Scientific Journal Impact Factor (SJIF) : 4.364
International Journal of Modern Trends in Engineering and Research
ISSN (ONLINE) : 2349 - 9745 ISSN (PRINT) : 2393 - 8161

AN ENHANCED TEXT MINING TECHNIQUE USING BIG DATA FOR STUDENTS ATTENDANCE MANAGEMENT SYSTEM

Deepankar Bharadwaj[1] and Dr. Arvind Kumar Shukla[2]
[1]*Research Scholar IFTM University*
[2]*Associate Professor, IFTM University*

Abstract- Attendance Management System plays a significant role in managing the attendance record of Institutes and Universities students. There are lots of methodologies in the current scenario which provides computerized system for attendance data management. These automated systems works on database architecture for storing the attendance data. The entire perception behind these systems is to store and to use the data in databases with different technologies. With all these inputs and our survey it is found that in old concept, there is problem of the size of these databases. In this manuscript an attempt has been made to introduce a new concept of storing the data in the database for attendance management system which decrease its size and also reduce data redundancy in a very efficient manner.

Keywords- Attendance Management System, Big Data, Text Mining, Data Mining

I. INTRODUCTION

Student is the largest union in the study environment of a university, so it is hard for managing student things. Especially, in the respect of student class attendance, the original named style is hard to response the really situation of student attendance [5]. Now-a-days, there are many Universities, around the world that are having huge number of students enrolled. [2] To handle these large set of students, it is very difficult to keep their records in manual state. It is a very big problem especially to get the attendance of these students. In earlier days, the attendance were taken manually and the process was that whenever a lecturer comes to class, he came with a register, manually takes attendance by calling roll-numbers and at the end of month, he has to calculate the monthly attendance of each and every student. The major problem is that the lecturer or teacher had to take care of the register and enter the attendance into the log (or) data-base, calculate the attendance percentage and give the marks on its basis. It is also difficult to find out those students whose attendance is below average. This would be a big problem in Universities. The suitable solution for this problem is by designing a system that will store record of attendance in the system and automatically calculate the desired results from it. In today' s world, automated attendance is generally taken by various sources such as Bio Metric devices, RFIDs, Face Recognition Kits etc. There is a rapid increase in the data in databases of Attendance Management Systems with these sources that are current being used. This paper consists of a study to improve the current concept of student attendance system by providing new ideas of keeping the record of the students whose attendance is being captured.

II. DRAWBACKS IN PRESENT SYSTEM

In the current system we found some disadvantages on various cases. Following are the disadvantages we found

1. Data Redundancy.
2. Data Inconsistency.
3. Data Corruption.
4. Data Size.
5. Data Efficiency.

Data redundancy is a very common attribute of the modern computer database. Fundamentally, it is a repetition of data within a database. It can appear within a table as repeated data, or it can appear by itself. It is typically not something that the programmer intended to be part of the system; however, it is a by-product of programming without forethought [6]. The duplicated data, also known as the redundant data, creates unreliable information because the chances of having a value changed in one file are high, but on the other file the value remains the same [7]. This condition of inconsistency is often experienced when using the traditional file processing, and it is very expensive and difficult to rectify such inconsistencies.

Data Corruption means the errors in the data stored and creates error while producing outputs. If the system is having duplicated or redundant data then there are much more changes of error in the data.

The redundant data means duplicate data in database. Consider the example of a employee table having his personal records and due to some reasons, the employee data is duplicated. Then the size of the database will increase automatically.

Data Efficiency [8] refers to efficiency of the many processes that can be applied to data such as storage, access, filtering, sharing, etc., and whether or not the processes lead to the desired outcome within resource constraints

III. LITERATURE REVIEW

All the universitites, institutions are capturing attendance record of students and it is very important for various purposes and its one of the important criteria is to follow for students. In older days capturing student attendance record was manual approach. This method consumes lots of time and it was a very tough task to capture and maintain attendance data. Keeping the record and maintaining the record years to years and finding the values as per user's requirement, is a very hard.

This method was not able to work efficiently in the case of students where student count is in bulk. These issues leads to the generation of automated attendance management systems which automates the whole process of capturing attendance very fast and in effective manner.Initially these systems work on manual process to capture the attendance but slowly technology has change the method of capturing student attendance. Initially the user manually enters the attendance of each and every student and system stores the data in database but now a day's the systems captures all the required information automatically and with the use of latest technologies and devices, a complete automated system can keep the attendance records.

