Intelligence Performance In Students Absence System With Predicted Information By Data Mining Algorithms

Ali Fattah Dakhil*[1], Waffa Muhammad Ali[2], Asseel Jabbar Almahdi[3]
[1]University of Sumer, Faculty of Computer Science and Information Technology, Computer Science Dept., Thi-Qar, Iraq
E-mail: a.dakhil@edu.salford.ac.uk,
[2]University of Thi-Qar, Educational College of Pure Science, Computer Science Dept., Thi-Qar, Iraq
E-mail: wafaam20142015@gmail.com,
[3]University of Thi-Qar, Computer Science and Mathematics College, Computer Science Dept., Thi-Qar, Iraq
E-mail: asseelalmahdi@gmail.com

Abstract: *There are many databases projects used by the numerous number of organizations. However, embedding business intelligence, BI, the technique is a qualitative impact and a quite important factor to improve such project's results and performance. Modern data mining algorithms have dramatically changed our business work, data model and the way we develop software projects. A database application that manages students' attendance used in university classes is the objective scope that we adopt and work on in this paper. The main key of interest in this research is to improve such attendance system by participating one of data mining classification technique in which we then have useful learned information and predicted reports about future students' attendance. Beside this intelligent trait, our work would be crowned with pictorial analytic results that encourage us to have modern and well-mannered intelligent database application.*

Keywords- *Intelligence, Prediction, Data analysis, Students' absences, BI database, Forecasting*

1. INTRODUCTION

Database projects are widely used all around the world, however, most of them are still tradition to deliver the beneficiary needs as basic database tasks. Developing and involving BI tools that are established on a database project, would have significant leverage to ameliorate the desired outcome of these applications. It is commonly understood that data mining is scientific revolution [1] in the discipline of data science and statistic fields. Data mining has many kinds of classification techniques and algorithms. This work focuses on well-known classification methods called decision tree. Later in this research, we would spot light of how to use the decision tree in a proper way to discover the inferred students' absences status in the future. Students' attendance practical project has been made for a collage in order to manipulate absences records during a year of study. After a couple of years in use, we found out that it is necessary to have predicted information and proactive decision besides the tradition reports that the system already have. The outline of this work can be drawn in this preamble by beginning with the used software then students' absence system infrastructure. After that, we launch to discuss and examine our favourite and discriminative algorithm of the decision tree to know then who will attend and who is not under particular circumstances on a day in the future. Later of this research, some analytic tools are applied in order

to garland this work with BI benefits and statistic information. In addition, we will work on measuring an overall trend for a period of time basis on forecasting technique. Also, there are couple of challenges and proposed solutions have been made when practicing this empirical work as well as future needs enhancements. Finally, several results and information are obtained to provide this research with sufficient support to business modelling and organizations.

2. PROPOSAL STATEMENT AND MOTIVATIONS

Using intelligent software is so urgent matter in modern foundations. So, in order to step forward beyond the conventional software, this practical work and theoretical research stand up to employ data mining approaches and proactive solutions in students' absences database. The actual need for this work is to support collages and schools with foretelling and well-analyzed decision. Real case scenario happens occasionally in our example case, such as Universities of Sumer and Thi-Qar, where we work with, when we always suffer from postponing exam date because sometimes plenty of students are absent, and that is because we do not have enough expectation of students' status in that fixed date of exam. Thus, when this work supports us with information about how many students would come, and how many do not. If this comes true, it would be a great job we are looking for. In addition, in order to have a trend of attendance conditions during a period of time, so this will guide us for proper managing on student's attendance and then to manipulate such situations with correct decisions. So, this research and application will support such educational organizations like universities to have a prior guess about the student attendance in a specific date. Therefore, a college staff can manage their work based on this fact.

Two examples that practiced this concept and have inspired us to adopt this research and apply it in the educational field. A research paper based on real scenario application [2] exerts such algorithm and technique on students' records who enrolled in the first year, in order to predict and having prior knowledge of these students for future performance and classify them based on the results of the algorithm. Other related work, an online e-commerce website developed by the researcher [3] to group buyers and their interested goods based on their history purchase and activities. The algorithm thy used is to raise the proper advertisements and promotions for oriented merchants. Those examples and others [4] have influenced us to use and apply such data mining algorithm.

3. SOFTWARE USED

This software project has many tools to produce such required system. MS Access database is the core components to hold data. Access has been chosen as it has high integration with MS Excel. Excel is also used for data analysis and statistic modeling. Recent research by Dakhil A. et al [5] shows how Access and Excel communication has a power of both programs benefits. Also, data mining is the best part that is used as the primary tool, so that with data mining, we apply the key feature of this paper. As mentioned previously, this part has all our mission to work on. Data mining algorithms and techniques are very wide in use, so in this work, we particularly applied decision tree. C# is a visual handy software tools is also used to implement our algorithm and exert the decision tree mechanism.

