

KDD CUP 2015- Predicting Dropouts in MOOC'S

Priti Kadam , Jayashree Palve , Kranti Kusale & Nikita Sankhe
Department of Information Technology, Mumbai, India

Abstract: *There are education related problems today that have a direct or an indirect connection to the respective dropout students. Efficient methods of the same and spreading proper awareness about the online courses associated with education is the today's need. The idea that a "Better use of technology leads to a Better solution" is the inspiration of the proposed system . The proposed system advises the user according to his/her requirement. The requirement ranges from gaining knowledge, maintaining progress and reducing drawbacks of existing system. The proposed system uses data mining as the tool to bridge the gap between the education and the users .The proposed system uses a classification algorithm to efficiently direct the respective user to the best possible solution. The WEKA software is used for the study implementation since it is freely available to the public and is widely used for research purposes in the data mining field*

Keywords: *weka, ID3algorithm, Eclipse.*

I. INTRODUCTION

The proposed system uses a classification algorithm which gives information about no.of users intrested or not,it builds a decision tree from set of data and the resulting tree is used to classify future samples. Today the primary problem of education issues is based on exclusively considering the traditional teaching system. The proposed system is an upgrade to this approach, it considers various online educational factors("clickstream data") along with the MOOCS to efficiently groups the data according to behaviour of the users at the time of surfing . According to the system places the user in one of the following two groups: Intrested,Non-Intrested. The proposed system uses a classification algorithm which gives information about no.of users intrested or not,it builds a decision tree from set of data and the resulting tree is used to classify future samples.

METHODOLOGY

The proposed system efficiently uses classification algorithms to effectively deliver the best possible expert advice to the use's problem .the computing for teachers MOOC was designed and produced by the Department of computer science at The university of warwick .the purpose of the CFT MOOC was to assist school teacher in the delivery of the increased computing contain being introduced to U.K schools.The course ws to delivered in two different modes : "Tradtional" MOOC mode with peer support and automated assessment options, and a paid "supported" mode which also included on-line tutor support and tutor supported forums and resources.

SYSTEM DESIGN

1.1 UML Diagrams

UML stands for Unified Modeling Language which is used in object oriented software engineering. Although typically used in software engineering it is a rich language that can be used to model an application structures, behavior and even business processes.

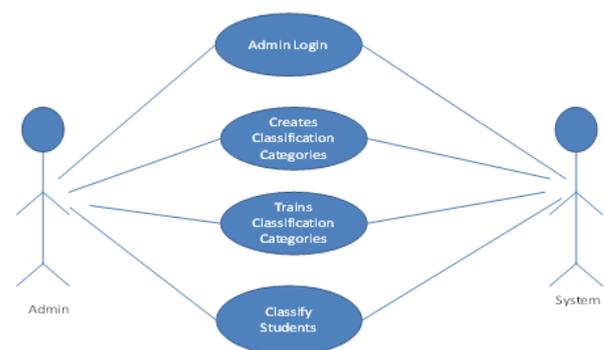


Fig. 1.1 Use case

1.2 Data Flow Diagrams

A data flow diagram (DFD) illustrates how data is processed by a system in terms of inputs and outputs. As its name indicates its focus is on the flow of information, where data comes from, where it goes and how it gets stored.

1.2.1 DFD Level 0:

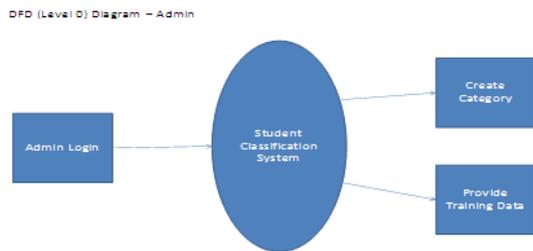


Fig. 1.2.1 Data Flow Diagram Level 0

Description:

1. The user Admin logs in with his/her login id and password.
2. Admin classified student data.
3. Create category on student such as Interested or not Interested.
4. Provide student data to train the data.

1.2.2 DFD Level 1:

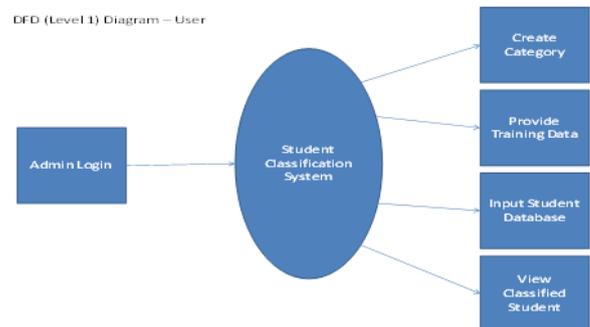


Fig. 1.2.2 Data Flow Diagram Level 1

Description:

1. The user Admin logs in with his/her login id and password.
2. Admin classified student data.
3. Create category on student such as Interested or not Interested.
4. Provide student data to train the data.
5. Admin can add new entry of student into the dataset after classification.
6. Admin can view classified student.

