Assistant Professor, Department of CSE** Daffodil International University. I also declare that neither this thesis nor any part of this thesis has been submitted elsewhere for award of any degree or diploma.

**Supervised by:**



---

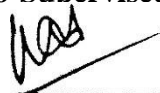
**Abdus Sattar**

Assistant Professor

Department of Computer Science and Engineering

Daffodil International University

**Co-Supervised by:**



---

**Dr. Md. Ismail Jabiullah**

Professor

Department of Computer Science and Engineering

Daffodil International University

**Submitted by:**



---

**Sumon Deb**

ID: 183-25-714

Department of Computer Science and Engineering

Daffodil International University

## ACKNOWLEDGEMENT

First, I express my heartiest thanks and gratefulness to almighty God for His divine blessing makes me possible to complete the final thesis successfully.

I really grateful and wish my profound my indebtedness to **Abdus Sattar, Assistant Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of my supervisor in the field of “*Human Computer Interaction in Education*” to carry out this thesis. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this thesis.

I would like to express my heartiest gratitude to **Dr. Syed Akhter Hossain, Professor and Head**, Department of CSE, for his kind help to finish my thesis and also to other faculty member and the staff of CSE department of Daffodil International University.

I would like to thank my entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, I must acknowledge with due respect the constant support and patients of my parents.

## **ABSTRACT**

Data mining models of education system is limited to relatively small scope of researches and analysis. In the information age, it is the vital issue to retrieve and processing available data in real time. The education system based on relational database has lack capabilities of processing real time data for its large volume and unstructured format. On the other side, NoSQL repositories have strong architecture, it can process big data efficiently in real time by using appropriate model. In this thesis, we proposed an improved model for data mining by NoSQL repositories and transformation of data from unstructured form to structured form using NoSQL technologies. Double view of big data collected in the database use data mining technology to retrieve required data. Appropriate process of query a database allows to effective retrieval of imported data for analytical systems. Data is available for analysis instantly after submitted by users. This model is suitable for any institutes or organizations, which have a large amount of data.

# TABLE OF CONTENTS

<b>CONTENTS</b>	<b>PAGE</b>
Board of examiners	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
<b>CHAPTER</b>	
<b>CHAPTER 1: INTRODUCTION</b>	<b>1-3</b>
1.1 Introduction	1
1.2 Motivation	2
1.3 Rationale of the Study	2
1.4 Research Questions	3
1.5 Expected Output	3
1.6 Report Layout	3
<b>CHAPTER 2: BACKGROUND</b>	<b>4-6</b>
2.1 Introduction	4
2.2 Related Works	4
2.3 Research Summary	4
2.4 Scope of the Problem	5
2.5 Challenges	6
<b>CHAPTER 3: RESEARCH METHODOLOGY</b>	<b>7-13</b>
3.1 Introduction	7
3.2 Research Subject and Instrumentation	7

3.3 Data Collection Procedure	7
3.4 Usage Scenario and Design Principle	9
3.5 Description of System Architecture	10
3.6 Proposed Model Algorithm	12
3.7 Implementation Requirements	13
<b>CHAPTER 4: PERFORMANCE OF THE PROPOSED MODEL</b>	<b>14-24</b>
4.1 Introduction	14
4.2 Experimental Results	14
4.3 Descriptive Analysis	23
4.4 Summary	24
<b>CHAPTER 5: SUMMARY AND CONCLUSION</b>	<b>25-26</b>
5.1 Summary of the Study	25
5.2 Conclusions	25
5.3 Recommendations	26
5.4 Implication for Further Study	26
<b>REFERENCES</b>	<b>27</b>

## **LIST OF FIGURES**

<b>FIGURES</b>	<b>PAGE NO</b>
Figure 3.3.1: Big Data Collect from various Sources	8
Figure 3.5.1: Architecture of Proposed System	10
Figure 3.6.1: Flow Chart of Proposed System	12
Figure 4.2.1: Collection of Data in JSON Format	14
Figure 4.2.2: Graphical User Interface of Our Developed Model (GUI)	15
Figure 4.2.3: Data Import Process	16
Figure 4.2.4: Displaying NoSQL Data	17
Figure 4.2.5: Displaying NoSQL Data in SQL Format	18
Figure 4.2.6: Box Chart of Analyzed Data	19
Figure 4.2.7: Line Chart of Analyzed Data	20
Figure 4.2.8: Bar Chart of Analyzed Data	21
Figure 4.2.9: Area Chart of Analyzed Data	22
Figure 4.2.10: Pie Chart of Analyzed Data	23



# CHAPTER 1

## INTRODUCTION

### 1.1 Introduction

Big Data is now very common word and it has become popular in scientific community for its grate impact. There is no universal accepted definition for big data. It can be describe by 5V model, which are Velocity, Varity, Volume, Value and Veracity [1]. Velocity is the speed of shifting digital content from data set to data stream. Varity is different kind of natures of data such as texture data, multimedia data etc. Volume describes the size of data, which can be terabyte to petabyte. Value determines by the cost of data and noise renovation of data stream is called veracity.

Generally, Relational Database Management System is very popular and common in this time, but it has become unfit for work day by day. The main drawback of relational warehouse is limitation of its storage and data mining capability. On the other side, there is no particular and significant structure of big data. For this, it is a big trouble to handle big data using RDBMS. In this case, NoSQL has various databases such as graph, object, multimodal etc. Moreover, the main advantage of NoSQL is various type of data storage capability such structured, semi structured and unstructured data. The repository can store real data in its storage like as file, image, attachment etc.

In educational world, institutions gather a large amount of structured and non-structured data about students, teachers and other employees as well as course materials. Which can be than used to improve educational systems, as related personnel such as teachers can depends not only their own knowledge but also acquire knowledge from others experiences. Nowadays, it is the big problem to perform data miming in such large storage. There are various kinds of data and the major problem is summarization of these data.

In this thesis, we proposed a data-mining model using NoSQL, which can perform to generate analytical reports efficiently. In practical world, we need structured data in everywhere. The model has given an accumulated structured data from NoSQL

repositories. The model is designed using MongoDB (NoSQL Database) for its powerful data engine.

Thus, the proposed data-mining model can provide better output in short time. It can be fruitful for educational personnel such students, teachers and so on.

## **1.2 Motivation**

The world requirement is a superior educational environment. An organized educational system is the basement of a nation. The volume of data in educational world is increasing day by day. Globally those data become in large volume and these varieties are increasing continuously. Therefore, a problem has arisen to retrieve and mining of these data using current system. For the necessity of superior education system, we approached to thesis in this area and work with educational institutions. The data mining model could be very useful for education system to analysis their data. It can provide a big support to educational personnel such as teachers, students, stuffs. They can get analytical reports by the model and can improve education system by analysis those.

## **1.3 Rationale of the Study**

This is not something new to research in educational world. Many people worked on this area. To generating the knowledge, it is the best platform. For necessity of improve education system we tried to focus on data mining system on global educational data.

The research objectives as shown below:

- To improve education system, need more analysis of existing data.
- To provide an appropriate model for education system.
- To combine all data of both structure and unstructured format.
- To get proper analytical reports of global education system.
- To find out the major issue in the system.
- To identify weak students for caring him.
- To identify weakness of students.
- To decrease the dropout rate of students.
- To evaluate performance of particular intuitions.

## 1.4 Research Questions

These are some questions focused in this thesis as shown below:

- Why cannot be improved education system?
- Why student perform bad result continuously?
- How to improve the education system?
- What is the problem in data mining of current system?
- How to accumulate structured data and unstructured data?

## 1.5 Expected Output

In this thesis, we tried to analysis educational using NoSQL database with appropriate model. The proposed model will provide a lot of support to improve the educational system. The educational personnel can analysis the output for their necessary step. The research expected output:

- Structured data of education system from unstructured data.
- Accumulated global data in one database.
- Analytical reports from NoSQL database.
- Reports in graphical format from NoSQL database.

## 1.6 Report Layout

The report is designed in one way that all of the issue can determined.

**Chapter 1:** The chapter contains introduction, motivation, rational of study, research questions, expected output and report layout.

**Chapter 2:** It describes background of this thesis, related work, research summary, problem discussion.

**Chapter 3:** It describes system architecture of the proposed model.

**Chapter 4:** All over performance is described in this chapter.

**Chapter 5:** It describes result of the proposed system.

**Chapter 6:** It describes the conclusion of thesis.

## **CHAPTER 2**

### **BACKGROUND**

#### **2.1 Introduction**

In information age, educational data field, which is essential to work on it. Data mining process that produce knowledge from dataset, as system needs and disclose hidden patterns. To improve education field, many researcher tried to establish better system by the virtue of data mining technology.