4. PROJECT MECHANISM AND STRUCTURE

The basic frame for this project lays on some components. Back to what we introduced that this work mainly designed for applying intelligent performance to tradition database systems. So, here we clarify what this project does in a very brief description and leave the major part of this paper to

our goal which is embedding data mining algorithm within developed database application in next sections. Figure 1 shows the components infrastructure of this project.
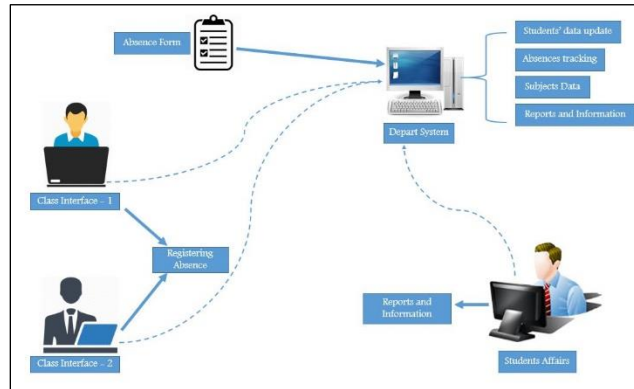

Figure 1**:** Students' Attendance System Infrastructure

In a concise description, this project has some interfaces reside in various locations. The main one locates at the head of the department, so that a user can use it and utilize its functionality. Other interfaces are placed in classrooms so a lecturer can use the system when they read out students' names. The environment frame for this project depends on basis of LAN network so that all clients and terminals are connected and communicated for data saving and retrieving in a live session. Students' affairs department is also provided with an interface to interact with the system. In fact, this project is used by several departments and each one has many of classrooms with at least four years of study in an undergraduate university degree. Also, much of lecturers are practicing this application in daily use. Consequently, there are tens and hundreds of operations are being performed against this database.

4.1 Department Interface

This portion has the main project functionality. These jobs which illustrated in figure 2 are what users can access and use. This is the main interface for the project. As we can see there are many functions this project delivers. First, registering students' absences for four years. Second, managing students' information. Third, some application tasks such as reporting, administrations, and subjects (material) data.
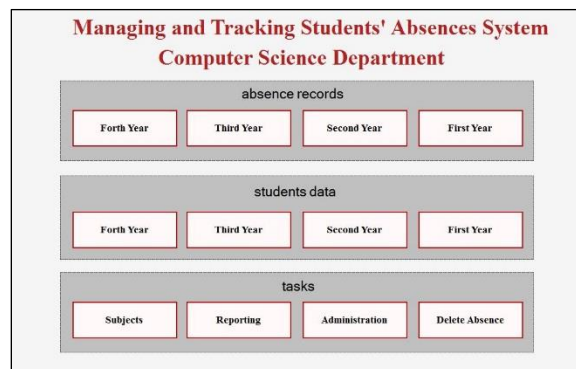

Figure 2: Main System Interface

4.1.1 Absence records Interface

This is to deal with absences records database table. It performs all database tasks, insert, delete and update. This window as in figure 3, is the direct input for data set which we are going to use for mining the data and training tasks later. Some information are involved here that related to student's absences, such as subject (study material), type of lecture, group, date and time of the lecture, lecturer and hours.

Figure 3: Absences Managing Interface

These features are quite enough to use them as an input data set by the chosen data mining algorithm, however, there are some key points are worth to be included about those data set in the algorithm. Later we explain how they play such a rule. So, via this form (interface) we manage the most required information: absences status, students' data, and subjects' information. These three main resources are considered as a data warehouse to fill our data entry. In fact, not all the data inside database tables is used. As mentioned in the introduction, these database tables have been taken advantage to recruit data source for data mining algorithm, so only some data table are used. In the next section, we are going to explain, in details, what they are and why.

4.2 Reporting Service

As a database system, the major mission of its work is reporting. Reporting is the core task in this project. Any empirical application has specific job to deliver, so the reporting services are this project's duty that was developed to. There are some kinds of reports as following:

4.2.1 Students Absences Report

This kind of report focuses only on a specific student based on their #no. A very detailed report is generated which has all the attendance events for each specific class with the following information:

- Date and time of class.
- Subject and material title.
- How many hours that a class lasts.
- Lecturer name, who registers an absence.

So, we can have plenty of information about each student and every subject for a single class. During the time, a student would have a long record that can be used as a source of data for desired report service as well as for data mining applications and forecasting heuristic.

4.2.2 Punishments Reports

The other kind of reports are designed into categories of punishments types. They vary into.