Some of the common used techniques are: Bio-Metric Devices for Finger Print, Bio Metric Devices for Face & Retina Scan, RFIDs. [1] A fingerprint device was introduced to capture the human attendance based on its fingure print. The system initial stores the fingure print of all and on the basis of input of prints, the system marks attendance. The technology is samrt to decide the which data needs to be captured and it becomes very easy to capture the attendance data. [4] Another method introduced was Face & Retina Scan when a user enters in field of view of the webcam, the system detects the face, and human eye ratina. Similar process of capturing the attendance can be by face and ratina recognition.

The face verification result is valid for a period of time, after which a new verification is needed. [3] Next and very effective used technology is RFID. RFID (Radio Frequency Identification) is a technology that uses radio waves to transfer data from an electronic tag – called an RFID tag or label, which is attached to an object – through a reader for the purpose of identifying and tracking the object.

RFID systems have been widely used in many different application areas, such as: product tracking through manufacturing and assembly, control of inventory, parking lot access and control, container tracking, ID badges and access control, equipment tracking in hospitals, etc.

IV. ATTENDANCE MANAGEMNT SYSTEM

In any automated system, database design plays a key role in development as it is the back bone for any automated system to work effetively and efficiently. In case of automated attendance management system it is also important to design the database in such a manner that it will work effetively and efficiently. The database is designed with the attributes in such a manner which decrease data duplicacy and maintain data integrity. These attributes will work as a structure of proper functioning of attendance management system and also provides the good solution. But as per our study in the existing architecture of any database structure of attendance, the database occupies much space in the memory. As per our survey the data in these tables is redundant and inconsistent. There is repetition of data in the database tables that increases database size which may compromise data security. Bulk amount of data is stored according to the old methods which may raise the issues of data redundancy and data inconsistency. As per the inputs from other attendance management softwares, the database has minimum following attributes to keep the record of the attendance.

Table 1: Structure of Attendance Table

ATTRIBUTE NAME	DATA TYPE	CONSTRAINT
ID	INT	PRIMARY KEY
STUDENT_ID	INT	FOREIGN KEY
COURSE_ID	INT	FOREIGN KEY
BRANCH_ID	INT	FOREIGN KEY
SEM_ID	INT	FOREIGN KEY
LEC_NO	INT	NOT NULL
DATE	DATE	CURRENT DATE
TIME	TIME	CURRENT TIMESTAMP
STATUS	TINYINT	DEFAULT 0

Following is the Data Flow Diagram of a basic automated attendance management system. It is clear with the diagram that there are basic two modules in this system: Attendance Entry & Report Generation. After attendance entry, the data is being stored in the database which can be further used to generated the MIS reports.

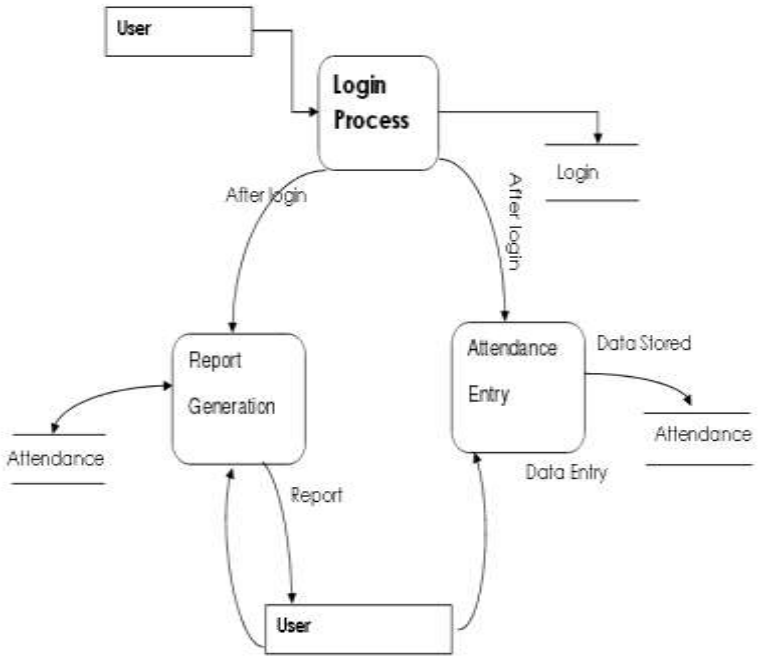

FIGURE 1 data flow diagram for attendance system

Students Attendance

Sno	Batch Name	Timings	Take Attendance
1	Batch-1	XX to XX	Take Attendance
2	Batch-2	XX to XX	Take Attendance
3	Batch-3	XX to XX	Take Attendance

FIGURE 2 Snapshot 1 For Attendance Management System

Take Attendance

Sno	Image	Enrollment No	Student Name	Father Name	Present	Absent
1		015020	ANJALI	MADAN LAL	Present	Absent
2		0130099	ASHISH CHITRSNSH	PREAM PRAKASH SAXENA	Present	Absent
3		0130099	ASHISH CHITRSNSH	PREAM PRAKASH SAXENA	Present	Absent

Upload

FIGURE 3 Snapshot 2 For Attendance Management System

Above are some snapshots of an attendance management system which captures the attendance record.