#### **2.2 Related Works**

There are huge research work on data mining technique with structured and unstructured data. Many researchers work with data mining for relational database. Till now, it is not popular to practice with NoSQL.

Marcin Mazurek proposed a model titled to combine both RDBMS and NoSQL technologies in an analytical system [2].

A B M Moniruzzaman and Dr. Syed Akhter Hossain has given a discussion to provide classification, evaluation and characteristics of NoSQL databases for Big Data Analysis [3].

Aryan Bansel, Horacio Gonz´alez-V´elez, Adriana E. Chis tried to provide data standardization and classification stages with efficient mapping and they discuss different between cloud-based and NoSQL data stores [4].

Benymol Jose and Sajimon Abraham tried to solve the problem of insrtng unstructured data. They discussed flexibility of NoSQL to various sorts of data, structured data, semi structured data or unstructured data [5].

F.N. Chernilin, M.M. Rovnyagin, A.V. Guminskaia, O.V. Myltsyn, V.M. Kinash, A.V. Kuzmin and A.P. Orlov have given disscussion to accelerate NoSQL-system and amplify their functionality [5].

The mentioned papers are helpful for my research work. In these papers, they worked with Relational Database Management System (RDBMS) and NoSQL. Moreover,

they tried to provide acknowledgement and depth information of structured and unstructured database to analysis Big Data. In this thesis, we proposed a model for data mining with NoSQL repositories efficiently. Applying the model, a better output could be achieved for proper analysis of educational big data. It will be more useful for educational data analysis.

## **2.3 Research Summary**

We are creating a large amount of data in educational repositories that is not being utilized perfectly. We can get advantage from the large educational dataset, by using data mining processes, which serve as stronger tools for characterization, investigation, analysis and prediction. Nowadays, many people are using data mining techniques in various fields but it is limited used in educational perspectives. Moreover, it is more limited for NoSQL repositories. In our proposed model, we used NoSQL database, which provide an efficient way to retrieve data from database that contains a large volume of data. Who need to analysis data from such database can be able to retrieve data in a short time to use that information for analysis, characterize, investigation and preparing valuable reports. This model can be helpful for any organizations especially for educational institutions, which play with large volume of data. This model will provide advantages to the both students and teachers also and they can support their institutions by quality assurance of education system. It is more effective to improve our education system.

## **2.4 Scope of the Problem**

Student's performance evaluation of the global educational systems is now one of the vital issue. In this time educational institutions are fail to proper guide their students and teachers as well as related stuff due to lack of information and training. To overcome these problems educational system must need a better model for assessment, analysis the information to provide proper training to improve system. Their need a combined system that play significant role in data analysis, characterized, clustering, classification, regression and visualization. Proper data mining technique need to ensure compute, analysis and prediction of available data. To compute and analysis data, there are required structured and organized data.

However, in real life, each organization or institutes store data in different structure. Therefore, it is more difficult to analysis accumulated-data of all institutions together.

## **2.5 Challenges**

The most challenges are data collection from different perspectives. There are huge data in different structures, which are accumulating and manipulating is big problem. The combined of different structured data produce an unstructured dataset. We need to accumulate all available data to analysis and classify global education system.

## **CHAPTER 3**

### **RESEARCH METHODOLOGY**

#### **3.1 Introduction**

Data mining techniques and concepts are used in various perspectives. Educational System, Customer Relationship Management, Healthcare, Banking, Fraud Detection, Market Analysis, Real Estate, Criminal Investigation, Manufacturing, Engineering, Education are the most common Data Mining applications. Educational data mining is now one of the most demanding issue. There are so many techniques are used for educational data mining such as Decision Tree, Naïve Bayse, Regression, Neural Networks, K-Nearest Neighbors and so on.

These techniques provide us various knowledge to discover such as association rule, classification, regression and clustering. These techniques can be useful us to retrieve data using NoSQL.

#### **3.2 Research Subject and Instrumentation**

In this section, we talk about the requirements tools for implementation. We collect data from various university and some other educational institutions. To develop this system, here we used Python programming language. The collected data we are stored in JSON format. Then we produce the dataset using MongoDB. Then we produce expected output by implementing the data mining technique on the generated dataset using “Matplotlib” library of Python. Here we also used “Numpy”, “Seaborn” and “Bokeh” library of python in implemented time. The models are assessed by calculating accuracy through sklearn preprocessing. We created various graphical user interfaces (GUI) for displaying the data and evaluating the models using MongoDB. All of the implementation processes and procedures are done by using Ubuntu 18.04.

#### **3.3 Data Collection Procedure**

First, we identified the Big Data and determine the origins from which come. Suppose that we dealing with Daffodil International University, which is very similar to any other universities. After collecting data of one university, we have tried to collect data

from other educational intuitions. Main differences start to emerge at the level of elementary and secondary schools. We separate this data on various groups based on its origins, as shown in Figure 3.3.1.

- Documents in offline form
- Server's Logs
- Public portal
- University information
- Data from social network

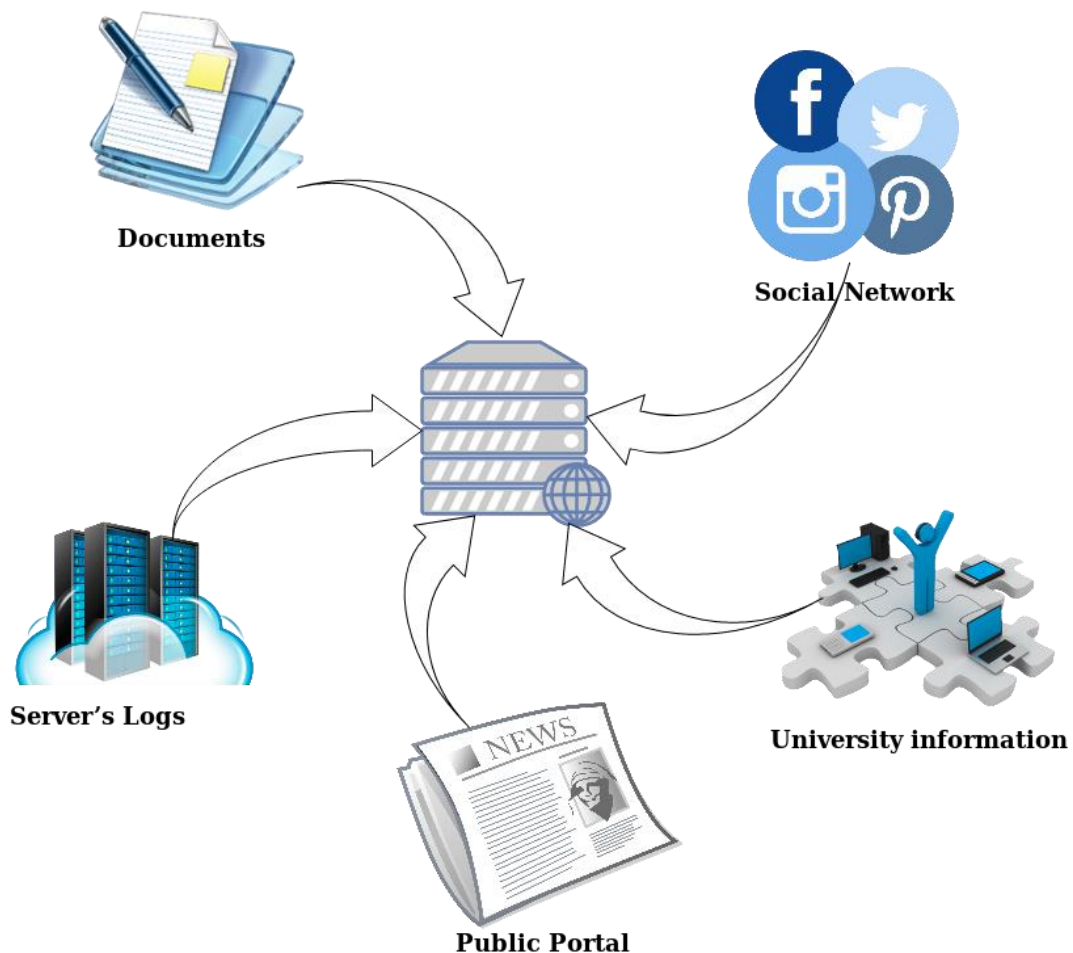


Figure 3.3.1: Big Data Collect from various Sources



### 3.4 Usage Scenario and Design Principle

The system targets various groups. This system will be useful to researchers, teachers, students and system of education. We discussed various use-case scenario for the system:

1. Education Sector
  - a. Actors: Teachers, Students, Staff.
  - b. Input: Students Information, Teacher's Information, Students attendance results, others education information.
  - c. Output: Big data analytical report, structure dataset, aggregated statistics for educational system, visualized pattern.
  - d. Tools: Desktop application.
2. Data distribution from large amount of both structured data and unstructured data.
  - a. Actors: Database administrator.
  - b. Input: Stored data from various repositories.
  - c. Output: Extracted dataset of structured and unstructured data, algorithms.
  - e. Tools: NoSQL repositories such as MongoDB, RDBMS.
3. Data Analysis.
  - a. Actors: Data scientist.
  - b. Input: Dataset of structured and unstructured data.
  - c. Output: Data analytical reports, visualization patterns, algorithms.
  - d. Tools: Statistical packages of programmatic language like Python or R.
4. Predictive Modeling.
  - a. Actors: Users, Researchers.
  - b. Input: Relational repositories, Anomaly detected data.
  - c. Output: Graphical Report, Prediction.
  - d. Tools: Data mining tools, visualization package of Python.

To achieve the goal of our proposed system, the above usage scenario has been determined.

### 3.5 Description of System Architecture

There are two types of dataset in our proposed system, which are structured and unstructured dataset. The reason is large volume and variety of dataset. There are many institutions, which have various kinds of data in their own format. These have become an unstructured data when all data are combined in a single database system.

In this thesis, our prime concern is to retrieve those data from large volume of data efficiently. To do this, data are stored in NoSQL repositories, then classify and cluster these dataset using data mining technology, as shown in Figure 3.5.1.

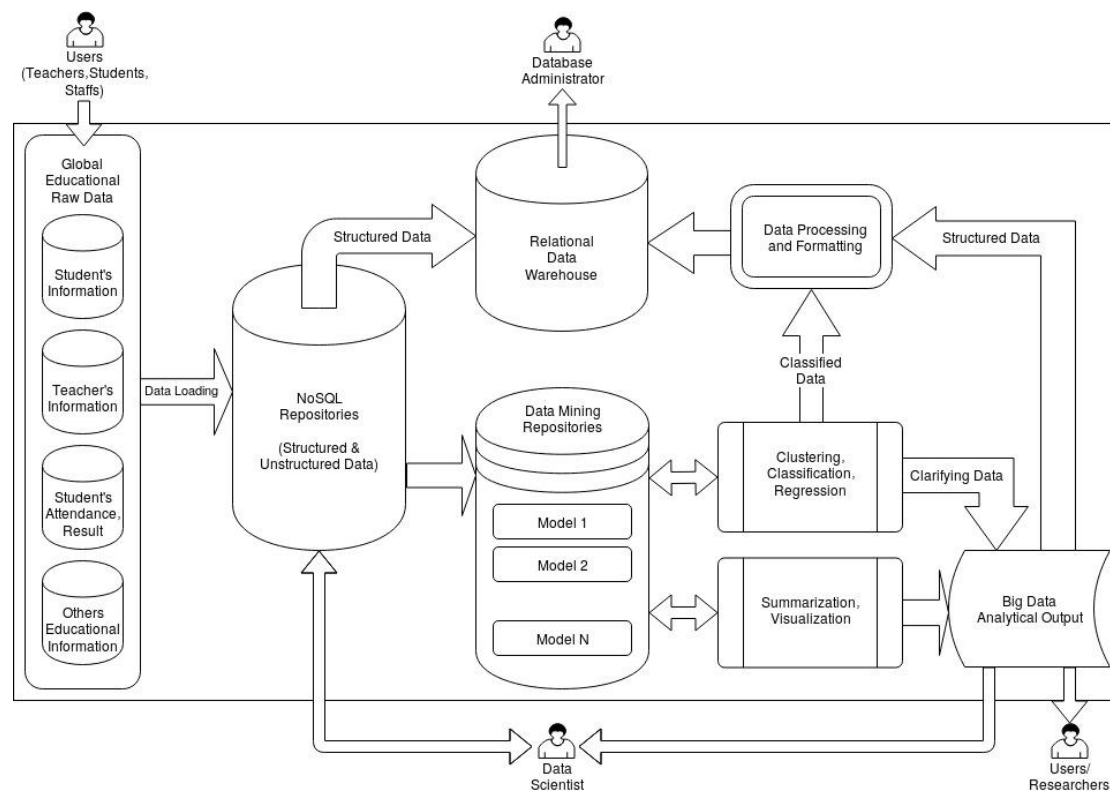


Figure 3.5.1: Architecture of Proposed System

The architecture focuses our proposed model, by which, users can retrieve and mining data efficiently and they get all analytical reports in the system. The architecture is performing with proper algorithm, which has given in next section. The working procedure has shown below:

## **Data Sources**

Data stored by different actors such as teachers, students, and stuffs. They collect data from various sources such as universities, schools; educational institutions and so on about student's information, teacher's information, student's attendance, results etc. They are mainly collected as data. Collected data can be structured or unstructured form.

## **NoSQL Repositories**

Global educational raw data is loaded in NoSQL repositories like as MongoDB. Which is able to store both kinds of structured data and unstructured data in Database. In NoSQL database, data are stored in JSON format. For this, user can able to store any kinds of data as JSON format. The NoSQL database can also provide user with relational data view of stored data.

## **Data Mining Repositories**

After retrieved data are stored in data mining repositories, which is then process by data mining model. Data are manipulated using data mining processes are include clustering, classification, regression, summarization and visualization.

## **Relational Data Warehouse**

It is only able to gather structured data from repositories. And passes unstructured data to Hadoop for transform into structured data by perfuming some operations.

## **Big Data Analytical Output**

It is the final output of our system, which provide analytical output to users in various form that depends on the user's requirements; such as graphical, tabular, SQL and so on. The generated output, which is structured form, can also be store in relational data warehouse.

### 3.6 Proposed Model Algorithm

After find and identify a data source we can begin processing data with them. The algorithm of our propose system is given bellow, as shown in Figure 3.6.1.

