International Journal of Innovations in Applied Sciences &Engineering http://www.ijiase.com

(IJIASE) 2018, Vol. No. 4, Jan-Dec **e-ISSN: 2454-9258, p-ISSN: 2454-809X**

INFORMATION MINING: EMPLOYABILITY OF ID3 & C4.5 ALGORITHMS IN ENHANCING CLASSIFICATION CRITERIAS

Karan Sablok

Student, Delhi Public School

ABSTRACT

An instructive foundation needs an inexact earlier learning of selected understudies to foresee their execution in future scholastics. This encourages them to distinguish promising understudies and furthermore gives them a chance to focus on and enhance the individuals who might most likely get bring down evaluations. As an answer, we have built up a framework which can anticipate the execution of understudies from their past exhibitions utilizing ideas of information mining systems under Classification. We have dissected the informational index containing data about understudies, for example, sexual orientation, marks scored in the board examinations of classes X and XII, stamps and rank in placement tests and results in first year of the past group of understudies. By applying the ID3 (Iterative Dichotomiser 3) and C4.5 grouping calculations on this information, we have anticipated the general and individual execution of newly conceded understudies in future examinations.

1. INTRODUCTION

Consistently, instructive organizations concede understudies under different courses from various areas, instructive foundation and with changing legitimacy scores in placement tests. In addition, schools and junior universities might be subsidiary to various sheets, each board having distinctive subjects in their educational program and furthermore unique level of profundities in their subjects. Investigating the past execution of conceded understudies would give a superior viewpoint of the likely scholastic execution of understudies later on. This can possibly be accomplished utilizing the ideas of information mining.

For this reason, we have investigated the information of understudies selected in first year of designing. This information was acquired from the data given by the conceded understudies to the establishment. It incorporates their full name, sexual orientation, application ID, scores in board examinations of classes X and XII, scores in placement tests, classification and affirmation compose. We at that point connected the ID3 and C4.5 calculations subsequent to pruning the dataset to anticipate the aftereffects of these understudies in their first semester as exactly as could be allowed.

2. LITERATURESURVEY

2.1 Data Mining

Information mining is the way toward finding fascinating learning, for example, affiliations, designs, changes, critical structures and peculiarities, from a lot of information put away in databases or information stockrooms or other data archives [1]. It has been broadly utilized as of late because of the accessibility of colossal measures of information in electronic shape, and there is a requirement for transforming such information into helpful data and learning for huge applications. These applications are found in fields, for example, Artificial Intelligence, Machine Learning, Market Analysis, Statistics and Database Systems, Business Management and Decision Support [2].

2.1.1 Classification

Arrangement is an information mining procedure that maps information into predefined gatherings or classes. It is a directed learning strategy which requires marked preparing information to produce rules for arranging test information into foreordained gatherings or classes [2]. It is a two-stage process. The principal stage is the learning stage, where the preparation information is examined and characterization rules are created. The following stage is, where test information is ordered into classes as per the produced tenets. Since characterization calculations necessitate that classes be characterized dependent on information property estimations, we had made a trait "class" for each understudy, which can have an estimation of either "Pass" or "Come up short".

2.1.2 Clustering

Bunching is the way toward gathering an arrangement of components so that the components in a similar gathering or bunch are more like each other than to those in different gatherings or groups [1]. It is a typical method for measurable information investigation utilized in the fields of example acknowledgment, data recovery, bioinformatics, machine learning and picture examination. Grouping can be accomplished by different calculations that vary about the similitudes required between components of a bunch and how to discover the components of the groups proficiently. Most calculations utilized for bunching attempt to make groups with little separations among the group components, interims, thick zones of the information space or specific measurable conveyances.

2.2 Selecting Classification over Clustering

In grouping, classes are obscure Apriori and are found from the information. Since we will probably anticipate understudies' execution into both of the predefined classes - "Pass" and "Fizzle", bunching is certifiably not an appropriate decision thus we have utilized arrangement calculations as opposed to grouping calculations.

2.3 Issues Regarding Classification

2.3.1 Missing Data

Missing information esteems cause issues amid both the preparation stage and to the characterization procedure itself. For instance, the explanation behind non-accessibility of information might be expected to [2]:

2.3.1.1 Hardware glitch

2.3.1.2 Cancellation because of irregularity with other recorded information

2.3.1.3 Non-entry of data due tomisunderstanding

2.3.1.4 Certain data considered unimportant at the time ofentry

2.3.1.5 No registration of data or itschange

This missing data can be handled using following approaches [3]:

2.3.1.6 Data miners can ignore the missingdata

2.3.1.7 Data miners can replace all missing values with a single globalconstant

2.3.1.8 Data miners can replace a missing value with its feature mean for the givenclass

2.3.1.9 Data miners and domain experts, together, can manually examine samples with missing values and enter a reasonable, probable or expectedvalue

For our situation, the odds of getting missing qualities in the preparation information are less. The preparation information is to be recovered from the affirmation records of a specific organization and the qualities considered for the contribution of order process are obligatory for every understudy. The tuple which is found to have missing an incentive for any characteristic will be overlooked from preparing set as the missing qualities can't be anticipated or set to some default esteem. Thinking about low odds of the event of missing information, disregarding missing information won't influence the exactness antagonistically.