V. DRAWBACKS OF OLD METHOD

Let us consider a case of an institute or a university having students in thousands for ex: 20,000. In universities there are six to seven lectures scheduled everyday for each and every class. And on average for each class there are approximately 50 students present daily. As per old method, the system keeps a single record entry for every single student for every single lecture. According to study, and as discussed earlier about the minimum 10 attributes required. [9] From the study we found that 6 attributes are of INT type each of which occupies 4 bytes per attribute per record. 2 are of DATE & TIME type which occupies 3 bytes per attribute per record, 1 is of TINYINT type which occupies 1 byte per attribute per record.

From the above inputs if we calculate the size of a record with these 10 attributes, it will be as follows:

INT:	4 Bytes * 6 Fields	24 Bytes
DATE:	3 Bytes * 1 Field	03 Bytes
TIME:	3 Bytes * 1 Field	03 Bytes
TINYINT:	1 Bytes * 1 Field	01 Byte
		31 Bytes

These 31 bytes data will be stored for per student, per lecture, per day. As mentioned earlier that we are consedering the case of colleges, institutions, universities having thousands of student there are six to seven lectures daily which will be calculated to 217 bytes for 1 student per day. If we calculate it for a class having 50 students for a day, it will be 10850 bytes or 11 KB per day per class. In universities, the student count is in thousands then this will be 43,40,000 bytes or 4238 kb or 4.13 MB per day for 20,000 students. For a month with 25 days working, it will be 103 mb in a month and same for a year, considering 200 days as working, it will be approx 0.81 GB in one year. This means approx 1 GB data is storing every year for every university who are using the automated attendance systems. To overcome this issue we have introduced a new concept of storing the data in databases which not only decrease the database size but also reduce data redundancy and data inconsistency.

VI. PROPOSED CONCEPT

We have introduced a new concept of storing the data of an automated attendance management system in such a manner that it helps in reducing database size, data redundancy and data inconsistency. From this manuscript an attempt has been made to introduce a new methodology that can easily keep the record of all student attendance in all lectures / periods with reducing database size and data redundancy. To implement the proposed concept, any database can be used. In this paper, we have used an open source database mysql, which can work on apache environment with some minimum system requirements.

Table 2 Proposed Table Architecture

Attribute Name	Data Type	Constraint
Id	INT	PRIMARY KEY
Student_Id	INT	FOREIGN KEY
Course_Id	INT	FOREIGN KEY
Branch_Id	INT	FOREIGN KEY
Sem_Id	INT	FOREIGN KEY
Attendance_No	INT	DEFAULT 0
Date	DATE	CURRENT DATE
Time	TIME	CURRENT TIMESTAMP

In Table 2 architecture, there will be single record for a single student in a single day. As mentioned earlier that in universities, there are maximum six to seven lectures daily. Consider a binary number of eight (8) bit say 00000000. Each and every bit represents the lecture no. Any student is present in Ist, IInd, IIIrd and Vth lecture then for that student the binary no will be like 00010111. Equivalent decimal representation of this binary number is 23. This 23 will be stored in the table in Attendance_No as integer value.

Whenever system mark attendance of a student, the system fetches the Attendance_No from the database, convert the value in binary number, change the bit number of the lecture no to 1 (if mark as present) or 0 (if mark as absent), convert to decimal and store in the database.

In the old system, there will be four (4) entries in the database with the above case but in this concept, there will be only one entry per day per student.

INT:	4 Bytes * 6 Fields	24 Bytes
DATE:	3 Bytes * 1 Field	03 Bytes
TIME:	3 Bytes * 1 Field	03 Bytes
		30 Bytes

These 30 Bytes is of per student in per day. If we calculate it for a class having 50 students for a day, it will be 1700 Bytes or 2 KB per day per class. If we take 20,000 students then this will be 68,000 Bytes or 66 KB per day for 20,000 students. For a month with 25 days working, it will be 2 MB in a month and same for a year, considering 200 days as working, it will be approx 13 MB in one year. Data has been reduced from GB to MB. Following is the representation of the proposed concept in tabular format.

VII. ALGORITHM IMPLEMENTD

In this concept, the following algorithm has been used.