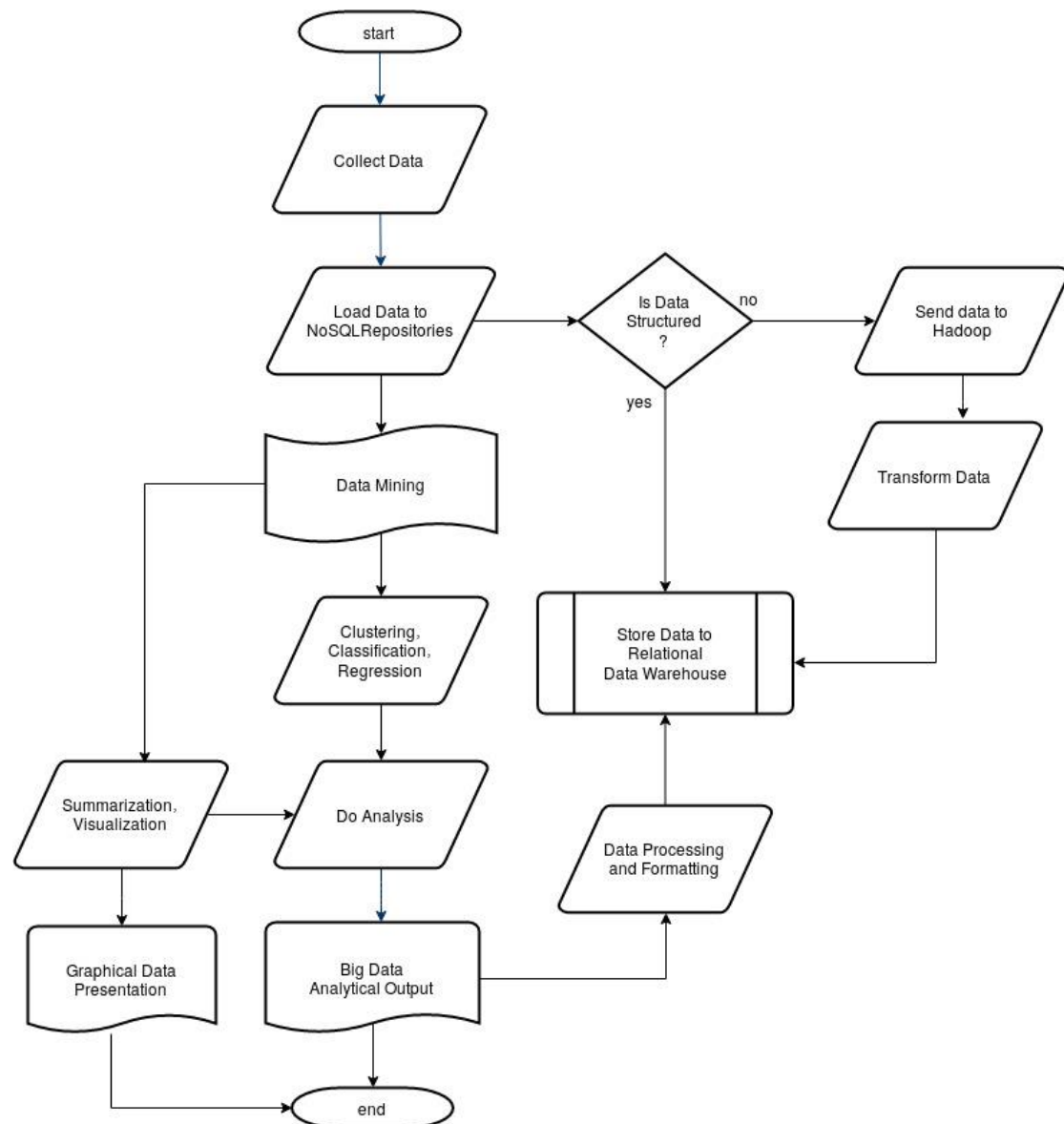


Figure 3.6.1: Flow Chart of Proposed System

This figure shows that, first collect data from various sources, then load data into NoSQL database. The collected data can be in both form, which are structured or

unstructured format. After loading data into NoSQL database, data processed by data mining techniques. There are various processing techniques such as clustering, classification, regression, Summarization, visualization etc. After processing data, processed data can be provide to Data Scientists or Data Researchers for analysis. Users can also receive processed data in graphical form. After analysis data, system provides big data analytical output. The output can be also stored in data warehouse by processing and formatting data.

Relational data warehouse can gather only structured data, so after loaded data into NoSQL repositories, data can be directly stored in relational data warehouse if data is in structured form. Otherwise, data is send to Hadoop for structuring data and then after transformed data into structured form, it can be stored in relational data warehouse.

### **3.7 Implementation Requirements**

In this section, we talk about the requirements tools for implementation. We collects data from various university and some other educational institutions. To develop this system, here we used Python programming language. The collected data we are stored in JSON format. Then we produce the dataset using MongoDB. Then we produce expected output by implementing the data mining technique on the generated dataset using “Matplotlib” library of Python. Here we also used “Numpy”, “Seabon” and “Bokeh” library of python in implemented time. The models are assesses by calculating accuracy through sklern preprocessing. We created various graphical user interfaces (GUI) for displaying the data and evaluating the models using MongoDB. All of the implementation processes and procedures are done by using Ubuntu 18.04.

## CHAPTER 4

### EXPERIMENTAL RESULTS AND DISCUSSIONS

#### 4.1 Introduction

In general, there are different kinds of data in real life and all over data produced in unstructured form. To evaluate performance, need collected data in JSON format. For performing the model, we have collected data from various universities. These data have some similarity. All those data, we have converted in JSON format and given in produced system. After that, the system has given expected outcome.

#### 4.2 Experimental Results

Experimental input and output analysis of our developed model is illustrates in follows:

For experiment the results of the proposed model, at first we collected data from various sources such as university and other educational data from some educational institutions, which is stored in NoSQL database with JSON format, as shown in Figure 4.2.1.

```
1  [
2  {
3      "Name": "Md. Rakibul Islam",
4      "Studen_ID": "101-15-945",
5      "Result": "4",
6      "Passing_Year": "2009",
7      "Varsity": "DIU",
8      "Address": "Dhaka"
9  },
10 {
11     "Name": "Md. Nazim Uddin",
12     "Studen_ID": "101-35-107",
13     "Result": "3.74",
14     "Passing_Year": "2009",
15     "Varsity": "DIU",
16     "Address": "Dhaka"
17 },
18 {
19     "Name": "Mohammed Zahid Hasan",
20     "Studen_ID": "103-15-1107",
21     "Result": "3.76",
22     "Passing_Year": "2009",
23     "Varsity": "DIU",
24     "Address": "Dhaka"
25 }
26 ]
```

Figure 4.2.1: Collection of Data in JSON Format

To import data in our model we developed a Graphical User Interface (GUI) using Python Qt5 package, which includes ten buttons to perform various operations, as shown in Figure 4.2.2.

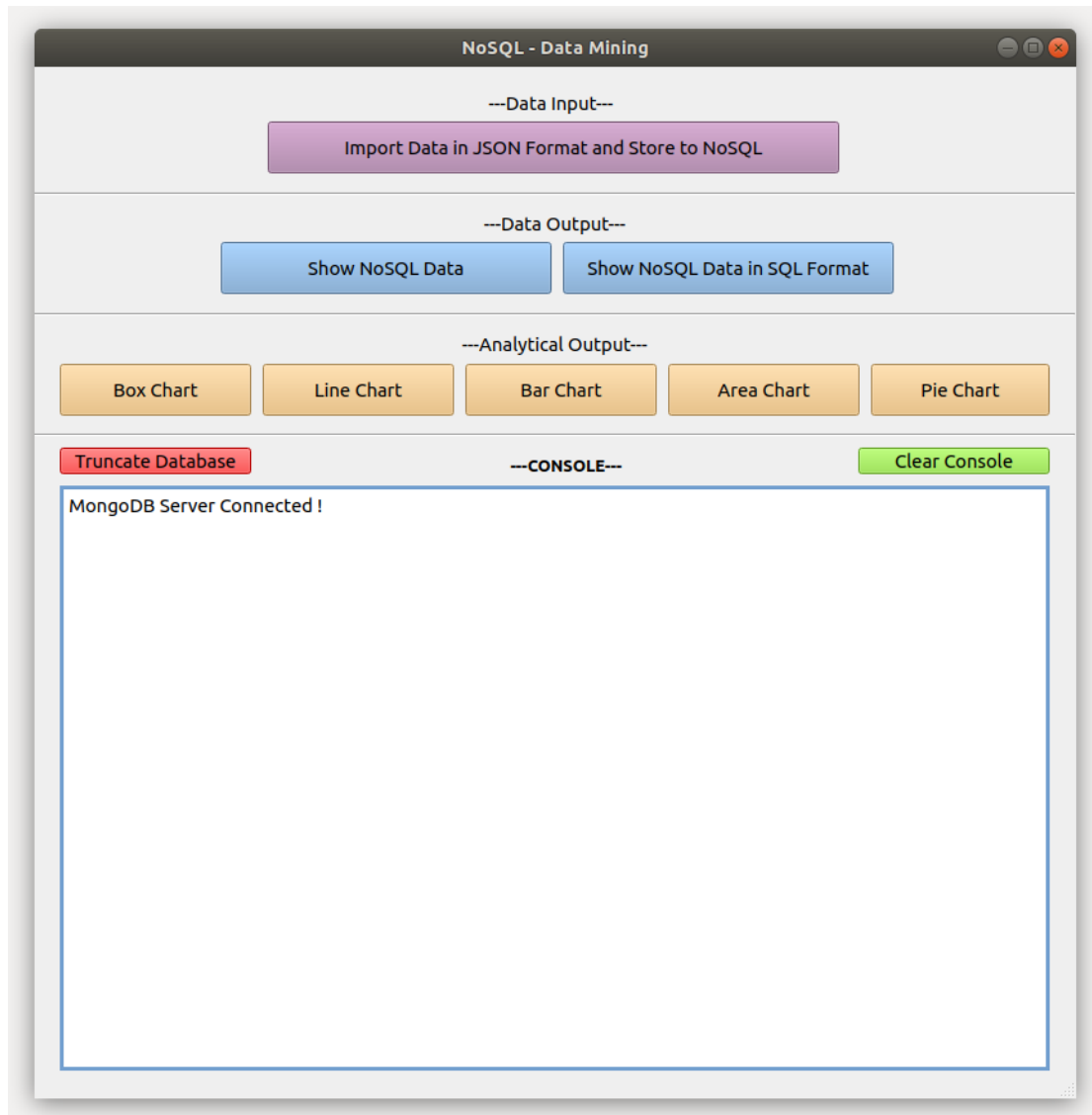


Figure 4.2.2: Graphical User Interface of Our Developed Model (GUI)

To perform the developed model, first we need collected data as input, which can be structured, semi structured or unstructured data. We need input data in JSON format. Therefore, we need to transform all data into JSON format for input of the system.

