Analysis and Prediction of students behaviour by using classification as an data mining techniques

Mayur R. Saner
Shiwani Suryavanshi
Vinaya patil
Bhushan Sarode
Viplav Patil
Under the Guidance of,
Mr.Sushant S. Bahekar

S.S.B.T.COET.Jalgaon

November 10, 2016

Abstract

- Students opting engineering as their disciple is **increasing rapidly**. But due to various factors and inappropriate primary education in India **dropout rates are high**.
- With the help of **data mining techniques we can predict the performance of students** in terms of grades and dropout for a subject.
- There are various data mining methods such as **classification, clustering and association.**
- Classification is supervised learning method that **builds a model to classify a data item** into a particular class label.The aim of classification is **to predict the future outcome based on the current available data.**

- data analysis task as an Classification
  E.g.
  A bank loan officer wants to analyze the data in order to know which customer (loan applicant) are risky or which are safe.
  In this e.g a model or classifier is constructed to predict the categorical labels which are **risky or safe** for loan application data.

- We have the following four goals concerned with Educational Data Mining (EDM):
  - **Predicting** student's future learning behaviour
  - **Discovering** or **Improving** Educational domain needs
  - Studying the **factors affecting Educational support**
  - Providing the **Advanced scientific knowledge parameters** to the Educational Institutes

Problem In Existing Environment

- Education is the key to prosperity of any nation.
- Unfortunately due to **lack of quality education at primary level**, socio- economic, psychological and other diverse factors, students **drop out rates** are high and performance is low.
- Everyone talks about the problem, no one is thinking about the solution.
- Everyone is blaming to schools, colleges.

Solution Of The Problem

- To **improve the quality of engineering** graduates such cases of dropout and poor performance must be monitored proactively.
- Developement of an data mining tool which helps to finds the behaviour of student using various Classification algorithms.
- This help us to get insight of data and reach to meaningful conclusion.