- Alarm, for all those whose absences have reached 3% to 5%.
- First warning, for all those whose absences have reached 5% to 7%.
- Second warning, for all those whose absences have reached 7% to 10%.
- First failed, for all those whose absences have reached more than 10%.
- Second failed, for all those whose absences have reached 15% with exception form first failed group.

These categories are important to classify such students' status. We can get the advantage of these reports by conforming and accordance them with the discovered information. In order to cover our research topic in focus, we only mentioned the basic and the used parts in the system that can be applied in an intelligent manner, otherwise, this system is massive and has a lot do. So, in the next section, we launch to introduce heuristic and analysis part of this work.

5. DATA ANALYSIS AND PREDICTION

In this part of the research, we maintain the significant task which this work aims to convey its main mission. The process of analyzing and categorizing data into useful information is called data mining [6]. Data minding is computerized operation that discover useful information and guides us for patterns in the large set of data. It can sieve messy and noisy data. It helps to find out relevant data in order to guide useful outcomes [7]. So, data mining would hasten the step toward developing knowing decisions. There are seven data mining techniques, association, classification, outlier detection, prediction, regression and clustering [8]. Those various techniques are so diverse and manifold with their details. To concentrate on the paper major idea, we are going to use classification technique as our chosen method and the general topic underneath it, we accomplish our purpose and practice the desired results. Classification is a complex technique among them. It needs to collect the variety of information attributes into categories so that it used to draw predictable results [9]. Also, classification has some techniques that used to implement its tasks. These are the rule-based method, neural network, Bayesian networks, memory-based learning, support vector machines and decision tree [10]. Back to focus in this work topic and in order not to elaborate in the wide area, data mining and classification with the decision tree induction strategy are employed and adopted here in this work as a selected method to apply the title.

6. DECISION TREE

Decision tree DT is the most widely used technique to achieve supervised classification method. Decision tree inference encompasses of classification and learning. DT can be employed in any domain of discipline. It is built with a structure of a tree model. It is designed into smaller subsets incrementally getting deeper [11]. At the end, it would have leaf nodes and decision nodes as in Figure 4.
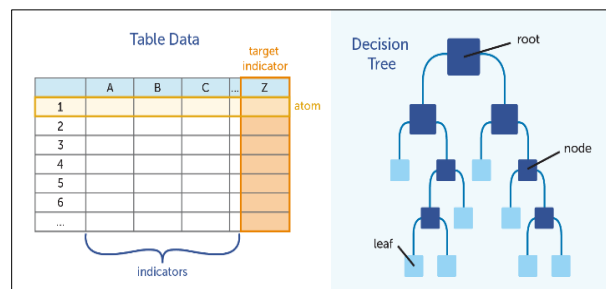

Figure 4: Decision Tree

There are some benefits of using decision tree:

- It is a visualized structure, so it understandable.
- It can be applied by numerical and categorical data when other techniques are used with only by one type.
- Its cost is measured by logarithmic in the value of data entry used with tree training.
- It works by logical output operation condition, because commonly there are two results, for instance, yes or no.

Besides to those benefits, it is worth to mention that in data mining there are two commonly used decision trees. First, **classification** tree, in this tree the data belongs to a class represents the predicted outcome which analysis for. An example of this, a student can be absent or attend. Second, **regression** tree, with this tree the analytic results is numeric only, for an example, student's scores. Both of them have their own characteristics [12]. Our work of research adopts the classification decision tree as it fits our needs.

So, how does it work? It splits and divides the population of data set into smaller sets. In figure 4, we have set of data as a table of columns. Decision tree depends on a class (normally as one column) and the other columns would be structured as a tree based on the specific algorithm.

6.1 Algorithm

Some algorithms are used with the decision tree, which are CART, C4.5, Random Forest and ID3. In fact, these algorithms are so wide in use and applications. They all have their own advantages and disadvantages regards to different criteria. In this section, we are not going to discuss them individually as much as mentioning their names at least. However, this paper adopts the ID3 algorithm to implement data mining classification technique [13]. There are some reasons have encouraged us to use ID3:

- The training data is used to create understandable prediction rules.
- It builds the fastest as well as a short tree.
- ID3 searches the whole dataset to create the whole tree.
- It finds the leaf nodes thus enabling the test data to be pruned and reducing the number of tests.
- The calculation time of ID3 is the linear function of the product of the characteristic number and node number.

6.2 ID3 Process

ID3 (Iterative Dicodetomiser 3) is one of decision tree algorithms. It was produced by Ross Quinlan [13] to build a decision tree by a given dataset. It works in a top-down manner suing greedy search via branches (dataset) without backward. As mentioned before, the decision tree converts a table of data into tree model via an algorithm [14]. Here, ID3 is our chosen tools to implement decision tree. So, how does ID3 work? To answer this question, below we list all its application processes.