First button is use for import collected JSON data in NoSQL database and others button is used for showing output. In this GUI we used console for displaying data and message. Here “Truncle Database” button is used to erase database and “Clear Console” button is for blank the console.

To import and stored data we need to click “Import Data in JSON Format and Store to NoSQL”, as shown in Figure 4.2.3.

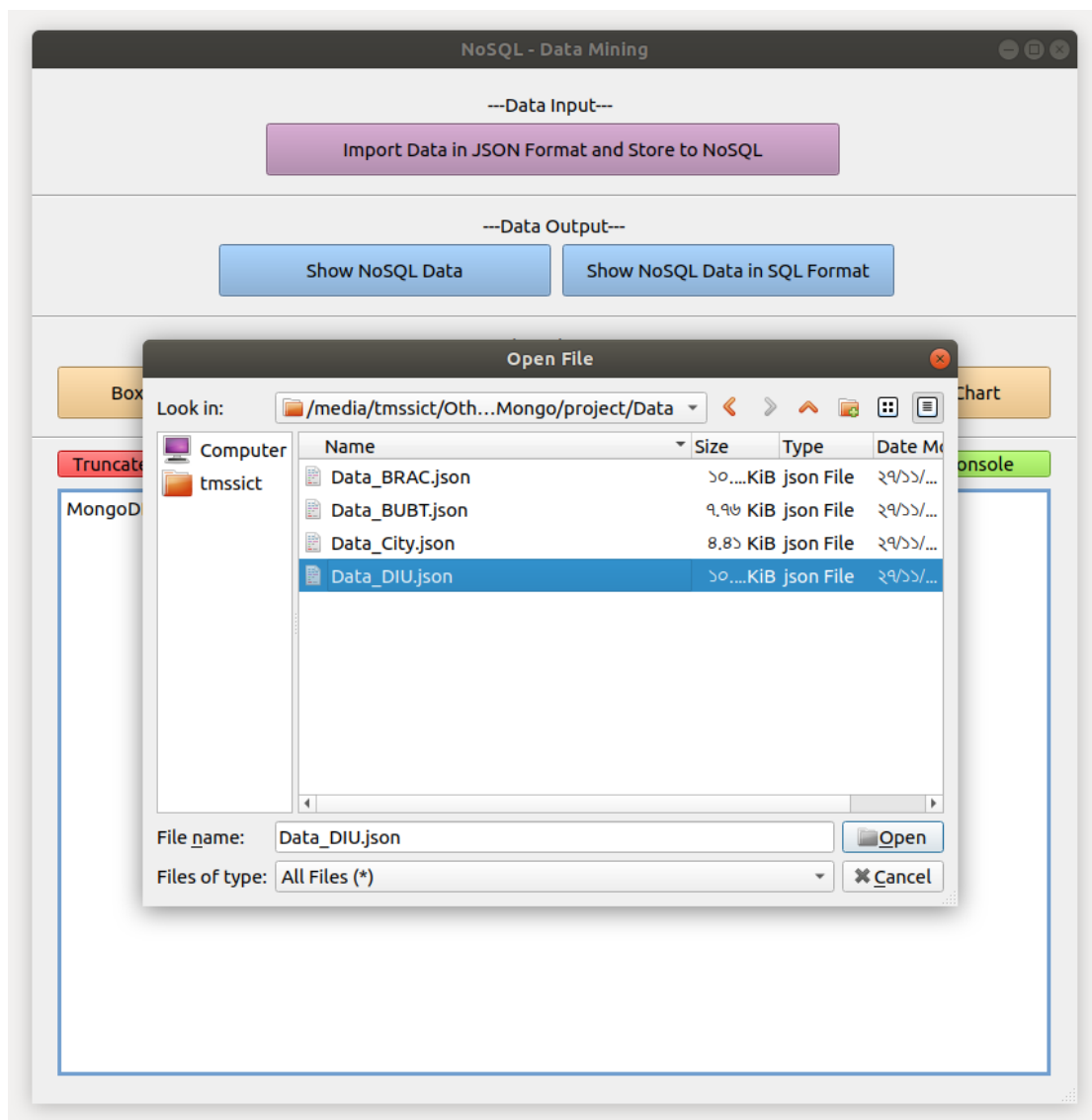


Figure 4.2.3: Data Import Process



After importing collected data, that stored in NoSQL database that can be displayed by clicking on “Show NoSQL Data”, as shown in Figure 4.2.4.

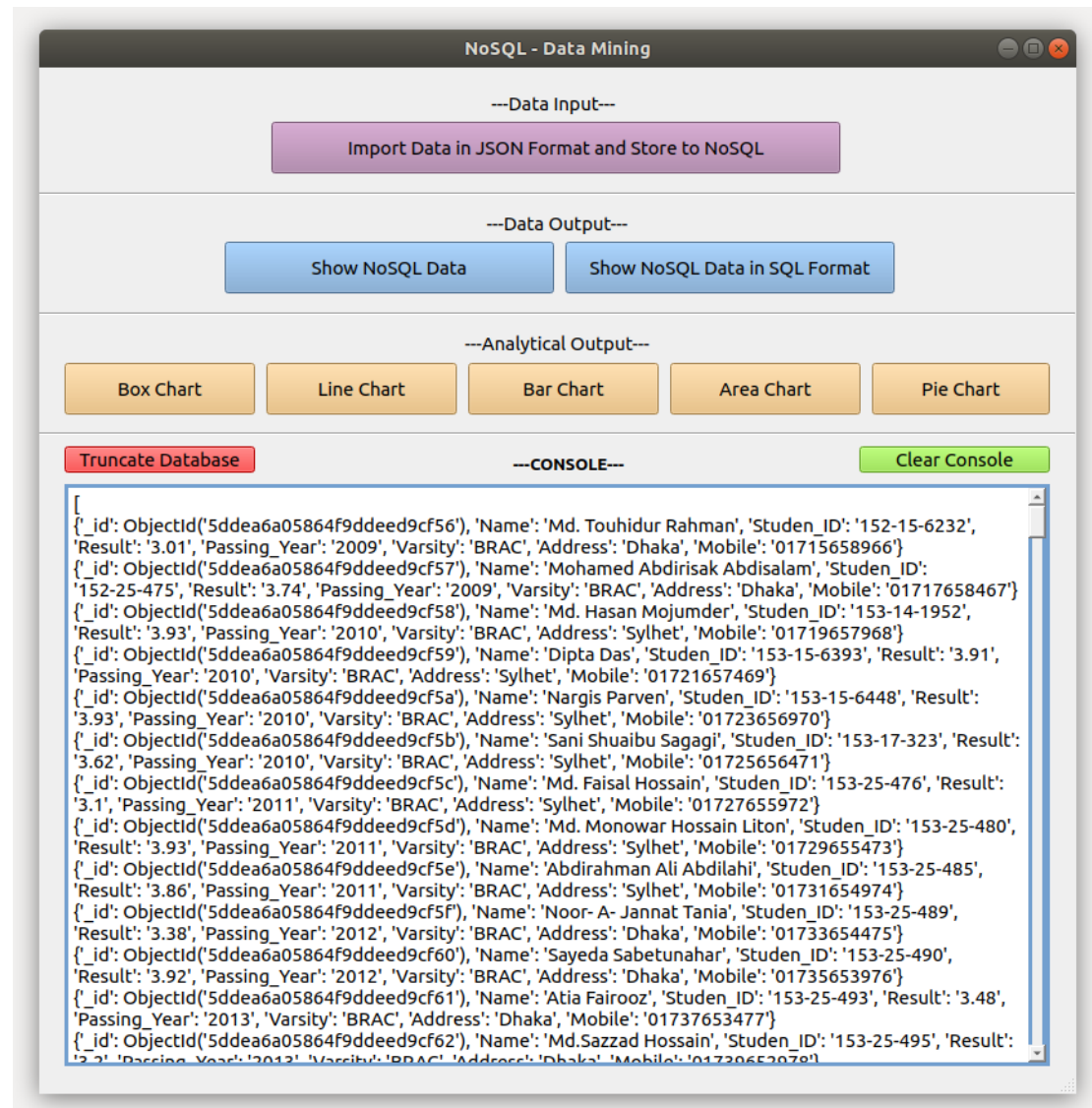


Figure 4.2.4: Displaying NoSQL Data

In the system, there are used MongoDB as NoSQL database. And the outputted NoSQL JSON data directly retrieve from MongoDB.