2.3.2 Measuring Accuracy

Figuring out which information mining strategy is best relies upon the elucidation of the issue by clients. As a rule, the execution of calculations is analyzed by assessing the precision of the outcome. Characterization precision is ascertained by deciding the level of tuples put in the right class. In the meantime there might be an expense related with a mistaken task to the wrong class which can be overlooked.

2.4 ID3Algorithm

In choice tree learning, ID3 (Iterative Dichotomiser 3) is a calculation concocted by Ross Quinlan used to produce a choice tree from the dataset. ID3 is regularly utilized in the machine learning and normal dialect preparing spaces. The choice tree procedure includes building a tree to show the arrangement procedure. Once a tree is manufactured, it is connected to each tuple in the database and results in arrangement for that tuple. The accompanying issues are looked by most choice tree calculations [2]:

- Choosing part properties
- Ordering of part properties
- Number of parts to take
- Balance of tree structure and pruning
- Stopping criteria

The ID3 calculation is a grouping calculation dependent on Information Entropy, its essential thought is that all models are mapped to various classifications as indicated by various estimations of the condition property set; its center is to decide the best order characteristic shape condition quality sets. The calculation picks data gain as characteristic determination criteria; ordinarily the quality that has the most elevated data gain is chosen as the part property of current hub, with the end goal to make data entropy that the partitioned subsets require littlest [4]. As per the diverse estimations of the quality, branches can be built up, and the procedure above is recursively approached each branch to make different hubs and branches until the point that every one of the examples in a branch have a place with a similar class. To choose the part characteristics, the ideas of Entropy and Information Gain are utilized.

2.4.1 Entropy

Given probabilities $p_1, p_2, \ldots, p_s$, where $\sum p_i = 1$, Entropy is defined as

$$H(p_1, p_2, \ldots, p_s) = \sum - (p_i \log p_i)$$

Entropy finds the measure of request in a given database state. An estimation of H = 0 distinguishes a splendidly arranged set. At the end of the day, the higher the entropy, the higher the possibility to enhance the characterization procedure.

2.4.2 Information Gain

ID3 picks the part characteristic with the most astounding increase in data, where gain is characterized as contrast between how much data is required after the part. This is figured by deciding the contrasts between the entropies of the first dataset and the weighted total of the entropies from each of the subdivided datasets. The equation utilized for this intention is:

$$G(D, S) = H(D) - \sum P(D_i)H(D_i)$$

2.5. C4.5

C4.5 is an outstanding calculation used to create a choice trees. It is an expansion of the ID3 calculation used to beat its detriments. The choice trees created by the C4.5 calculation can be utilized for order, and thus, C4.5 is likewise alluded to as a measurable classifier. The C4.5 calculation rolled out various improvements to enhance ID3 calculation [2]. A portion of these are:

- Handling preparing information with missing estimations of qualities
- Handling contrasting cost properties
- Pruning the choice tree after its creation
- Handling qualities with discrete and constant qualities

Give the preparation information a chance to be a set S = s1, s2 ... of officially ordered examples. Each example Si = xl, x2... is where xl, x2 ... speak to traits or highlights of the example. The preparation information is a vector C = c1, c2..., where c1, c2... speak to the class to which each example has a place with.

At every hub of the tree, C4.5 picks one trait of the information that most adequately parts informational collection of tests S into subsets that can be one class or the other [5]. It is the standardized data gain (distinction in entropy) that outcomes from picking a trait for part the information. The quality factor with the most elevated standardized data gain is considered to settle on the choice. The C4.5 calculation at that point proceeds on the littler sub-records having next most noteworthy standardized data gain.

3. TECHNOLOGIESUSED

3.1 HTML and CSS

HyperText Markup Language (HTML) is a markup dialect for making pages or other data to show in an internet browser. HTML enables pictures and protests be incorporated and that can be utilized to make intelligent structures. From this, organized records are made by utilizing basic semantics for content, for example, headings, joins, records, passages, cites and so forth.

CSS (Cascading Style Sheets) is intended to empower the division between archive content (in HTML or comparative markup dialects) and report introduction. This method is utilized to enhance content availability additionally to give greater adaptability and control in the particular of substance and introduction attributes. This empowers various pages to share organizing and lessen redundancies.

