

Chapter 1

Analysis of BIIS Data

1.1 Scope

BUET offers educational facilities for more than 5000 students in an academic session. These students are under departments. Number of students in a respective department is different according to the department facility. Department of Electrical & Electronic Engineering, Department of Civil Engineering, Department of Mechanical Engineering can afford 195 students per batch. Department of Computer Science & Engineering has 120 students. Department of Architecture afford 55 students, Department of Water Resource Engineering, Department of Materials & Metallurgical Engineering, Department of Urban & Regional Planning and Department of Industrial & Production Engineering hold 30 students each. Department of Naval Architecture & Marine Engineering and Department of Chemical Engineering has 60 students each [1]. These numbers have been changed throughout year by year but it represent the period on which this research has been more focused. In this research BIIS data of 10 already graduated batches have been used which represents almost 10 thousand students. Academic performance of no current student has been brought under this research.

1.2 Database Structure

There are 8 parts of BIIS data and each data sheet has 65,000 tuples. Each tuple represents the performance of a student in a particular course. The attributes of the tuple are described below

- Serial : It is the serial number of a particular student. It is not actually the student id, but it's a unique constraint and primary key of the database table.

- Department : Respective department of the student.
- HallStatus : Indicates whether the student is hall resident or attached.
- District : Home district of the student.
- Thana : Thana of the student.
- Gender : Gender of the student.
- PlaceOfBirth : Birth place of the student.
- StartingOfAcademicYear : Indicates on which academic year the student has started undergrad.
- AdmissionDate : Admission day at BUET.
- CourseName : The particular course taken by student.
- LetterGrade : Lettergrade of the respective course.
- ClassAttendanceMark : It represents the attendance mark of the student in respective course. The attendance mark is allocated 10% of total course mark. E.g., For 3 credit courses, it's marked out of 30, for 4 credit course it's marked out of 40.
- ClassTestMark : Class test mark is allocated 20% of the total number of a course. E.g., For 3 credit course, highest ct mark is 60, for 4 credit course it's 80.
- PartAMark : For theory courses the term final examination holds 70% weight of total course. This huge exam is divided into 2 parts. Generally internal examiner handles Part A. E.g., For 3 credit course, part A holds 105 marks, for 4 credit courses it's 140.
- PartBMark : Same as part A, generally external examiner handles part B.
- TotalNumber : Total mark obtained by a student in a particular course..
- LevelName : The level respective course has been taken.
- TermName : The term respective course has been taken.
- GPA : GPA of student in particular term.
- CGPA : CGPA upto the respective term.

- TermCount : Term count of the student.
- CreditHourEarned : Credit hour earned by the student in the respective term.
- TotalCreditHourCompleted : Total credit hour earned by the student upto respective term.

1.3 Problems in Existing Structure

The universal database holds raw academic data of the students. There are plenty of disturbance in this database. They are described below :

- The first and most important problem is, all the eight data sheets don't have same attributes among them. For example, Sheet 1 doesn't have a *CourseName* column whereas Sheet 2 has one. There are many more incidents where attributes don't match.
- One student enrolls into a number of courses in academic career which can vary from 65 to 80. So, there are multiple tuple for one student which is not suitable for classification. Determining status from multiple tuples is very difficult.

Serial	Department	...	CourseName	LetterGrade	...
2	CE	...	PHY143	A-	...
2	CE	...	MATH141	B+	...
...

Table 1.1: Multiple entries for single student in the Universal Database

- The database is in a format of full outer join. So, there are plenty of blank attributes. For example, in sessional courses, there is no Part A or Part B. So, *PartAMark* or *PartBMark* remains void for sessional courses. Besides, there are also some special course for which maximum attributes remain blank.

Serial	Department	...	CourseName	PartAMark	PartBMark	...
65	CSE	...	CSE100	-	-	...
65	CSE	...	MATH141	93	112	...
...

Table 1.2: Blank attributes in Database

- There are some attributes, for which there is no value was really inserted ever. In some datasheets *PlaceOfBirth, District, Thana* are blank.
- In some data sheets, *CGPA & GPA* have rounded value. But, for accurate results, it is very important to get the accurate *CGPA & GPA* upto two floating points at least.

Serial	Department	...	GPA	CGPA	...
4478	NAME	...	4	3	...
4478	NAME	...	3	3	...
...

Table 1.3: Rounded up value of CGPA and GPA

- There is no *CourseCredit* attribute in the database. But for proper reasoning, *CourseCredit* is very important.
- Some attributes like *TotalCreditHourCompleted, AttendanceMarks, ClassTestMark* don't have the same range of value for all the tuples. So, any reasoning from these values is very difficult.

Serial	Department	...	Attendance Mark	ClassTest Mark	Total CreditHour Completed	...
65	CSE	...	27	44	160	...
120	ARCH	...	40	71	190	...
...

Table 1.4: Different range of values

- There are many confusing attributes like *TermCount, AdmissionDate* which don't actually make sense.
- Reasoning for Theory and Sessional courses are entirely different, but this database don't indicate whether a course is Theory course or not.

Bibliography

- [1] Bangladesh University of Engineering and Technology, General Information on: <http://www.buet.ac.bd>