The stored data also can be displayed in SQL format by clicking on the “Show NoSQL Data in SQL Format”, as shown in Figure 4.2.5.

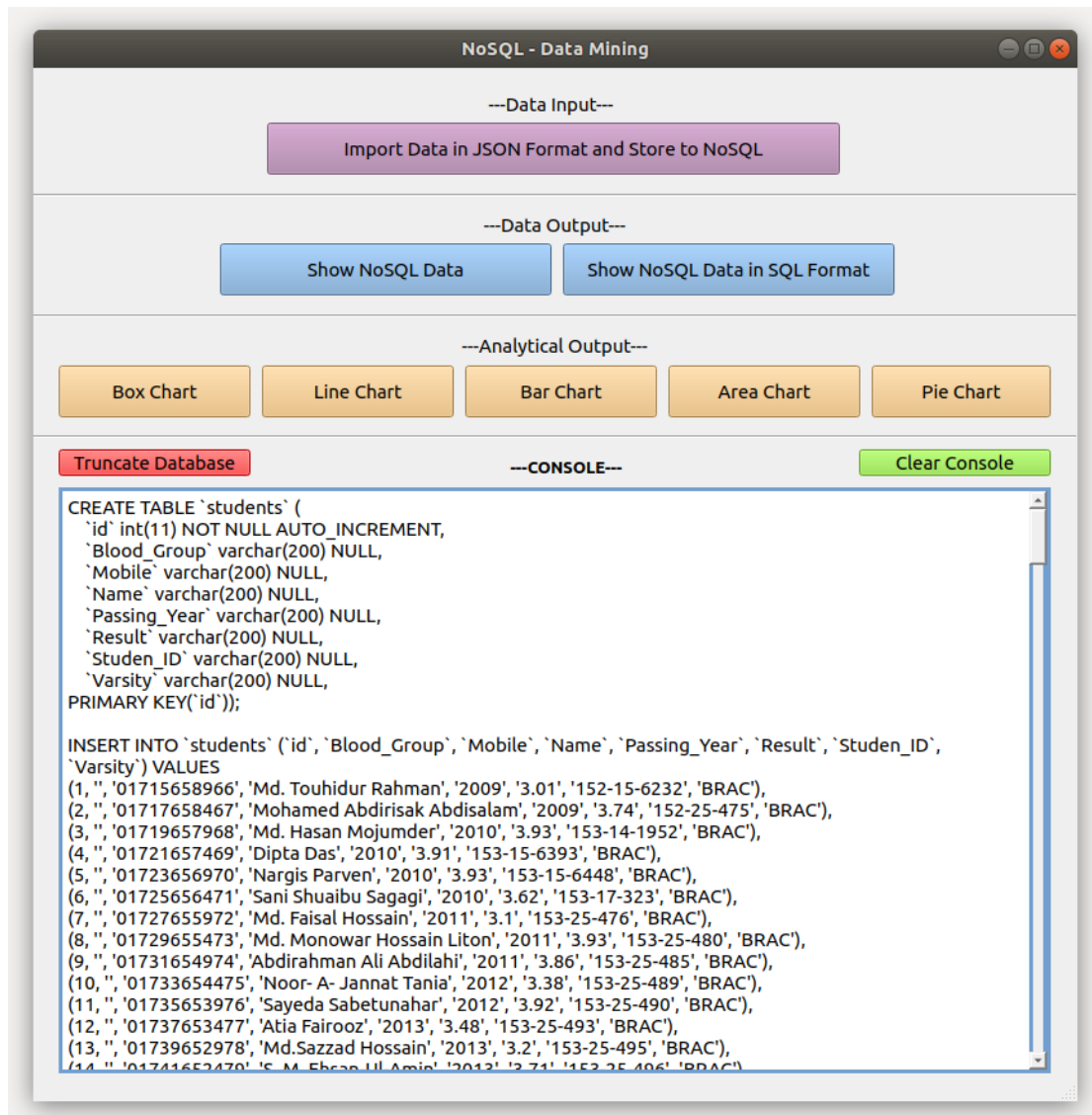


Figure 4.2.5: Displaying NoSQL Data in SQL Format

For analysis the collected data in graphical form, we using data mining technique to represents these information. Figure 4.2.6 shows the Box Chart based on the analyzed data.

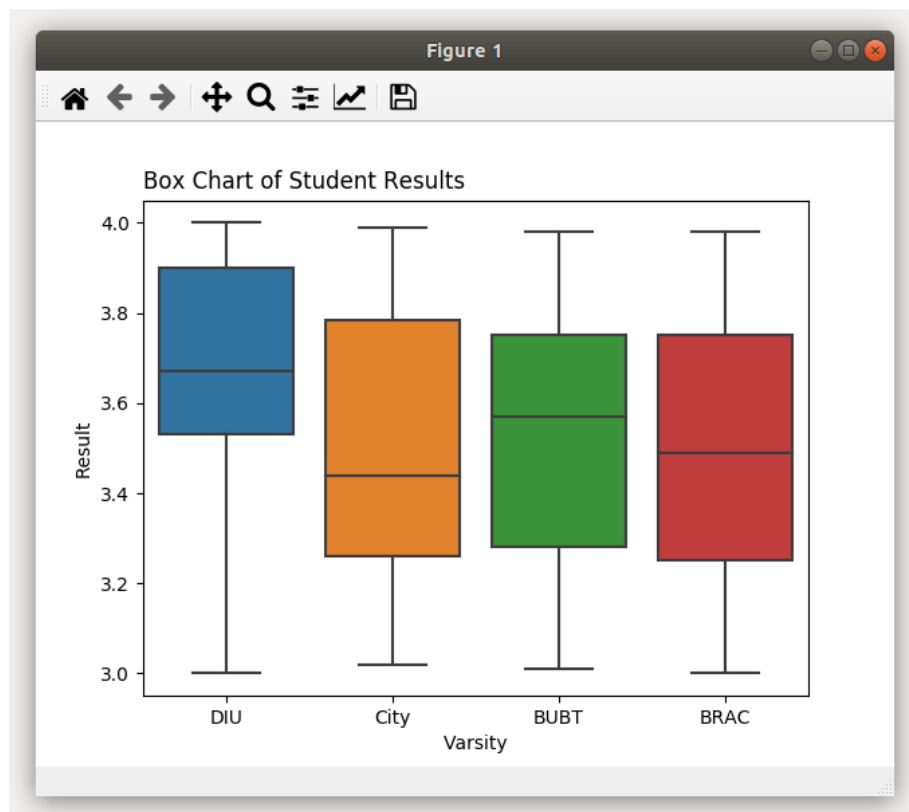


Figure 4.2.6: Box Chart of Analyzed Data

For displaying analyzed data via line chart, at first we need to click “Line Chart” button, which represents the passing year of students from different university, as shown in Figure 4.2.7.