3.2 PHP and the CodeIgniter Framework

PHP (recursive acronym for PHP: Hypertext Preprocessor) is a broadly utilized open source universally useful server side scripting dialect that is particularly suited for web improvement and can be installed into HTML.

CodeIgniter is an outstanding open source web application system utilized for building dynamic web applications in PHP [6]. Its will probably empower engineers to create extends rapidly by giving a rich arrangement of libraries and functionalities for regularly utilized assignments with a basic interface and legitimate structure for getting to these libraries. CodeIgniter is inexactly founded on the Model-View-Controller (MVC) example and we have utilized it to manufacture the front end of our execution.

3.3 MySQL

MySQL is the most famous open source RDBMS which is upheld, dispersed and created by Oracle [8]. In the execution of our web application, we have utilized it to store client data and understudies' information.

3.4 RapidMiner

RapidMiner is an open source information mining instrument that gives information mining and machine learning strategies including information stacking and change, information preprocessing and representation, displaying, assessment, and arrangement [7]. It is composed in the Java programming dialect and makes utilization of taking in plans and trait evaluators from the WEKA machine learning condition and measurable displaying plans for the R-Project. We have utilized RapidMiner to produce choice trees of ID3 and C4.5 calculations.

4. IMPLEMENTATION

We had separated the whole usage into five phases. In the main stage, data about understudies who have been admitted to the second year was gathered. This incorporated the points of interest submitted to the school at the season of enrolment. In the second stage, superfluous data was expelled from the gathered information and the pertinent data was bolstered into a database. The third stage included applying the ID3 and C4.5 calculations on the preparation information to acquire choice trees of both the calculations. In the following stage, the test information, i.e. data about understudies as of now enlisted in the main year, was connected to the choice trees. The last stage comprised of building up the front end as a web application.

These stages of implementation are depicted in Figure 1.

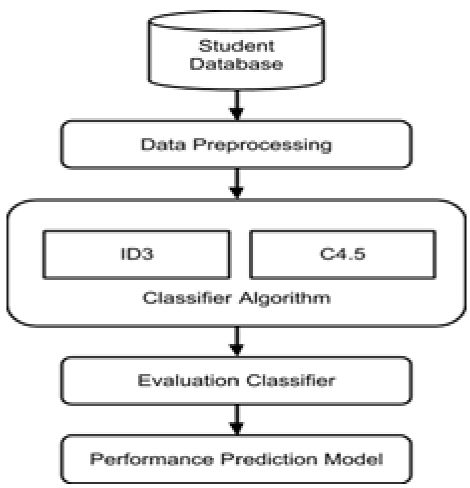


Figure 1. Processing model

4.1 Student Database

We were furnished with a preparation dataset comprising of data about understudies admitted to the principal year. This information was as a Microsoft Excel 2003 spreadsheet and had points of interest of every understudy, for example, full name, application ID, sexual orientation, rank, level of imprints got in board examinations of classes X and XII, level of imprints got in Physics, Chemistry and Mathematics in class XII, marks got in the placement test, confirmation compose, and so on. For simplicity of performing information mining activities, the information was filled into a MySQL database.

4.2 Data Preprocessing

When we had subtle elements of the considerable number of understudies, we at that point fragmented the preparation dataset further, considering different attainable part characteristics, i.e. the qualities which would higherly affect the execution of an understudy. For example, we had considered 'area' as a part characteristic, and after that portioned the information as indicated by understudies' region.

A preview of the understudy database is appeared in Figure 2. Here, unimportant properties, for example, understudies private location, name, application ID, and so forth had been evacuated. For instance, the confirmation date of the understudy was immaterial in foreseeing the future execution of the understudy. The characteristics that had been held are those for legitimacy score or stamps scored in placement test, sexual orientation, level of imprints scored in Physics, Chemistry and Mathematics in the board examination of class XII and confirmation compose. At last, the "class" characteristic was included and it held the anticipated outcome, which can be either "Pass" or "Come up short".

