

NGO – CHILDREN’S FUTURE INDIA (CFI- MUMBAI UNIT)

Prof. Martina D’souza, Hrishikesh Alhad Chindarkar¹, Saurabh Subhash Gawade², Vinod Jotiba Bhogan³

^{1,2,3}Dept. of Information Technology, Xavier Institute Of Engineering, Mumbai

Martina.r@xavierengg.com

hrishichindarkar@gmail.com

gawadesaurabh5@gmail.com

vinodbhogan@gmail.com

Abstract— The main aim of any NGO is to provide maximum benefit to the needy ones who are being sponsored by that NGO. NGOs work in different sectors such as education, medicine, health care, etc. This paper discusses about the NGO - Children’s Future India (CFI), issues which are occurred in this NGO and several remedies to overcome the circumstances. Data mining is an essential part for classifying, predicting huge data sets. We will be focusing on ID3 as one classification algorithm to determine the student’s academic growth in each phase; so that instead of waiting for the end examination result, it will be possible to evaluate a student and improve the academic record on the basis of his/her performance in one phase before moving to the next phase. The website will consists of new front-end technology like Angular6, Bootstrap and back-end technology like MySQL. For database, scripts will be written in PHP.

Keywords— User responsive website, data mining, classification algorithm, ID3 algorithm.

I. INTRODUCTION

Non-Governmental Organisations (NGOs) are usually non-profitable and are active in humanitarian, educational, health care, social, human rights, public policy, environmental, and other areas to affect changes according to their objectives. NGOs have a hierarchy in each discrete unit. Just like a manager in a corporate world follows up each of his team member regularly to check if all of his team members are working to execute a given task in a scheduled time and each team member has to report to his senior about his work; NGO also functions in a proper hierarchy to carry out all the tasks smoothly.

Children’s Future India (CFI) is a secular and non-profit voluntary organisation that works dedicatedly for the betterment of underprivileged children, their families and their communities among various places in India.

II. PROBLEM DEFINITION

At CFI, each and every student’s separate file is being maintained manually from the period a student is enrolled. CFI provides various services such as scholarships up to graduation level, tuition facility by having tie-ups with various coaching classes, etc. There are about 250-300 students at CFI-Mumbai Unit and as a result, it becomes quite hectic to maintain their academic and financial records with them manually. The main objective of developing a functional website for them is to ease their manual efforts and mitigate time consumption. Effective source of communication among people across the globe, expansion of the organisation within less period of time, accuracy in evaluating students’ academic as well as organisation’s performance and providing forecasts to avoid circumstances ahead in future are some of the key aspects in this project.

III. RELATED WORK

1. **Chin-Chia Hsu and Tao Huang[1]** researched on the application of data mining for evaluating candidate’s academic performance through various means such as participation in extra-curriculum activities, improvement in communication skills and grades, etc.
2. **Osofisan and Olamiti[2]** investigated students’ performance in a Computer Science programme by applying C4.5 learning algorithm in developing the student’s performance model to determine the marks subject-wise.
3. **Ashily M Baby[3]** stated that the Decision Tree J48, i.e., C4.5 algorithm is better than other algorithms to predict faculty’s performance. Apart from C4.5, Naive Bayes classifier can also add up to the enhancement of the project.
4. **PunamKumari and RainuNandal[4]** proposed a brief idea about various steps involved in developing a website right from analysis, specification up to testing, security, maintenance and update. This paper also includes information about

XAMPP and PHP which lets developers for creating a local web server for website testing and development process and clarifies with the scripts written in PHP respectively.

5. **Satoto, Isnanto, Kridalukmana, Martono**[5] proposed a concise study on MySQL stating the reasons behind the most utilized database platform till today's date, role of database design in deciding system's performance, process optimization, instant requests processing to the server, etc.
6. **Brijesh Kumar Baradwaj and Saurabh Pal**[9] has published a research paper which shows classification of student's academic performance using decision tree(ID3) algorithm. In this paper, it uses information such as Attendance of students, Class test, Seminar and Assignment marks were collected from the student's previous database and predict the performance at the end of the term.
7. **V. Shanmugarajeshwari and R. Lawrance**[10] evaluated student's academic performance using classification technique, i.e., C5.0 algorithm. It also alerts the students as well as the teachers to enhance their academic performance.