1. Begin Process
2. Initialize the value to 0.
3. Take the Lecture No, Status from the Input.
4. Select the value from database of the student.
5. Switch (Lecture No)
6. Case 1:
a. bin = Call dec_to_bin (value);
b. change = Call change_bit (bin, 1, status);
c. dec = Call bin_to_dec (change);
7. Case 2:
a. bin = Call dec_to_bin (value);
b. change = Call change_bit (bin, 2, status);
c. dec = Call bin_to_dec (change);
8. Case N:
a. bin = Call dec_to_bin (value);
b. change = Call change_bit (bin, N, status);
c. dec = Call bin_to_dec (change);
9. End Switch
10. Update the value to database of same student.
11. End Process

VIII. RESULT

Details View Page (In Bytes)			
SNo	Particulars	Old Method	New Method
1.	One Attendance Record	16384 B	16384 B
2.	Record per day per student	16384 B	16384 B
3.	Per day per class of 50 students	49152 B	16384 B
4.	Per day full university of 20,000 students	9977856 B	1589248 B
5.	Per Month full record (25 Days)	249446400 B	39731200 B
6.	Per Year full record (200 Days)	1995571200 B	317849600 B

Figure 4 Result in BYTES (B)

Details View Page (In KBs)			
SNo	Particulars	Old Method	New Method
1.	One Attendance Record	16 KB	16 KB
2.	Record per day per student	16 KB	16 KB
3.	Per day per class of 50 students	48 KB	16 KB
4.	Per day full university of 20,000 students	9744 KB	1552 KB
5.	Per Month full record (25 Days)	243600 KB	38800 KB
6.	Per Year full record (200 Days)	1948800 KB	310400 KB

Figure 5 Result in KILOBYTES (KB)

Details View Page (In MBs)			
SNo	Particulars	Old Method	New Method
1.	One Attendance Record	0.015625 MB	0.015625 MB
2.	Record per day per student	0.015625 MB	0.015625 MB
3.	Per day per class of 50 students	0.046875 MB	0.015625 MB
4.	Per day full university of 20,000 students	9.515625 MB	1.515625 MB
5.	Per Month full record (25 Days)	237.890625 MB	37.890625 MB
6.	Per Year full record (200 Days)	1903.125 MB	303.125 MB

Figure 6 Result in MEGABYTES (MB)

Details View Page (In GBs)			
SNo	Particulars	Old Method	New Method
1.	One Attendance Record	1.52587890625E-5 GB	1.52587890625E-5 GB
2.	Record per day per student	1.52587890625E-5 GB	1.52587890625E-5 GB
3.	Per day per class of 50 students	4.57763671875E-5 GB	1.52587890625E-5 GB
4.	Per day full university of 20,000 students	0.0092926025390625 GB	0.0014801025390625 GB
5.	Per Month full record (25 Days)	0.23231506347656 GB	0.037002563476562 GB
6.	Per Year full record (200 Days)	1.8585205078125 GB	0.2960205078125 GB

Figure 7 Result in GIGABYTES (GB)

IX. CONCLUSION AND FUTURE DIRECTIONS

Through this new introduced concept, we have reduced the size of attendance table. This concept is very useful for industries that are developing automated attendance management solutions. This will help programmers to implement new technique of storing data in database and also able to reduce data redundancy and data inconsistency. Database management is a very broad and relevant area of research. The future directions in this topic may be that this technique may be applied on other modules also with some modification.

REFERENCES

[1] "An Attendance Monitoring System Using Biometrics Authentication" International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume 3, Issue 4, April 2013
[2] Student Management System based on RFID Technology. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), ISSN 2278-6856, Volume 2, Issue 6, November – December 2013
[3] Attendance Control System based on RFID-technology, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 1, May 2012 ISSN (Online): 1694-0814 www.IJCSI.org.
[4] Multimodal Biometrics - based Student Attendance Measurement in Learning Management Systems. 2009 11th IEEE International Symposium on Multimedia. 978-0-7695-3890-7/09 2009 IEEE DOI 10.1109/ISM.2009.25.
[5] The Research and Application of College Student Attendance System based on RFID Technology. International Journal of Control and Automation Vol. 6, No. 2, April, 2013
[6] http://www.ask.com/technology/disadvantages-data-redundancy-9a30353bc8edf864.
[7] http://www.ask.com/math/definition-data-inconsistency-5bc80c9fd30c5f1a.
[8] https://en.wikipedia.org/wiki/Data_efficiency
[9] https://dev.mysql.com/doc/refman/5.1/en/storage-requirements.html