Since the properties for imprints would have discrete qualities, to create better outcomes, particular classes were characterized. Consequently, the "justify" characteristic had an esteem "decent" if the legitimacy score of the understudy was 120 or above out of a most extreme score of 200, and was delegated "terrible" if the legitimacy score was underneath 120. Additionally, the esteem that can be held by the "rate" property of the understudy are three - "qualification" if the level of imprints scored by the understudy in the subjects of Physics, Chemistry and Mathematics was 70 or above, "first_class" if the rate was under 70 and more noteworthy than or equivalent to 60, at that point it was delegated "second_class" if the rate was under 60. The characteristic for confirmation compose is named "type" and the esteem held by an understudy for it tends to be either "AI" (short for All-India), if the understudy was admitted to a seat accessible for All-India competitors, or "OTHER" if the understudy was admitted to another seat.

sr_no	merit_no	merit_marks	app_id	name	gender	cast	location	percent	type
1	328	153.00	EN10205034	AKSHAY DEBNATH	Male	Open	Mumbai	95.66	AI
2	725	152.00	EN10279070	YEMPALLE SUSHMA BASWARAJ	Female	Open	Mumbai	86.66	AI
3	1066	143.00	EN10288911	KIRAN SUSHIL GRIFFITHS	Male	Open	Mumbai	96.00	AI
4	1294	136.00	EN10167854	WALCHALE ABHIJEET SUHAS	Male	Open	Mumbai	82.00	AI
5	1419	132.00	EN10255786	KUNAL JADHAV	Male	Open	Mumbai	80.33	AI
6	21566	109.00	EN10230782	KARKHELE RAVINDRAKUMAR VITTHAL	Male	NT 3 (NT-D)	Mumbai	83.66	GNT3H
7	3290	156.00	EN10172564	TALAWADEKAR ADITYA SHYAM	Male	OBC	Mumbai	89.33	GOBCH
8	5933	144.00	EN10264877	SONAWANE NIKHIL RAJENDRA	Male	SBC/OBC	Mumbai	89.66	GOBCH
9	6882	140.00	EN10196064	PATIL SUMEET BHAGWAN	Male	OBC	Mumbai	88.33	GOBCH
11	1456	168.00	EN10195904	LOHOTE PRANIT TANAJI	Male	Open	Mumbai	92.00	GOPENH
12	2158	162.00	EN10216545	IYER SIDDHARTH SUNDARAM	Male	Open	Mumbai	93.66	GOPENH
13	2519	160.00	EN10255191	GEORGE NISHANT JOSEPH	Male	Open	Mumbai	94.66	GOPENH

merit	gender	percent	type	class
good	Male	distinction	AI	pass
good	Female	distinction	AI	pass
good	Male	distinction	AI	pass
good	Male	distinction	AI	pass
good	Male	distinction	AI	pass
bad	Male	distinction	OTHER	pass
good	Male	distinction	OTHER	pass
good	Male	distinction	OTHER	pass
good	Male	distinction	OTHER	fail
good	Male	distinction	OTHER	pass
good	Male	distinction	OTHER	pass

Figure 2. Preprocessed student database

4.3 Data Processing Using Rapid Miner

The following stage was to nourish the pruned understudy database as contribution to RapidMiner. This helped us in assessing fascinating outcomes by applying order calculations on the understudy preparing dataset. The outcomes got are appeared in the accompanying subsections:

4.3.1. ID3 Algorithm

Since ID3 is a choice tree calculation, we acquired a choice tree as the last outcome with all the part traits and it is appeared in Figure 3.

4.3.2. C4.5 Algorithm

The C4.5 calculation too creates a choice tree, and we acquired one from RapidMiner similarly as ID3. This tree, appeared in Figure 4, has less choice hubs when contrasted with the tree for enhanced ID3, or, in other words Figure 3.

4.4. Implementing the Performance Prediction Web Application

RapidMiner helped altogether in finding concealed data from the preparation dataset. These recently learnt prescient examples for anticipating understudies' execution were then actualized in a working web application for staff individuals to use to get the anticipated consequences of conceded understudies.