IV. PROPOSED SYSTEM

1. Data Mining :

Data mining is a knowledge discovery process that involves mathematical and statistical analysis to determine the patterns and trends of data from large sets of data.[6] For an instance, to perform a survey of candidates appearing for the Common Proficiency Test (CPT), data mining can be used in order to determine the number of candidates appearing for the examination in the current year, the number of candidates appearing for the first attempt, second attempt and so on, who cleared the exam in their first attempt in last 4-5 years, ratio of total number of candidates appeared to the total number of candidates qualified for the next exam, etc. Data mining came into existence somewhere around in the year 1990. Applications of data mining include banking, retail, insurance, medicine, telecommunication, etc.

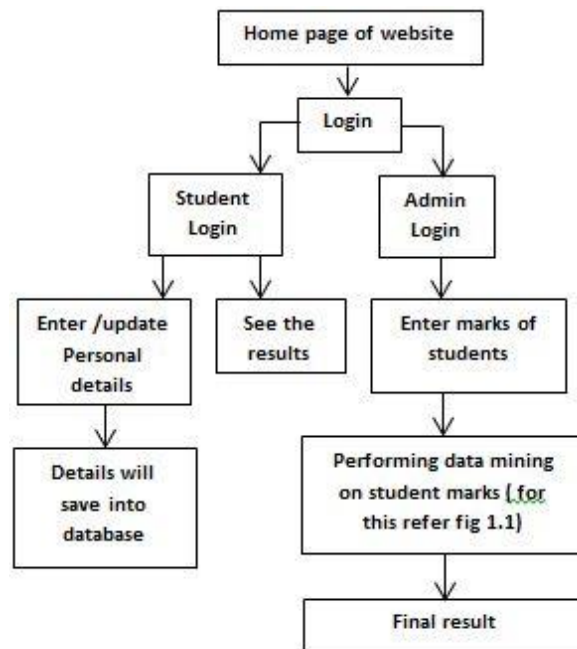


Fig. 1: Flow chart for Student performance assessment.

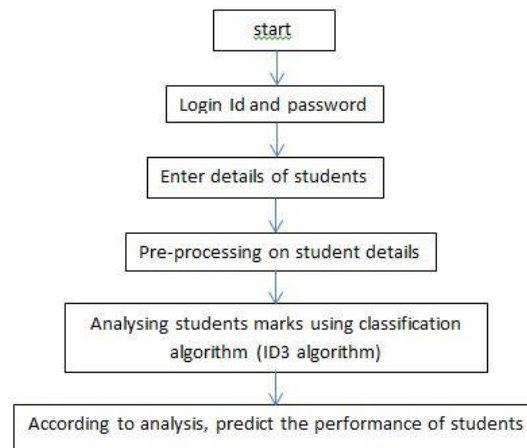


Fig. 1.1: Process to perform data mining on student's marks.

2. Classification Algorithm:

Classification is a major type of prediction problem where classification is used to predict discrete or nominal values. It is the use of prediction to predict class labels. Example: Group patients based on their known medical data and treatment outcome then this is called classification; whereas prediction is used to assess the value or value ranges of an attribute that a given sample is likely to have. Example: if a classification model is used to predict the treatment outcome for a new patient then it would be a prediction.[7]

Various classification techniques are:

- i. Regression
- ii. Decision tree
- iii. Rules
- iv. Neural networks

Classification using decision tree:

A decision tree uses a tree data structure or model of decisions and their possible effects. It consists of three types of nodes: decision node, chance node, leaf node (final decision). To construct decision tree we need to follow some greedy algorithm. It must be constructed by following a top-down, recursive divide-conquer approach.[7] Classification methods are:

- i. Decision Tree Induction: Attribute selection measures, tree pruning
- ii. Bayesian Classification: Naïve Bayes' classifier

Attribute selection measure:

- i. Gini Index(IBM Intelligent Miner)
- ii. Information gain(ID3/C4.5)
- iii. Gain Ratio

Iterative Dichotomiser 3 (ID3):

ID3 algorithm requires information gain as splitting criteria. It is a simple decision tree algorithm. It doesn't follow the pruning procedures to handle any missing value. It is heuristic in nature; prefer simplicity to avoid unnecessary assumptions.[6]

Formulae:[6]

- i. Expected information (Entropy) is needed to classify a tuple in D is given by:

$$\text{Info}(D) = -\sum_{i=1}^m p_i \log_2(p_i)$$
- ii. Information needed (after using A to split D into v partitions) to classify D is given by:

$$\text{Info}_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} * i(D_j)$$
- iii. Information gained by branching on attribute A is represented by:

$$\text{Gain}(A) = \text{Info}(D) - \text{Info}_A(D)$$

Pros:

- Helps to construct decision tree in minimum number of steps.
- Tree induction is determined by asking appropriate questions.
- The next level is benefited from the previous one.[8]

Cons:

- To make a decision, only one attribute is tested at a time.
- A new tree must be created; when new data is not classified correctly which in turn tree cannot be updated.[8]

Criteria	Result
>16	Next phase
>=12&&<=16	Next phase and revision of the previous phase
<12	Same phase

Fig 2.1: Rule Table

ID	Phase 1.1	Phase 1.2	Phase 1.3	Next Phase
1	>16	>=12&&<=16	>16	Yes
2	>=12&&<=16	<12	>=12&&<=16	No
3	>16	>=12&&<=16	>=12&&<=16	Yes
4	>=12&&<=16	>16	<12	No
5	<12	>=12&&<=16	>=12&&<=16	No
6	<12	>16	>=12&&<=16	Yes
7	>16	>=12&&<=16	<12	No
8	>16	>=12&&<=16	>16	Yes
9	>=12&&<=16	<12	>=12&&<=16	No
10	>=12&&<=16	>16	>16	Yes

Fig 2.2: Student Analysis Table (Phase 1)