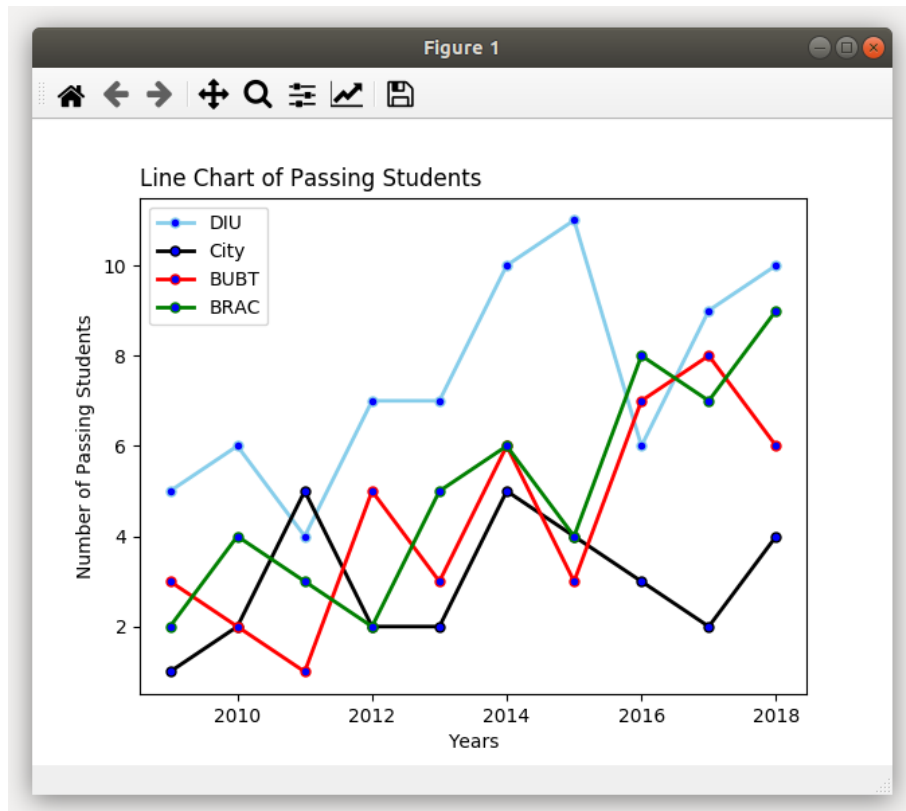


Figure 4.2.7: Line Chart of Analyzed Data

For displaying analyzed data through bar chart, at first we need to click “Bar Chart” button, which represents the area-based number of students of various universities, as shown in Figure 4.2.8.

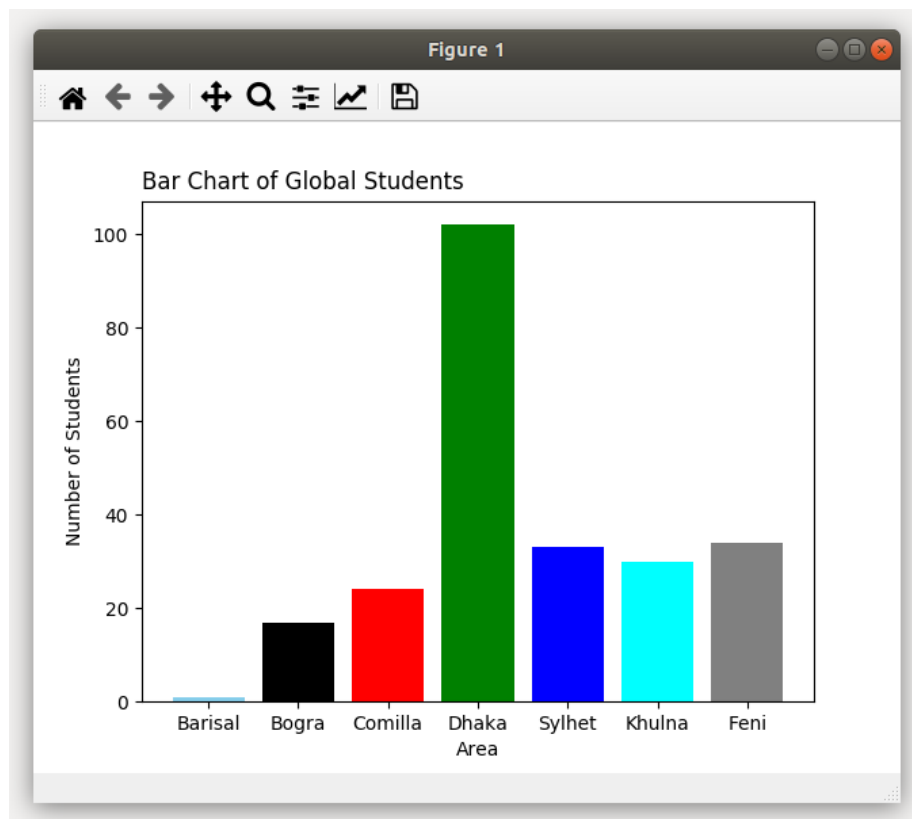


Figure 4.2.8: Bar Chart of Analyzed Data

For displaying analyzed data through area chart, at first we need to click “Area Chart” button, which represents the yearly number of passing students of various universities, as shown in Figure 4.2.9.

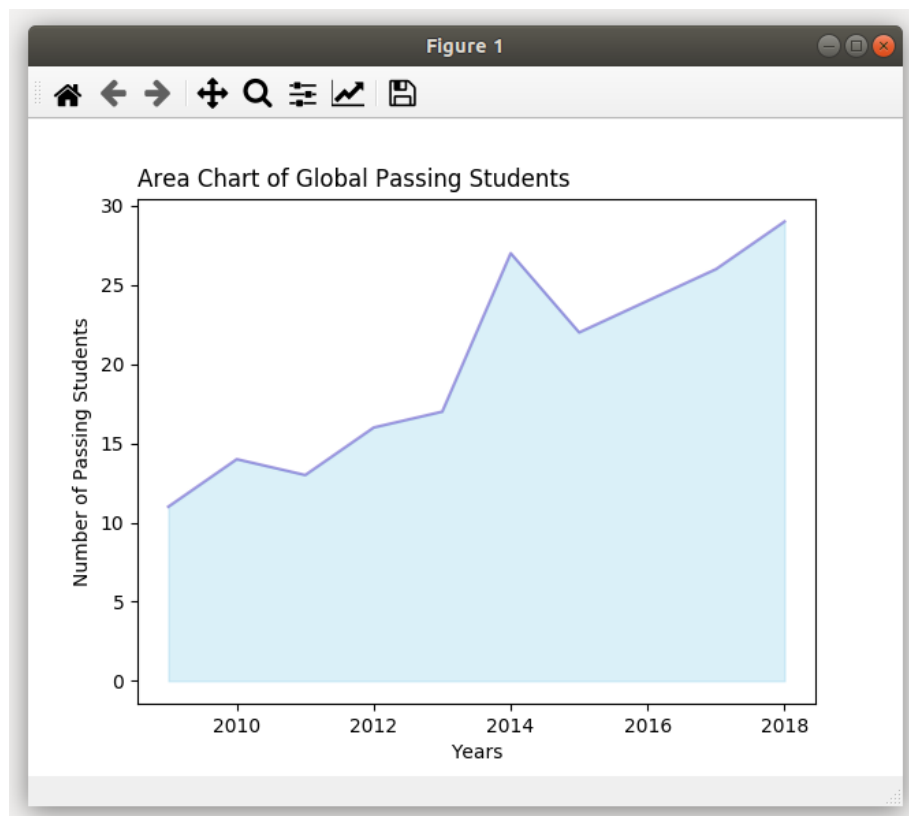


Figure 4.2.9: Area Chart of Analyzed Data

For displaying analyzed data through pie chart, at first we need to click “Pie Chart” button, which represents the number of students of various universities, as shown in Figure 4.2.10.