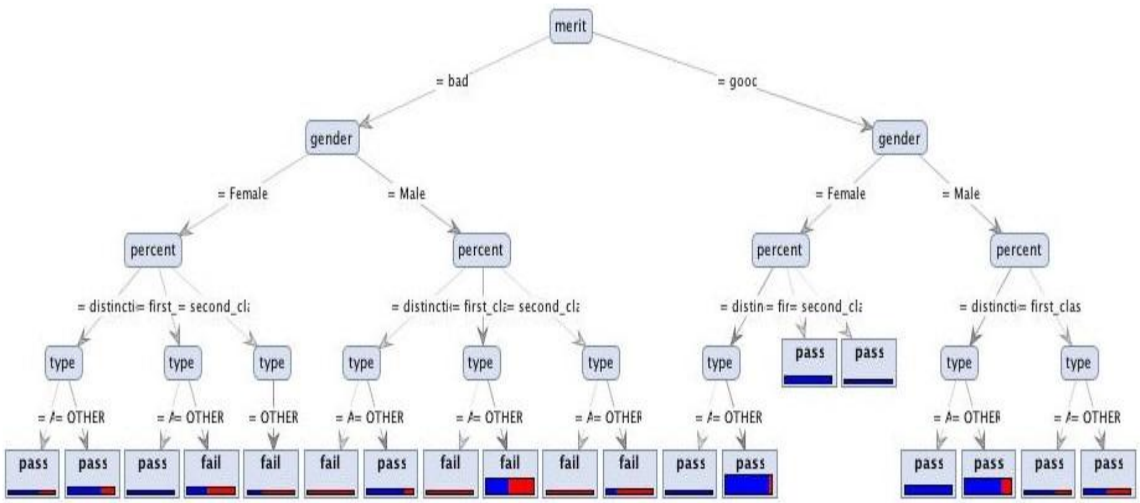


Figure 3. Decision tree forID3

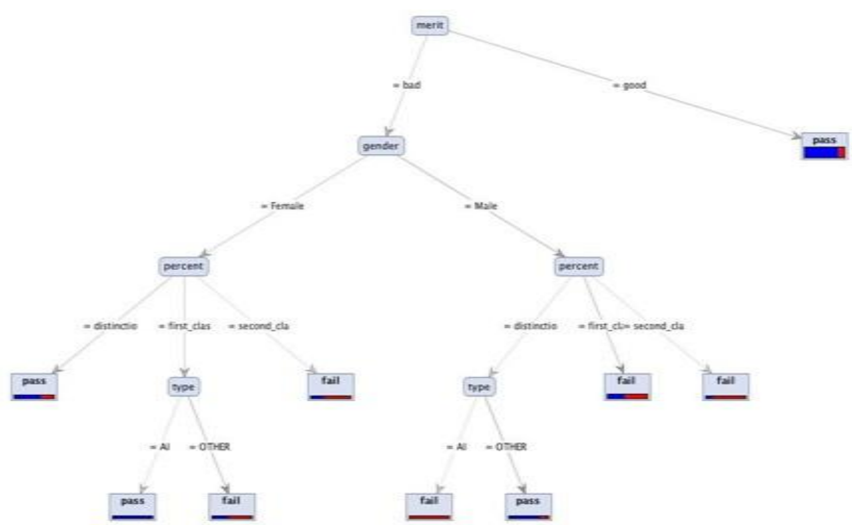


Figure 4. Decision tree forC4.5

4.4.1. CodeIgniter

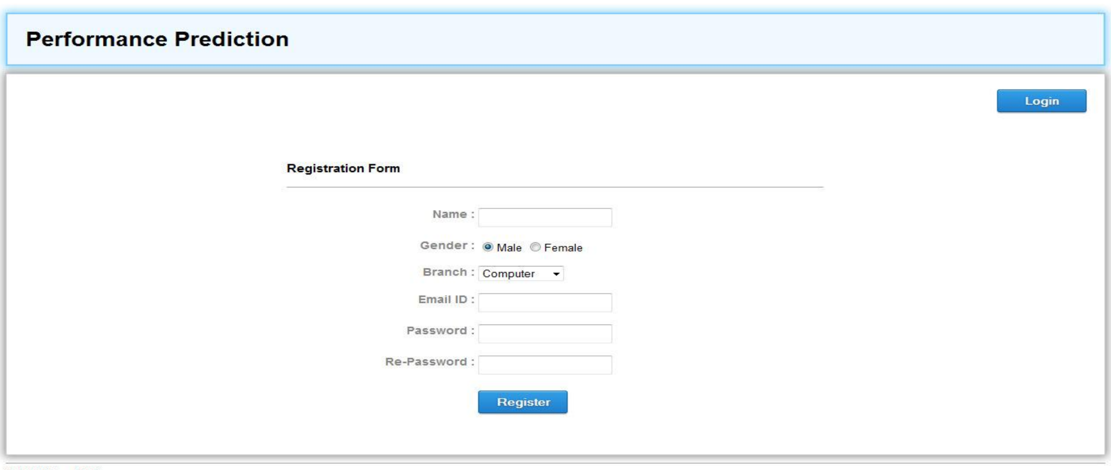
4.4.2. The web application was produced utilizing a well known PHP structure named CodeIgniter. The application has arrangements for different synchronous staff enlistments and staff logins. This guarantees crafted by no two staff individuals is intruded on amid execution assessment. Figure 5 and Figure 6 portray the staff enrollment and staff login pages separately.

4.4.3.

4.4.4. 4.4.2. Mapping Decision Trees to PHP

4.4.5.

4.4.6. The substance of the web application was to delineate outcomes accomplished after information handling to code. This was done in type of class strategies in PHP. The consequence of the enhanced ID3 and C4.5 calculations were as trees and these were meant code as if-else stepping stools. We at that point put these steps into PHP class techniques that acknowledge just the part characteristics - PCM rate, justify marks, confirmation compose and sexual orientation as strategy parameters. The class techniques restore the last aftereffect of that specific assessment, demonstrating whether that understudy would pass or flop in the main semester examination. Figure 7 demonstrates a class technique with the if-else stepping stool.