- First and foremost, we determine the indicators (all the table columns) and the target indicator (one column).
- Finding the best attribute (column in the table). Selecting such attributes depends on two main metrics: **Entropy** and **Information Gain**.

- Splitting decision node to its distinct values and loop again till last nodes to have pure data set of the selected class.
- Stop if a leaf node has Entropy 0, which means all values are the same (class, indicator target).

Thus, it is very necessary to explain what Entropy and Information Gain are with an example of our trained data.

7. TRAINING DATA SET

Students' attendance database was used as an entry data set for training purpose with decision tree ID3 algorithm. This database has several tables, views, and relationships among them. In the previous section, we noted that this data set has been collected from three main resources. In table 1, we have seven tuples. All of them are used as an indicator attributes. The class one is not explicitly shown here, which is named *attended*. It has only two values (yes = attend, no = absent). Below is a brief description of these tuples.

- Gender: Its values are male or female. It comes from the student's data table.
- Subject hours: class duration, comes from the subject table it could be 1, 2 or 3 hours. This saved in the subject table.
- Study time: If it was in the morning or afternoon (evening) time to attend the college. This is from the student table.
- Employment: if a student was an employee or not or even works as part time or full time. It comes from the student's data table.
- Event: if in that day there could be the national event or local ceremony or nothing (yes = an event, no = tradition day). It comes from absences table.
- Weather: Climate changes, it takes rain or normal. This comes from absences table.
- Exam: refers to whether there is a test or not in that particular day. This registered in absences table. It takes two values (yes = test day, no = no test).

The database tables have many data about students, subject and absence status. However, we only use those attributes because they have direct impact of future attendance status and influencing ID3 algorithm into right decision.

Table 1. Sample of used data set

Gender	Sub. hrs.	Study time	Emp	Event	Weather	exam
male	2	morning	none	no	rain	yes
male	2	morning	Full	no	normal	no
female	1	morning	part	yes	normal	yes
male	1	afternoon	Full	no	normal	no
female	2	morning	part	yes	rain	yes
female	2	afternoon	none	no	rain	no
male	1	afternoon	none	no	normal	no
female	2	morning	none	no	rain	no
female	1	morning	none	yes	normal	no
male	2	afternoon	part	yes	normal	yes
male	2	afternoon	part	no	normal	yes
female	2	morning	part	no	rain	no
female	2	morning	full	yes	rain	yes
male	1	morning	none	yes	normal	no

Table 1 shows a sample of data entry. In fact this comes from a view called '*vw_dt*' in the database.

8. IMPLEMENTING ID3 ALGORITHM

To begin executing our chosen algorithm, we need to follow its procedures as stated in the previous section, 6.2. First, we ensure that we have to have a data entry set like in table 1. Now, the most significant step that this algorithm depends on, is to find the best attribute among the seven ones. Selecting the best attribute will take us for best decision tree model so that finally we will have the right decision through the tree which we have built via this algorithm. In 6.2 section, we talked about Entropy and Information Gain. These two metrics are so crucial to find the best attributes [15].

8.1 Entropy (E)

When we choose a root node (decision node) as considered best-selected attribute, it will have different values related to the class (target indicator), so it would have some YESs and couple NOs, for example. Entropy is defined as a measurement to identify the homogeneity and uncertainty of these values/sample. The Entropy has a value between 0 and 1 only. If the sample is completely same, then E is 0, if it is equal, then it is 1 [16]. Thus, E increases with an increase in randomness, and it decreases with decreasing in randomness. The mathematic equation for E like:

$$E(S) = \sum -p_i \log_2 p_i$$

An example of this function in figure 5 shows E to the *employment* attributes, Later in table 2, we will have the results for all data sets.
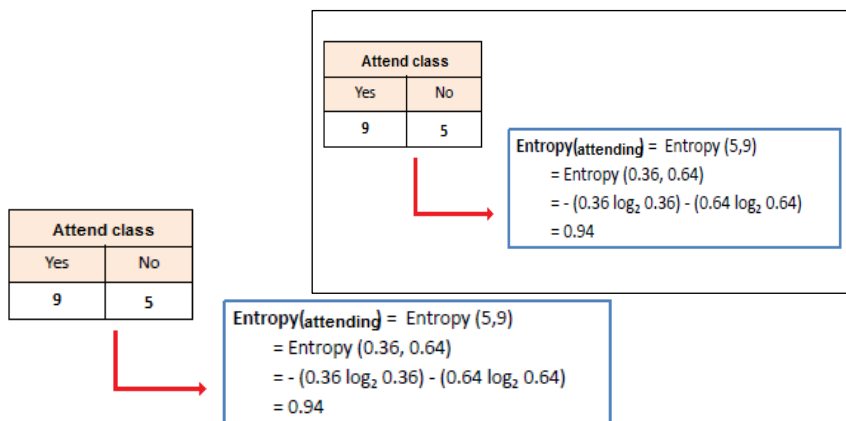