Phase	Gain Value
Phase 1.1	0.16
Phase 1.2	0.23
Phase 1.3	0.51

Fig 2.3: Gain Value

V. RESULTS



Fig. 3: Homepage

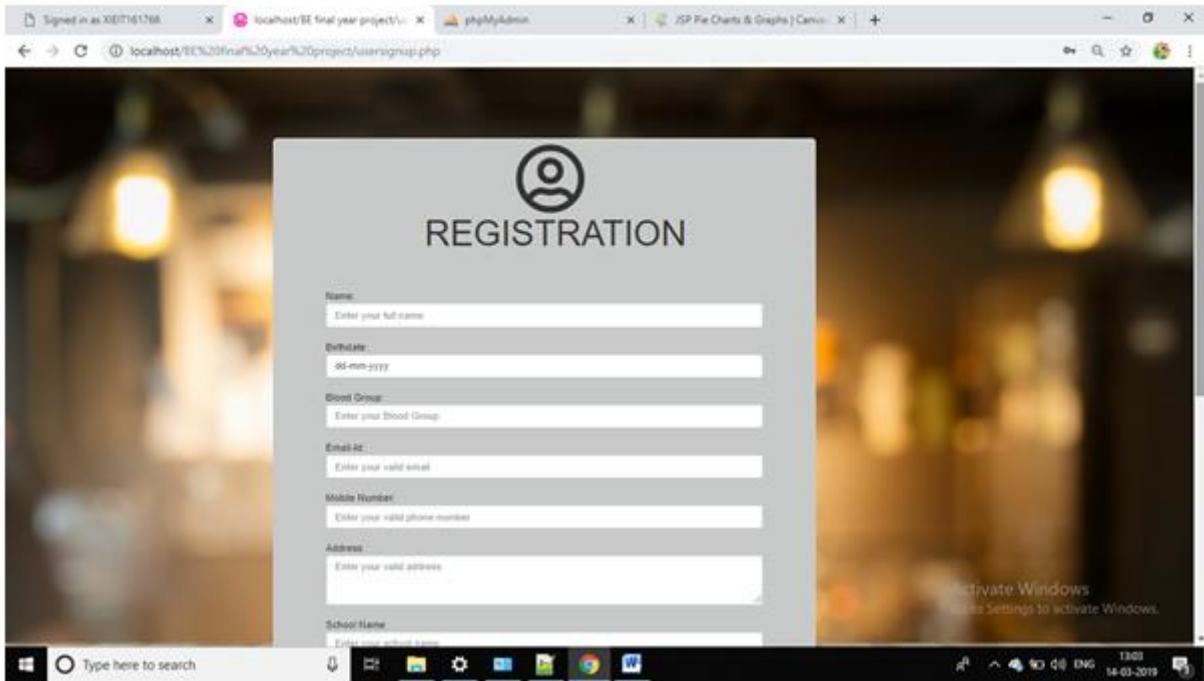


Fig 4: User Registration

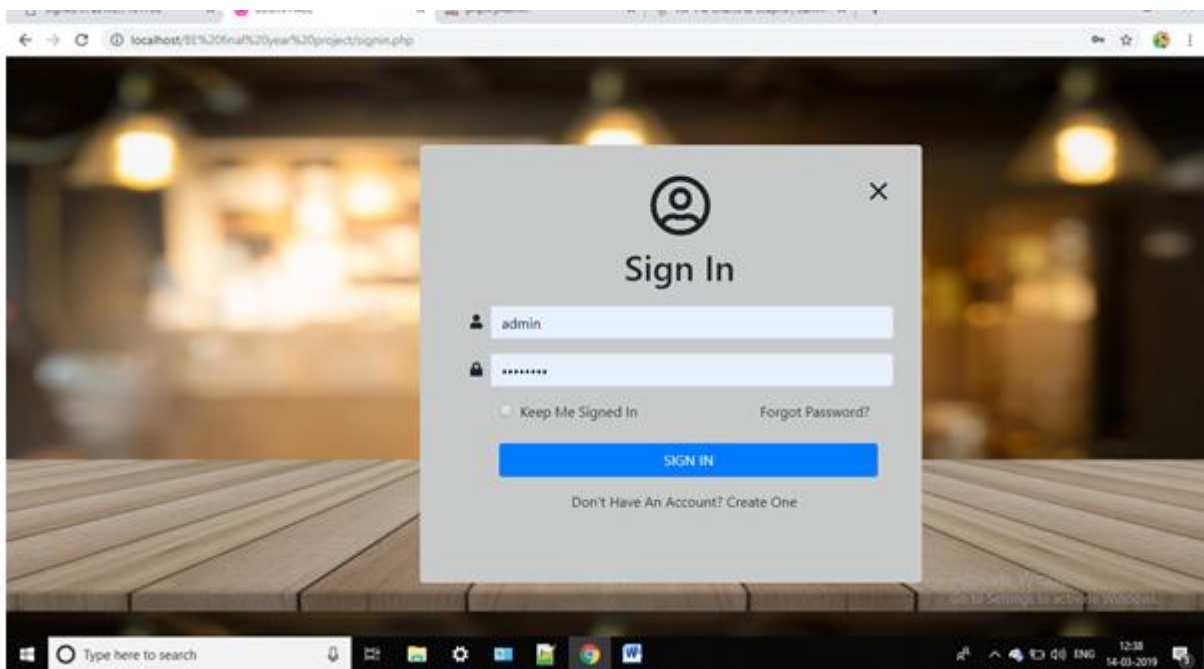


Fig 5: User Sign In

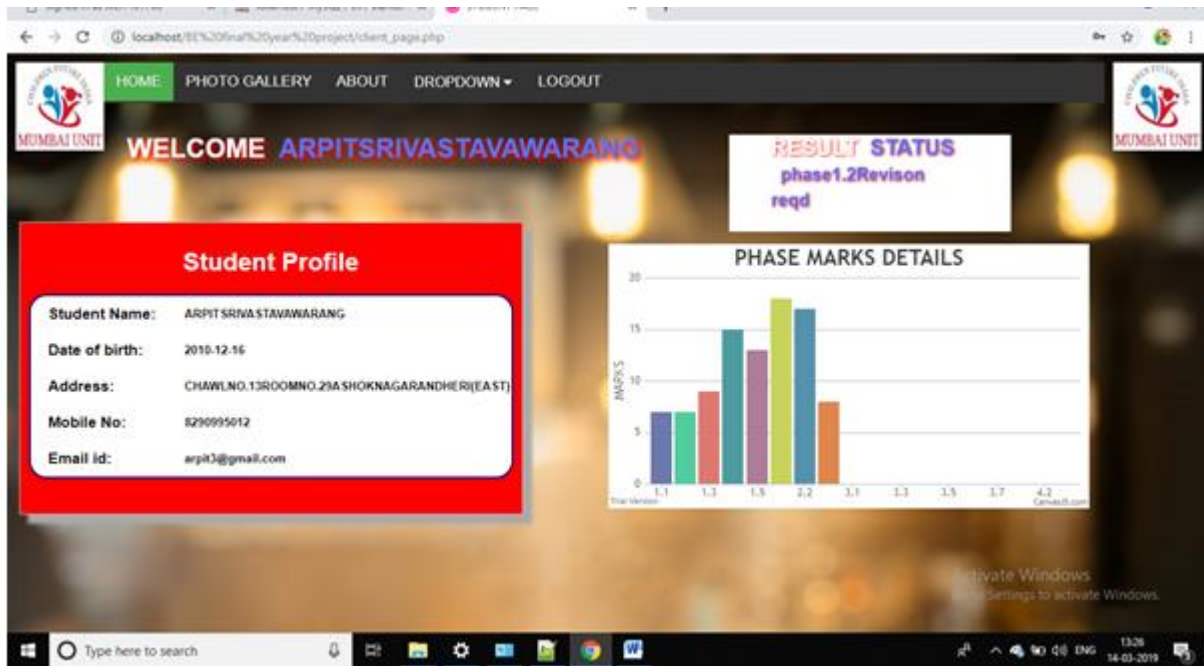


Fig 6: User Profile

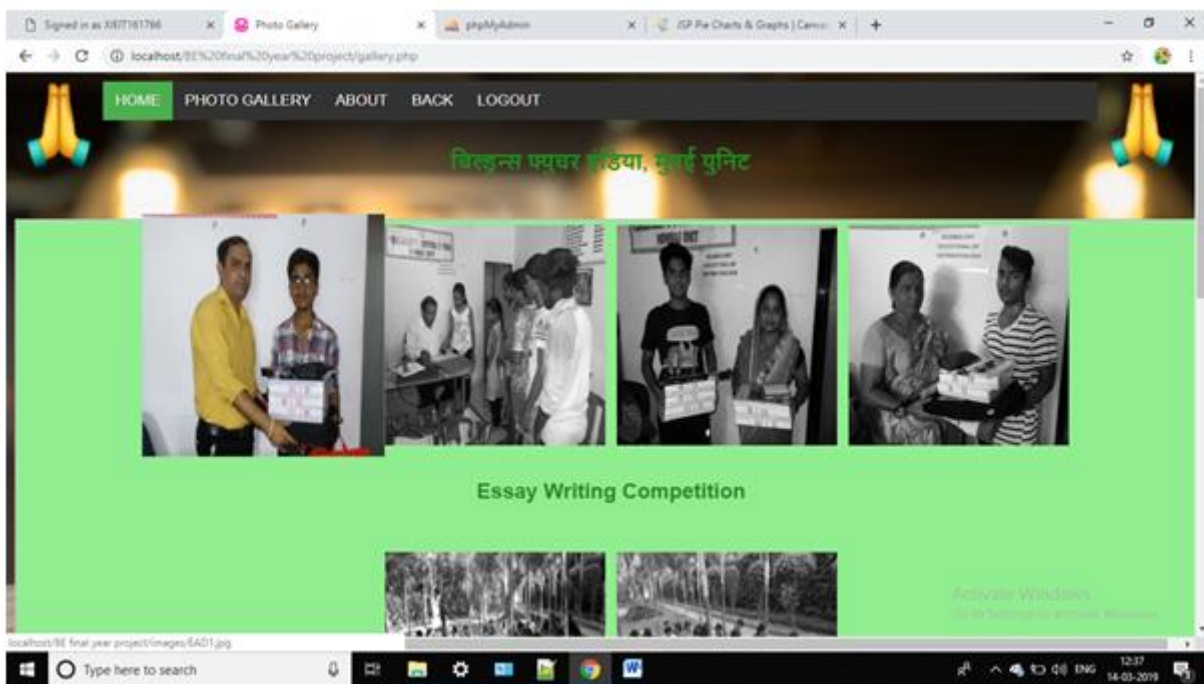


Fig 7: Photo Gallery

VI. CONCLUSION

As we know that a functional website not only depends on the front-end technologies but also on the back-end technologies. A basic webpage can be developed using HTML but adding couple of extra features to it using CSS, JavaScript will be a step ahead of its development. Other than SQL, MySQL is a better option for storing and referring data. To evaluate a student's academic performance, various classification algorithms like Naïve Bayes', C4.5 have been used. On a small-scale basis, we will be using ID3 as the classification algorithm in our project.