4.4.7.

Figure 5. Registration page for staff members

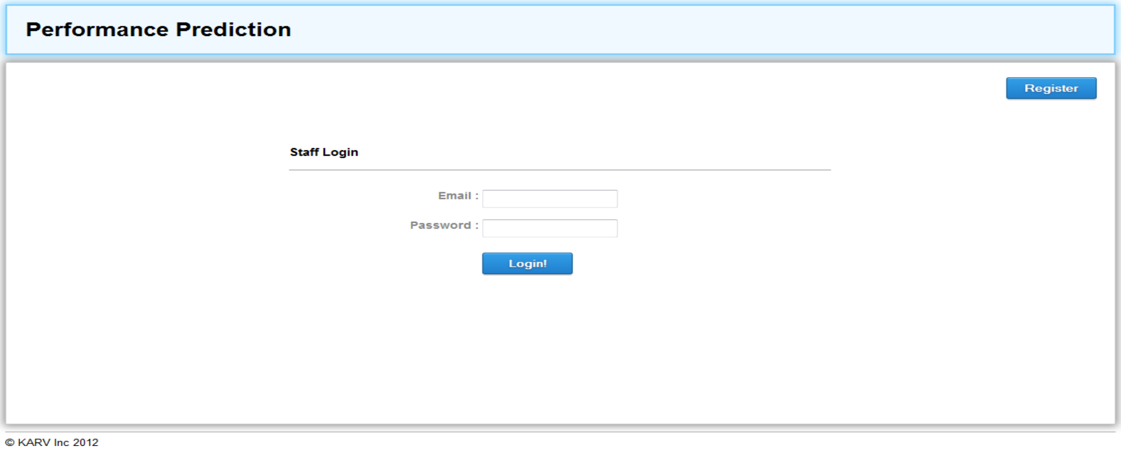


Figure 6. Login page for staff members

4.4.8. SingularEvaluation

Once the choice trees were mapped as class strategies, we manufactured a site page for staff individuals to sustain esteems for the name, application ID and part properties of an understudy, as can be found in Figure 8. These qualities were then used to foresee the consequence of that understudy as either "Pass" or "Fall flat".

4.4.9. Upload Excel Sheet

Particular Evaluation is helpful when the consequences of few understudies are to be anticipated, each one in turn. However, if there should be an occurrence of huge testing datasets, it is achievable to transfer an information document in an organization, for example, that of a Microsoft Excel spreadsheet, and assess every understudy's record. For this, staff individuals can transfer a spreadsheet containing records of understudies with qualities in a foreordained request. Figure 9 demonstrates the transfer page for Excel spreadsheets.

```php
public function dtalgo3($percent, $merit, $ad_type, $gender){
    if( $percent === "distinction" )
        return "pass";
    else{
        if( $percent === "first_class" ){
            if( $merit === "bad" ){
                if( $ad_type === "AI" )
                    return "pass";
                else
                    return "fail";
            }
            else
                return "pass";
        }
        else
            return "fail";
    }

}
```

Figure 7. PHP class method mapping a decisiontree

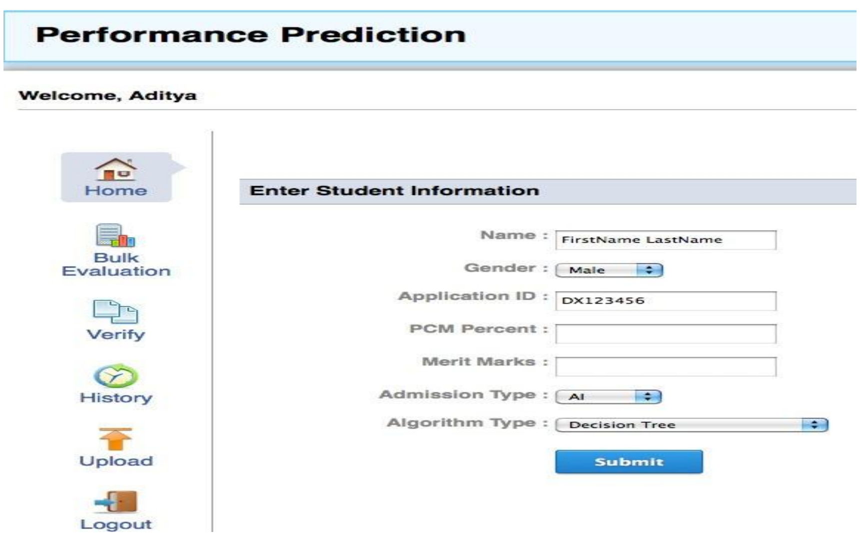


Figure 8. Web page for SingularEvaluation

4.4.9. BulkEvaluation

Under the Bulk Evaluation tab, a staff part can pick a transferred dataset to assess the outcomes, alongside the calculation to be connected over it. In the wake of presenting the dataset and calculation, the anticipated aftereffect of every understudy is shown in a table as the estimation of the trait "class". An example consequence of Bulk Evaluation can be found in Figure 10.