Figure 5: Example of Entropy on attendance values

8.2 Information Gain

The Second major factor is called Information Gain IG. IG is other mathematic operation needed to be calculated in order to determine what the selected attribute is.

$$Gain(T, X) = Entropy(t) - Entropy(T,X)$$

Where T denotes for target attribute and X is to a sub-set that was divided under an attribute such as *employment*. An example of this in our empirical work as following:

$$G(_{attending, Emp.}) = E(_{Attending}) - E(_{Attending, Emp.})$$
$$= 0.940 - 0.693 = 0.247$$

In fact, IG represents the decrease of E after splitting dataset into some attributes. Creating the decision tree needs to find an attribute that has maximum IG [17]. Thus we can select the best attribute which is *employment*. Table 1 illustrates the values of all Es and IGs about all the nodes (attributes and their sub-nodes). After calculating the Entropy and Information Gain for all the attributes, we can find that (attr, Employment) is the best choice for our data entry. Table 1 shows a brief of this conducted scenario.

Table 2. E and IG of all attributes

Attributes = Employment				
Sub-set	full	part	non	Emp.	IG
E	0.45	0.76	0.82	0.94	0.247
Attributes = Gender				
Sub-set	Male	Female	Gender		IG
E	0.93	0.77	0.93		0.071
Attributes = Weather				
Sub-set	rain	normal	weather		IG
E	0.56	0.81	0.88		0.213
Attributes = Exam				
Sub-set	yes	no	exam		IG
E	0.32	0.57	0.81		0.17
Attributes = Event				
Sub-set	yes	no	event		IG
E	0.67	0.9	.38		0.196
Attributes = Study Time				
Sub-set	morning	afternoon	Study time		IG
E	0.78	0.96	.58		0.093

From the above table, we can tell that *employment* attribute has the highest value. So, it was taken as root node. All other nodes were processed in the same scenario till reach pure node of certainty. As depicts in figure 6, the Emp. attribute has fallen into three values. Each value leads to a group of data with other tuples. We can see red rows indicates NO values in target class (attending column) and black ones for YES. These sample tables represent a node in the tree. ID3 algorithm loops via all possible nodes and applies same procedure in splitting and calculating E and IG each time for every single node till reach pure node with certain decision.
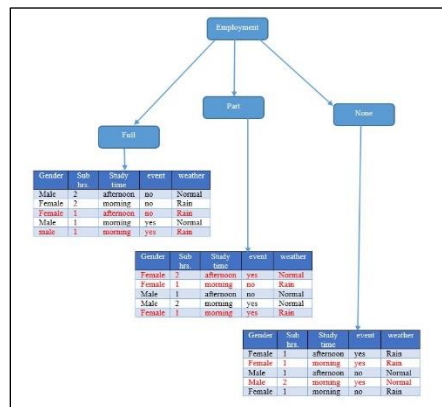

Figure 6: Root node with its split values

The final tree is illustrated in figure 7. This tree has the leaf nodes which carry the predicted decision we are looking for.
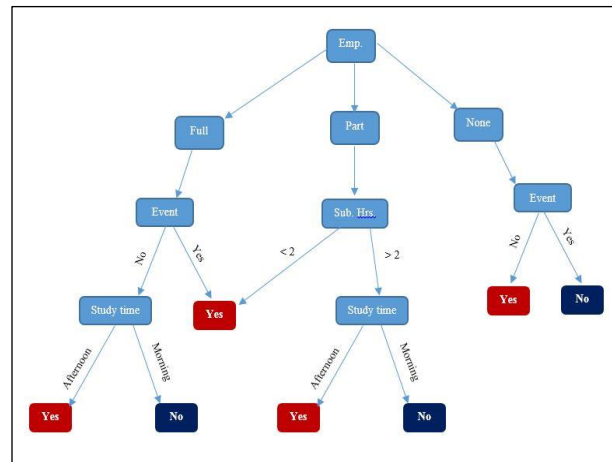

Figure 7: The Final Tree

9. DECISION RULES

The decision rules are inferred from the decision tree. We can reckon it as following, first three rules for not coming status.

- If a student (not the employee, an event happens) => they are absent.
- If a student (employee partially, class hours larger than 2 hours, the study in the morning) => they are not coming.
- If a student (employee full time, no event is there, the study in the morning) => they are not coming.