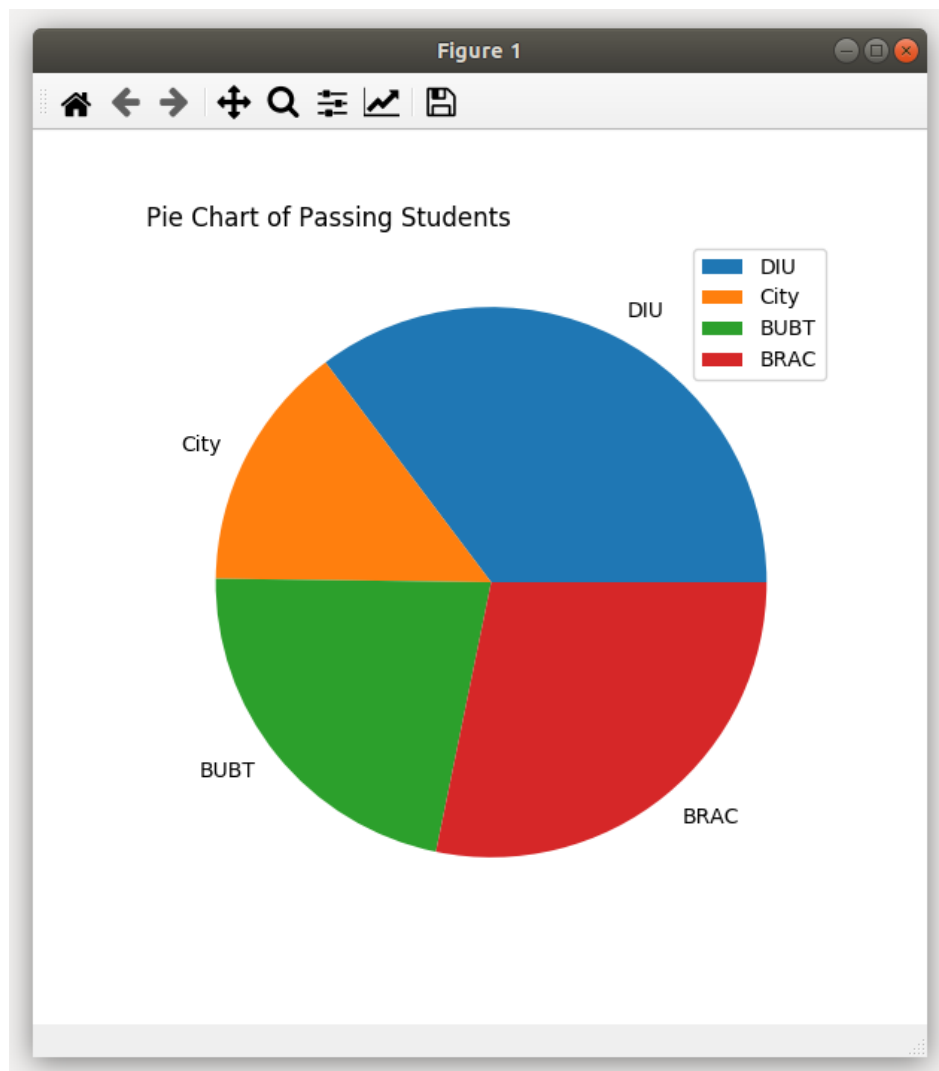


Figure 4.2.10: Pie Chart of Analyzed Data

### **4.3 Descriptive Analysis**

The thesis work gives us the highest accuracy comparing with other thesis. There are many research works, nearly closed to our work. Our developed model provides better result efficiently as compare to conventional data-mining model.

### **4.4 Summary**

In this chapter, we mainly discuss about experimental of our developed model and analyzed output result briefly. Our model performs in the satisfactory level.



## **CHAPTER 5**

### **SUMMARY AND CONCLUSION**

#### **5.1 Summary of the Study**

The proposed model architecture obtains from evaluation real life implementation of educational database. In this thesis, covered implementation of data repositories based on relational database. We are creating a large amount of data in educational repositories that is not being utilize perfectly. We can get advantage from the large educational dataset, by using data mining processes, which serve as stronger tools for characterization, investigation, analysis and prediction. Nowadays, many people are using data mining techniques in various field but it is limited used in educational perspectives. Moreover, it is more limited for NoSQL repositories. In our proposed model, we used NoSQL database, which provide an efficient way to retrieve data from database that contains a large volume of data. Who need to analysis data from such database can be able to retrieve data in a short time to use that information for analysis, characterize, investigation and preparing valuable reports. This model can be helpful for any organizations especially for educational institutions, which play with large volume of data. This model will provide advantages to the both students and teachers also and they can support their institutions by quality assurance of education system. It is more effective to improve our education system.

#### **5.2 Conclusions**

Applying data mining techniques with NoSQL repositories on educational data. Which discloses some significant region on education system, where analysis with data mining has acquired benefits, like identification performance of different student, comparison between different teachers, students and educational institutions, identify students satisfaction for a specific area, student evaluation, analysis students course registration planning, analysis the enrolment head count and so on. The thesis primarily investigated, analyzed and compare student performance with NoSQL database for achieving efficient data mining from various data sources. The study has developed a significant system, which can be apply for analysis student activity. The study of this thesis can be make easy to analysis student performances, global

educational environment evaluation and so on. It is more desirable to improve education system.

### **5.3 Recommendations**

This model can be applied anywhere to analysis both structured data and unstructured data. It is also helpful to data analyst for analyzed of big volume of different data in short time with appropriate analysis. New researchers, who want to work with NoSQL will be benefited from this thesis.

### **5.4 Implication for Further Study**

In this thesis, we have done some quality work distinctly. The thesis can be very useful to future research and project related to educational areas, especially for whom, who want to work with NoSQL repositories. The people interested to study on data mining, also get benefits from this thesis.

## REFERENCES

- [1]. P. C. ZIKOPOULOS, C. EATON, D. deROOS, T. DEUTSCH, AND G. LAPIS, "Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data," Published by McGraw-Hill Companies, 2012.
- [2]. Mazurek, Marcin., Applying NoSQL Databases for Operationalizing Clinical Data Mining Models, Springer International Publishing, 2014, pp. 527--536.
- [3]. Moniruzzaman, A B M & Hossain, Syed. (2013). NoSQL Database: New Era of Databases for Big data Analytics - Classification, Characteristics and Comparison. *Int J Database Theor Appl.* 6.
- [4]. Bansel, Aryan & Gonzalez-Velez, Horacio & Chis, Adriana. (2016). Cloud-Based NoSQL Data Migration. 224-231.
- [5]. B. Jose and S. Abraham, "Exploring the merits of nosql: A study based on mongodb," 2017 International Conference on Networks & Advances in Computational Technologies (NetACT), Thiruvanthapuram, 2017, pp. 266-271
- [6]. M. M. Rovnyagin et al., "Modeling NoSQL Systems in Many-nodes Hybrid Environments," 2017 IEEE 11th International Conference on Application of Information and Communication Technologies (AICT), Moscow, Russia, 2017, pp. 1-4.
- [7]. Arnaud Castelltort, Anne Laurent. Extracting fuzzy summaries from nosql graph databases. *FQAS: Flexible Query Answering Systems*, Oct 2015, Cracow, Poland. pp.189-200
- [8]. Bifet, Albert., Mining big data in real time. *Informatica (Slovenia)*, 2013, pp. 15-20.
- [9]. Bajarski, P., Augustyn, D.R., Bach, M., Brzeski, R., Duszeko, A., Aleksandra, W.: Databases vs. cloud computing. *Studia Informatica* 33(2A), 9–25 (2012)
- [10]. Emam, K.E.: Guide to the De-Identification of Personal Health Information. CRC Press (2013)
- [11]. Groves, P., Kayyali, B., Knott, D., Van Kuiken, S.: The 'big data' revolution in healthcare. Tech. rep., McKinsey & Company (January 2013)
- [12]. International Health Terminology Standards Development Organization (IHTDSDO): SNOMED CT.
- [13]. Khan, A., Doucette, J., Jin, C., Fu, L., Cohen, R.: An ontological approach to data mining for emergency medicine. In: 2011 Northeast Decision Sciences Institute Conference Proceedings 40th Annual Meeting, Montreal, Quebec, Canada, pp. 578
- [14]. Savage, N.: Better medicine through machine learning. *Commun. ACM* 55(1), 17– (2012), *Communications of the ACM*, January 2012, Vol. 55 No. 1, Pages 17-19 10.1145/2063176.2063182
- [15]. Yu, S., Rao, B.: Introduction to the special section on clinical data mining. *SIGKDD Explor. Newsl.* 14(1), 1–3 (2012), 2408736.2408738

# Sumon Deb

*by* Sumon Deb

---

**Submission date:** 02-Dec-2019 06:27PM (UTC+0600)

**Submission ID:** 1225068043

**File name:** Report-V2.pdf (1.21M)

**Word count:** 4177

**Character count:** 22994

# Sumon Deb

## ORIGINALITY REPORT

7%

SIMILARITY INDEX

3%

INTERNET SOURCES

0%

PUBLICATIONS

5%

STUDENT PAPERS

## PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	3%
2	savoirs.usherbrooke.ca Internet Source	1%
3	www.coursehero.com Internet Source	1%
4	Submitted to University of Glasgow Student Paper	1%
5	d-nb.info Internet Source	<1%
6	istqbexamcertification.com Internet Source	<1%
7	repository.dl.itc.u-tokyo.ac.jp Internet Source	<1%

8

Submitted to University of Nottingham

Student Paper

<1%

9

baadalsg.inflibnet.ac.in

Internet Source

<1%

10

docplayer.net

Internet Source

<1%

11

Submitted to The New Art College

Student Paper

<1%

Exclude quotes

Exclude bibliography

Off

On

Exclude matches

Off