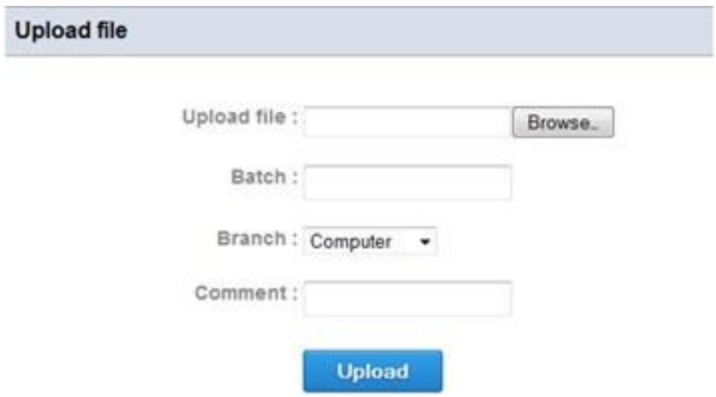


Figure 9. Page to upload Excel spreadsheet

merit_marks	app_id	name	gender	caste	location	percent	type	class
153	DX10205034	AKSHAY DEBNATH	Male	Open	Mumbai	95.66	AI	PASS
152	DX10279070	YEMPALLE SUSHMA BASWARAJ	Female	Open	Mumbai	86.66	AI	PASS
143	DX10288911	KIRAN SUSHIL GRIFFITHS	Male	Open	Mumbai	96	AI	PASS
136	DX10167854	WALCHALE ABHIJEET SUHAS	Male	Open	Mumbai	82	AI	PASS
132	DX10255786	KUNAL JADHAV	Male	Open	Mumbai	80.33	AI	PASS
109	DX10230782	KARKHELE RAVINDRAKUMAR VITTHAL	Male	NT 3 (NT-D)	Mumbai	83.66	GNT3H	PASS
156	DX10172564	TALAWADEKAR ADITYA SHYAM	Male	OBC	Mumbai	89.33	GOBCH	PASS

Figure 10. Page showing results after Bulk Evaluation

4.4.10. Verifying Accuracy of PredictedResults

The precision of the calculation results can be tried under the Verify tab. A staff part needs to choose the transferred confirmation record which as of now has the genuine outcomes and the calculation that must be tried for exactness. After accommodation the anticipated aftereffect of assessment is contrasted and genuine outcomes got and the exactness is computed. Figure 11 demonstrates that the exactness accomplished is 75.145% for both ID3 and C4.5 calculations. Figure 12 demonstrates the bungled tuples, i.e. the tuples which were anticipated wrongly by the application for the present test information.