Now for coming students can be under these situations:

- If a student (not the employee, no event there) => they are coming.
- If a student (employee partially, class hours larger than 2 hours, afternoon study) => they are attending.
- If a student (employee partially, class hours less than 2 hours) => they are attending.
- If a student (employee full time, no event is there, the study in the afternoon) => they are coming.

So far we have talked and gone through the decision tree and its algorithm ID3. Eventually, we hit the tree construction and its rules for predicted information. The next step is to use a forecasting strategy to portray attendance representation.

10. VISUAL ANALYTIC RESULT

We get an advantage of the major smooth integration between Excel and Access database. Massive resources talked about how to manipulate data between these programs as in [18]. In this work, we use excel to build visualized reports about the overall trends' behavior of students' absences that have already registered in the Access database. This topic addresses as possible how to employee information prediction from various perspective. In previous, we discussed how to apply ID3 algorithm as a decision tree model to predict students' absences status.

Now, we utilize analytic reports that visualize the actual fact of students' attendance. Before starting to talk about these visual statistic result as in figures 8, 9 and 10. It is important to clarify

how these trends were drawn and what their basic data is. Two universities have been chosen to exert such work, Sumer University and Thi-Qar University. Sumer depends on semester system, however, Thi-Qar works on annual approach. Both universities use this project with the two systems. After applying a suitable forecasting technique, we can easily find the right behaviour for present and future

10.1 Semester Terms

A college in the University of Sumer works in semesters approach in its taught program of education. In these two periods of time, first semester and second semester, a trend of attendances has been shown to illustrate the overall fact of students' absences, figure 8. The first term begins in September till the middle of January. Figure 8 depicts the ratio of students' attendance of particular year, 2017-2018, in our sample. In fact, the previous years are closely near to the same trend view with 2017- 2018.
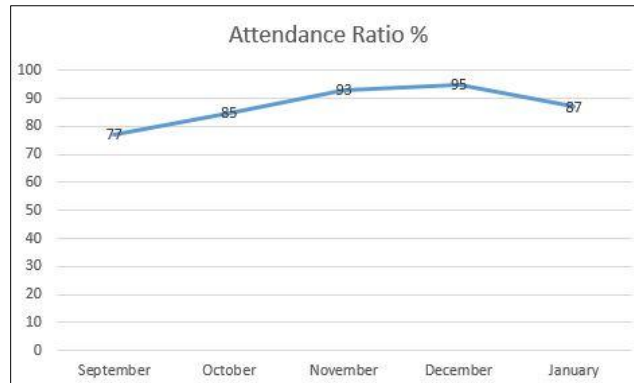

Figure 8: Attendance Trend for Five months Sep. to Jan. for 2017-2018

In figure 8, we can see a noticeable big difference from the trend in figure 9. A smooth change from 77% to 87% during five months with a little rise in November and December.
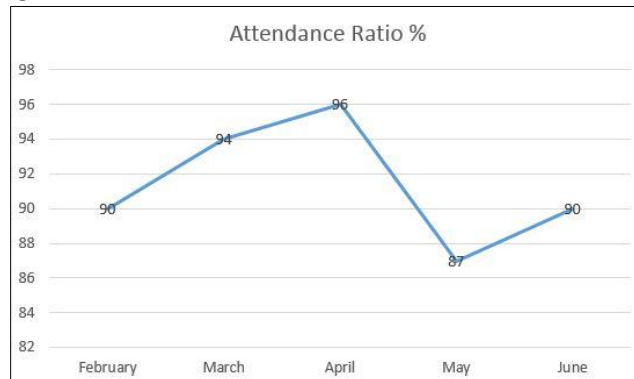

Figure 9: Attendance Trend for Five months Feb. to Jun. for 2017-2018.

In figure 9, there is a considerable ascent from February to April. Notwithstanding, from April to May there is an incredible decline. From May to June, gradual increasing has risen till we have 90% since 87%. These visual upshots tell us the past and current facts. But, when we need to have a discovering information, then we should adopt another solution like forecasting.

11. FORECASTING FEATURE AND FACTS

As previously discussed with the visualized reports which show the overview of attendances during the particular time. There are many forecasting algorithms and math approaches are applied to predict and estimate values based on historical data.

11.1 Forecast.ETS function Inner Calculation

This function has some parameters used to result from feature values. Before we start working with its function, let us take a glance at the general FORECAST function. The basic form of FORECAST function as following:

$$FORECAST(x, known\_y's, known\_x's)$$

These arguments are simply like:

- X: Is the data for that we predict a value.
- Known_y's: Is the dependent list of data.
- Known_x's: Is an independent list of data.

The actual math equation to implement this function is:

$$a + b\ x$$

Where a = $\overline{y}$ + b $\overline{x}$

And   b = $\frac{\sum(x-\overline{x})(y-\overline{y})}{\sum(x-\overline{x})^2}$

A simple example for this functionality is shown in table 2.