VII. ACKNOWLEDGEMENT

We express gratitude towards our project guide Prof. Martina D'Souza, project co-ordinator Prof. Suvarna Arango, Prof. Prachi Tawde and the entire IT department for helping and motivating us to gain insight of the implementation of the project.

VIII. REFERENCES

- [1] Chin Chia Hsu and Tao Huang, 2006 "The use of Data Mining Technology to Evaluate Student's Academic Achievement via multiple Channels of Enrolment", *An empirical analysis of St. John's University of Technology*.
- [2] Osofisan A.O. and Olamiti A.O, 2009 "Academic Background of Students and Performance Computer Science Programme", *Nigerian University*.
- [3] Ashily M Baby 2018 IOP Conf. Ser.: Mater. Sci. Eng. 396012024 "Pedagogue Performance Assessment (PPA)using Data miningTechniques".
- [4] PunamKumari and RainuNandal (2017), "A Research Paper On Website Development Optimization Using Xampp/PHP", *International Journal of Advanced Research in Computer Science; Volume 8, No. 5, May-June 2017; ISSN No. 0976-5697. PP[1231-1235]*
- [5] KodratImanSatoto , R. Rizal Isnanto, RintaKridalukmana, KurniawanTeguhMartono (2016), "Optimizing MySQL database system on information systems research, publications and community service", *3rd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE); INSPEC Accession Number : 16791795*.
- [6] S.K.Shinde and UddagiriChandrashekhar (2015), "Data Mining and Business Intelligence", *Dreamtech Press, NewDelhi*.
- [7] AartiDeshpande, "Data Mining and Business Intelligence", *Techmax Publication, Pune*.
- [8] <https://www.slideshare.net/HARDIKSINGH7/id3-algorithm-56632554>
- [9] Brijesh Kumar Baradwaj and Saurabh Pal, (2011), "Mining Educational Data to Analyze Students' Performance", *(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6,pp[63-69]*
- [10] V. Shanmugarajeshwari and R. Lawrance, (2016), "Analysis of Students' Performance Evaluation using Classification Techniques".