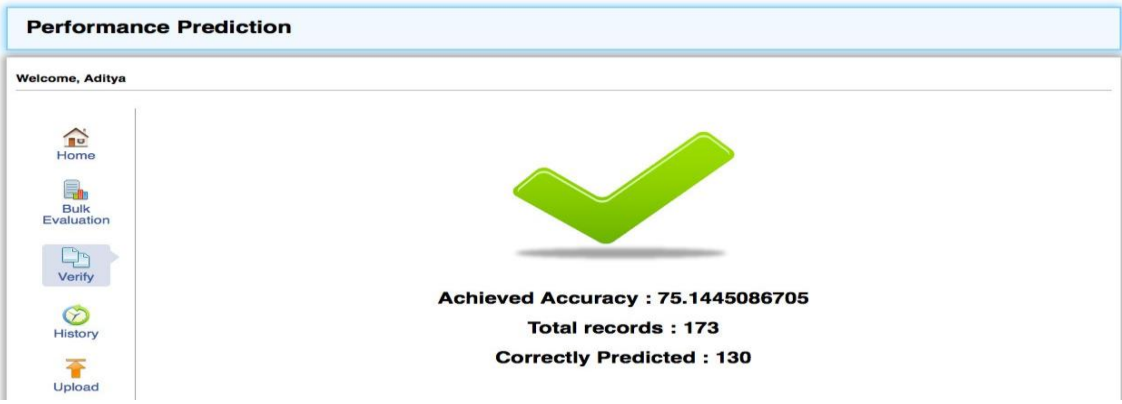


Figure 11. Accuracy achieved afterevaluation

merit_marks	app_id	name	gender	caste	location	percent	type	class	Predicted
140	DX10196064	PATIL SUMEET BHAGWAN	Male	OBC	Mumbai	88.33	GOBCH	fail	PASS
124	DX10297565	MAHAJAN NISHANT VIJAY	Male	OBC	North Maharashtra	58	GOBCO	fail	PASS
118	DX10356072	NARKHEDE JUHI RAJEEV	Female	Open	North Maharashtra	76.33	LOPENO	pass	FAIL
108	DX10149595	WAGHAMARE LAXMAN PANDURANG	Male	OBC	Shivaji + Solapur	74	GOBCO	pass	FAIL
153	DX10182982	JAISWAL ABHAY SHAILESH	Male	Open	Mumbai	75.66	GOPENH	fail	PASS
150	DX10193225	RAJPUT ABHISHEK DANSINGH	Male	Open	Mumbai	82	GOPENH	fail	PASS
93	DX10260441	RAMYA MACHERI	Female	Open	Mumbai	73	AI	pass	FAIL

Figure 12. Mismatched tuples shown duringverification

4.4.11. Singular EvaluationHistory

4.4.12. Utilizing the web interface, staff individuals can see every single Singular Evaluation they had led before. This is shown as a table, containing traits of the understudy and the anticipated outcome. Whenever required, a record from this table might be erased by a staff part. A preview of this table is appeared in Figure 13.

Application ID	Name	Gender	Percentage	Merit marks	Admission Type	Algorithm	Class	
DX123456	Aditya Gaykar	Male	89.17	157	OTHER	C4.5	pass	Delete
DX123456	Rahul	Male	123	89	OTHER	Decision Tree	pass	Delete
DX121312	Aditya Gaykar	Male	90.33	157	OTHER	Decision Tree	pass	Delete

4.4.13.

Figure 13. History of Singular Evaluations performed by staff members

5. FUTURE WORK

In this task, forecast parameters, for example, the choice trees created utilizing Rapid Miner are not refreshed progressively inside the source code. Later on, we intend to make the whole execution dynamic to prepare the forecast parameters itself when new preparing sets are bolstered into the web application. Additionally, in the present usage, we have not considered additional curricular exercises and other professional courses finished by understudies, which we accept may significantly affect the general execution of the understudies. Considering such parameters would result in better precision of forecast.

6. CONCLUSIONS

In this paper, we have clarified the framework we have used to foresee the consequences of understudies right now in the principal year of building, in view of the outcomes acquired by understudies as of now in the second year of designing amid their first year.

The aftereffects of Bulk Evaluation are appeared in Table 1. Irregular experiments considered amid individual testing brought about roughly rise to exactness, as showed in Table 2.

Table 1. Results of Bulk Evaluation

Algorithm	Total Students	Students whose results are correctly predicted	Accuracy (%)	Execution Time (in milliseconds)
ID3	173	130	75.145	47.6
C4.5	173	130	75.145	39.1

Table 2. Results of Singular Evaluation.

Algorithm	Total Students	Students whose results are correctly predicted	Accuracy (%)
ID3	9	7	77.778
C4.5	9	7	77.778

Thus, for a total of 182 students, the average percentage of accuracy achieved in Bulk and Singular Evaluations is approximately 75.275.

REFERENCES

[1] Han, J. and Kamber, M., (2006) *Data Mining: Concepts and Techniques*,Elsevier.

[2] Dunham, M.H., (2003) *Data Mining: Introductory and Advanced Topics*, Pearson

(IJIASE) 2018, Vol. No. 4, Jan-Dec **e-ISSN: 2454-9258, p-ISSN: 2454-809X**

EducationInc.

[3] Kantardzic, M., (2011) *Data Mining: Concepts, Models, Methods and Algorithms*, Wiley-IEEE Press.

[4] Ming, H., Wenying, N. and Xu, L., (2009) "An improved decision tree classification algorithm based on ID3 and the application in score analysis", Chinese Control and Decision Conference (CCDC),pp1876-1879.

[5] Xiaoliang, Z., Jian, W., Hongcan Y., and Shangzhuo, W., (2009) "Research and Application of the improved Algorithm C4.5 on Decision Tree", International Conference on Test and Measurement (ICTM), Vol. 2,pp184-187.

[6] CodeIgnitor User Guide Version 2.14,http://ellislab.com/codeigniter/user-guide/toc.html

[7] RapidMiner,http://rapid-i.com/content/view/181/190/

[8] MySQL – The world's most popular open source database,http://www.mysql.com/