Table 2. Example of Forecast.ETS data

Known Y	Known X
7	21
8	29
10	32
16	39
22	41

=FORECAST (30, [Y range], [X range]) = 10.89814

Now, after working with the basic form of FORECAST function, it is time to practice our mission of FORECAST.ETS function. This function calculates the predicted values based on historical data [19]. The algorithm applied in this function is the AAA version of Exponential Triple Smoothing (ETS). The predicted result or values is consequences of historical values for the particular date. Also, this function depends on a timeline to be utilized along with fixed steps between several points. It could be a yearly timeline, the 1st month for a year or even numeric references timeline. FORECAST.ETS form is:

FORECAST.ETS(target_date, values, timeline, [seasonality], [data_completion], [aggregation])

So, we have some parameters that such function uses. These arguments are:

- Target_date: data key that we want to predict.
- Values: the next forecasted points depends on these values which is historical.
- Timeline: numeric or dates values.

- Seasonality: It takes a value 0 - 8,760 it represents the number of date point which indicates the number of hours.
- Data completion: the function sets these params to automatically adjust up to 30% missing data to 0.
- Aggregation: values that have the same time stamp, will be aggregated with average if it is default with 0 left value.

Table 3 shows sample data to apply FORECAST.ETS algorithm on data like what was written. The orange cells are required to be predicted from 2021 to 2025 based on the historical data from 2010 to 2020. Figure 10 shows the extra capability of FORECAST.ETS algorithm of forecasting Lower Confidence Bound and Upper Confidence Bound.

Table 3. Historical data of Students' attendance

ear	actual
2010	90
2011	88
2012	91
2013	89
2014	92
2015	87
2016	92
2017	86
2018	90
2019	91
2020	89
2021	
2022	
2023	
2024	
2025	

We can see in figure 10 the three coloured trends, yellow one is for Upper Confidence Bound, orange trend is the normal predicted values and grey trend is for Lower Confidence Bound.
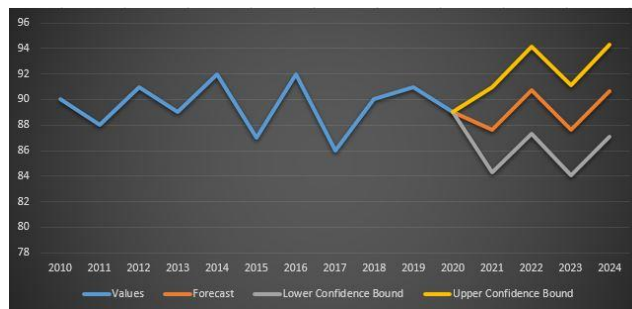

Figure 10: FORECAST.ETS Graphical values form 2010 till 2014.

Figure 11 shows the estimates values with Lower and Upper Confidant Pound. The years 2021 to 2024 have values comply with past years from 2010 till 2020.

83

Timeline	Values	Forecast	Lower Confidence Bound	Upper Confidence Bound
2010	90			
2011	88			
2012	91			
2013	89			
2014	92			
2015	87			
2016	92			
2017	86			
2018	90			
2019	91			
2020	89	89	89.00	89.00
2021		87.649143	84.32	90.97
2022		90.714665	87.28	94.14
2023		87.619909	84.09	91.15
2024		90.685432	87.05	94.32

Figure 11: Expected Values with FORECAST.ETS

12. CHALLENGES AND SOLUTIONS

Any developed software is possible to have some insufficient consequences or shortage. In fact, the obstacle we face with this work, is the disadvantages and short comes of ID3 itself. The most impediment situation is that ID3 applied only for one student each time we need to predict their future status, because it is executed on a student records data set. So, applying it for large number of students individually, is time wasting and computer resources consuming. Therefore, two remedies are suggested here.

- Performing DT task when it is required, but not all the time. That means the implementation occurs for a student or few number of students.
- Instead of creating data set (table of records) of each student and use ID3 with it, we shall create a data set of group of student. This is done by emulate group of student as one student by registering high number of whole class or group as an absent case, and low number of absence is attending. Thus, we could correspond group of students as one instance.

In future development and requirement expanding, some advice and recommendations are proposed, too.

13. CONCLUSION AND FUTURE WORK

At the end steps of this work and to recap the main outline points, it has been proven that heuristic applications and intelligent systems are needed and imperative to supply them into our business. In the empirical scenario, the database is used in addition to its basic job within data mining to advance such system qualifications and abilities. Among several data mining techniques, we have worked with classification method as it fits our goals.