1.3 Activity Diagram

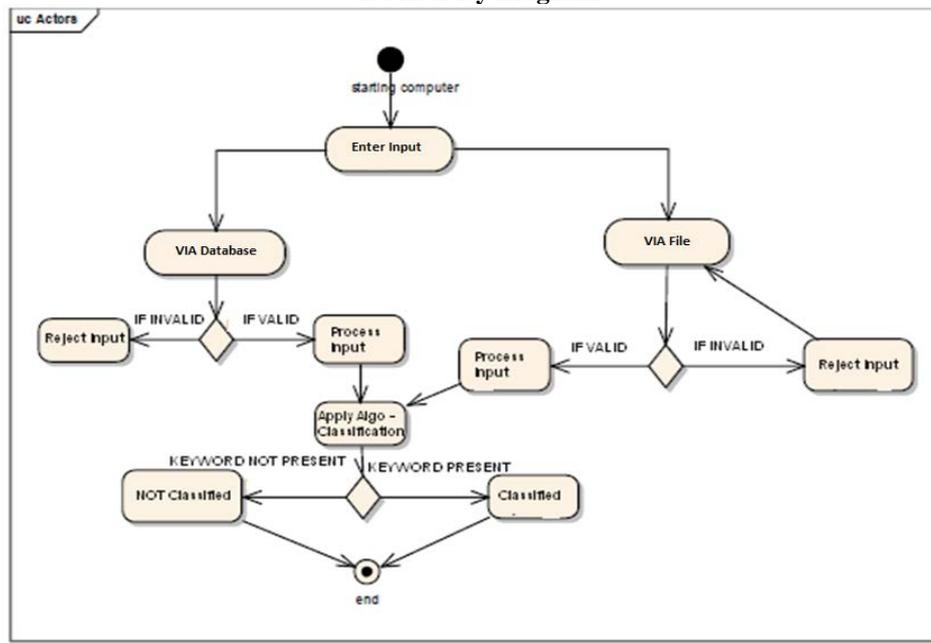


Fig. 1.3 Activity Diagram

1.4 SYSTEM BLOCK DIAGRAM:

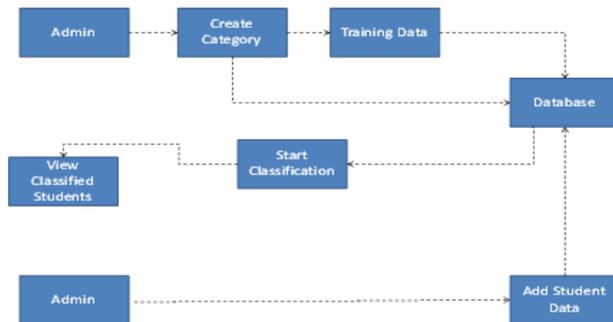


Fig.1.4 System Block Diagram

IMPLEMENTATION PLAN

In our project we are using ID3 algorithm, in our project firstly we are accepting course data and student data after that mapping of student Id and course Id is occur then it will start processing. Next step is to check view time and total time of mapping then it will compare with threshold. If it is greater than threshold set as Interested else not Interested. Further proceed with to checking total Interested column and compare with threshold. If it is greater than threshold set as not dropout else dropout.

Algorithm: ID3 Algorithm

- Step 1: start
- Step 2: Accept course data
- Step 3: Accept student data
- Step 4: Accept mapping data
- Step 5: start processing
- Step 6: Check view time and total time of mapping, compare with threshold.
- Step 7: If it is greater than threshold set as Interested else not Interested.
- Step 8: Check total Interested column and compare with threshold.
- Step 9: If it is greater than threshold set as not dropout else dropout.
- Step 10: Stop

ACKNOWLEDGMENT

It gives us great pleasure in presenting this project report titled “KDD Cup 2015 - Predicting dropouts in MOOC ”. On this momentous occasion, we

wish to express our immense gratitude to the range of people who provided invaluable support in the completion of this project. Their guidance and encouragement has helped in making this project a great success.

We express our gratitude to our project guide Prof Nutan Dhang mam , who provided us with all the guidance and encouragement and making the lab available to us at any time. We also would like to deeply express our sincere gratitude to Project coordinators.

We are eager and glad to express our gratitude to the Head of the Information Technology Dept. Prof Nileema Pathak , for her approval of this project. We are also thankful to her for providing us the needed assistance, detailed suggestions and also encouragement to do the project.

We would like to deeply express our sincere gratitude to our respected principal Prof. Dr. Shirking Kallurkar and the management of Atharva College of Engineering for providing such an ideal atmosphere to build up this project with well equipped library with all the utmost necessary reference materials and up to date IT Laboratories

We are extremely thankful to all staff and the management of the college for providing us all the facilities and resources required.

REFERENCES

- [1] Lee, Y., & Choi . A review of online course dropout research: Implications for practice and future research .Educational Technology researche and Development ,J.2011
- [2] Edin Osmanbegović , Mirza Sulji, DATA MINING APPROACH FOR PREDICTING STUDENT PERFORMANCE. Journal of Economics and Business, Vol. X, Issue 1, May 2012
- [3] Riti Lath, Manish Shrivastava, Analytical Study of Different Classification Technique for KDD Cup Data’99
- [4] Peterson, The Pros and Cons of MOOCS, Available at: <http://adulthood.about.com/od/Adult-Education-in-the-U.S./a/The-Pros-And-Cons-Of-Moocs.htm> D. (2014).