We applied decision tree methodology as one of classification manner to implement prediction and future discovering matter. The best algorithm to construct a decision tree was the ID3 and we successfully built DT on the base of mathematical operations as E and IG which are ID3 uses. A sample of data set matches a student's records in the database was utilized as an instance case to make DT. After processing such data and calculating ID3 parameters, we conducted the employment attributes is the major key that student absence case depends on. Consequently, we have made decision rules which are considered the inferred outcomes and the benefit of whole work that this paper motives to.

Not only this, but we also launched again to use visualized upshots and graphical views of students' absences over a span of months. In addition to these charts information, we have built

forecasting and estimating values and future expectation basis on historical events. The drawn trends have told us the overall scenarios and absences conductions.

In the future, we recommend using the enhanced ID3 algorithm to bypass some problems that we possibly encounter. The developed ID3 is called C4.5. However, C4.5 algorithm is not the optimal solution for ID3 as each has its own pros and cons. We can achieve the optimal use of ID3 by avoiding its disadvantages and apply it enhanced one the C4.5 algorithm as well as follow the guidance in section 12. We also advise to practice this paperwork in a way to match and measure how far the forecasting methods and DT rules are compiled and conformable with each one results.

REFERENCES

[1] Jiawei H., Data Mining: Concepts and Techniques, 3rd ed.,Morgan Kaufmann Publishers, USA: 2012: pp 4-10.

[2] Kalpesh A., Aditya G., Amiraj D., Rohit J., Vipul H., "PREDICTING STUDENTS' PERFORMANCE USING ID3 AND C4.5 CLASSIFICATION ALGORITHMS", International Journal of Data Mining & Knowledge Management Process (IJDKP), Vol.3, No.5, September

[3] Wang X.,Wang L.,Li N., "An Application of Decision Tree Based on ID3", 2012 International Conference on Solid State Devices and Materials Science, Physics Procedia 25 (2012) 1017 – 1021

[4] Yuesheng T., Zhansheng Q., Jingyu W., "Applications of ID3 algorithms in computer crime forensics ", 2011 International Conference on Multimedia Technology, 26-28 July 2011.

[5] Ali D., Waffa A, Ali A, "Prioritizing Software Capabilities and Focal Points of MS Access and Excel in Perspective of Data Management", *Applied Computer Science, vol. 14, no. 3, pp. 15–30.*

[6] Mohammed J., Wagner M., Data Mining and Analysis: Fundamental Concepts and Algorithms, 1st ed., Cambridge University Press, NY USA, 2014: pp 3-30.

[7] Kantardzic M., Data Mining: Concepts, Models, Methods, and Algorithms, 2nd ed., IEEE press, August 2018: pp 2-24.

[8] Bhumika G., et al "Analysis of Various Decision Tree Algorithms for Classification in Data Mining", *International Journal of Computer Applications*, 2017: Vol. 163, No 8.

[9] Rahman N., "Data Mining Problems Classification and Techniques", *International Journal of Big Data and Analytics in Healthcare*, January-June 2018: Vol. 3 No. 1, p 37-58,

[10] Brown L., Mues C., "An experimental comparison of classification algorithms for imbalanced credit scoring data sets", Expert Systems with Applications: February 2012, Volume 39, Issue 3, 15, Pages 3446-3453.

[11] Rokach L., Maimon O., "data mining with decision trees: Theory and Applications", 2014: 2nd. ed. World Scientific, London UK.

[12] Dai Q., et al "Research of Decision Tree Classification Algorithm in Data Mining", SERSC: IJDTA, 2016: Vol.9, No.5 pp.1-8.

[13] J. Ross Quinlan, "C4.5: Programs for Machine Learning" 1993:6th ed.,morgan kaufmann publishers

[14] Rupali Bhardwaj , Sonia Vatta "Implementation of ID3 Algorithm", International Journal of Advanced Research in Computer Science and Software Engineering. Vol 3, Issue 6, June 2013

[15] B. Hitarthi, M. Shraddha, R. Lynette. "Use of ID3 Decision Tree Algorithm for Placement Prediction", International Journal of Computer Science and Information Technologies. Vol. 6 (5), 2015

[16] Mathur, n.. et al "the base strategy for id3 algorithm of data mining using havrda and charvat entropy based on decision tree", international journal of information and electronics engineering, , march 2012: vol. 2, no. 2.

[17] Adhatrao k., et al "predicting students' performance using id3 and c4.5 classification algorithms", international journal of data mining & knowledge management process (ijdkp). september 2013: vol.3, no.5.

[18] Christopher C. Brown, Denise Pan, and Gabrielle Wiersma, "Advanced Data Analysis: From Excel PivotTables to Microsoft Access" (2014). Proceedings of the Charleston Library Conference. p 571.

[19] Rob J., Khandakar Y. "Automatic Time Series Forecasting: The forecast Package for R", Journal of Statistical Software July 2008, Vol. 27, No. 3.