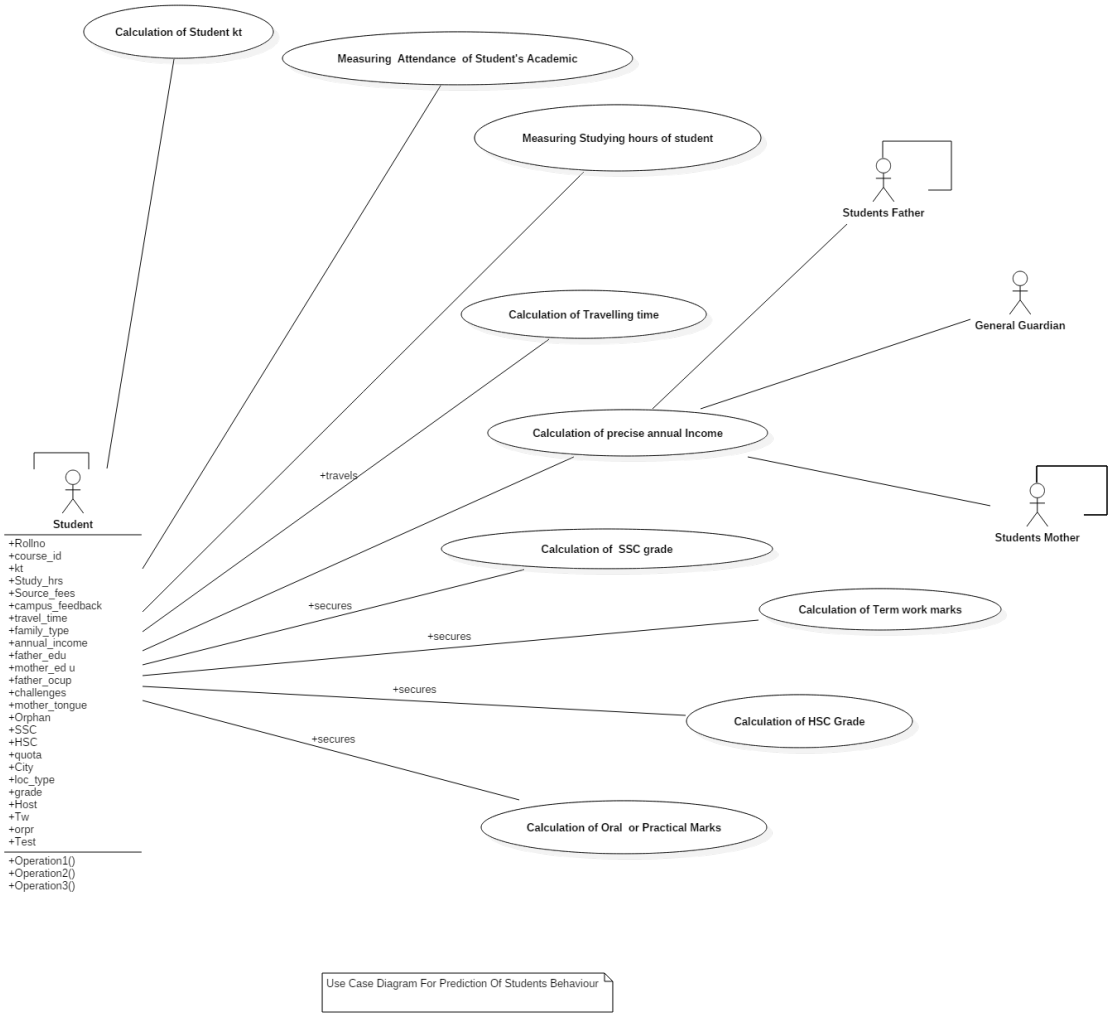
Methodology

- With the help of **classification of data mining techniques** we can predict the performance of students in terms of grades and dropout for a subject.
- In this project we are going to develop new technique by overcoming the disadvantages of present Classification Techniques such as **Random Forest, Naive Bayes and JRip** etc.
- We are **not going to concern any psychological factors** such as family issues, cast, gender, health, family type, etc in this project.

Modelling

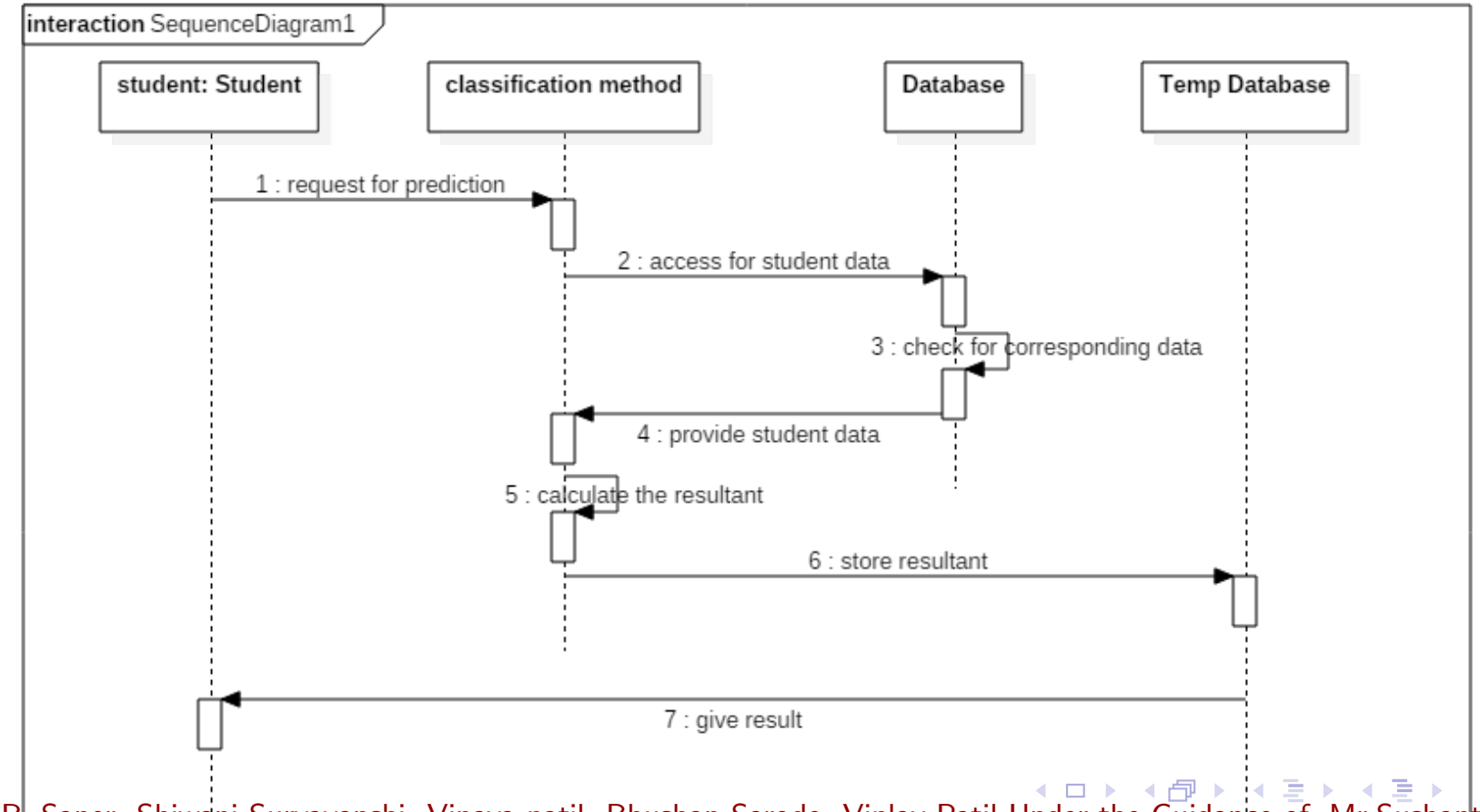
- In this Phase we have modelled an Use case,Sequence Diagram and Class Diagram.

Modelling

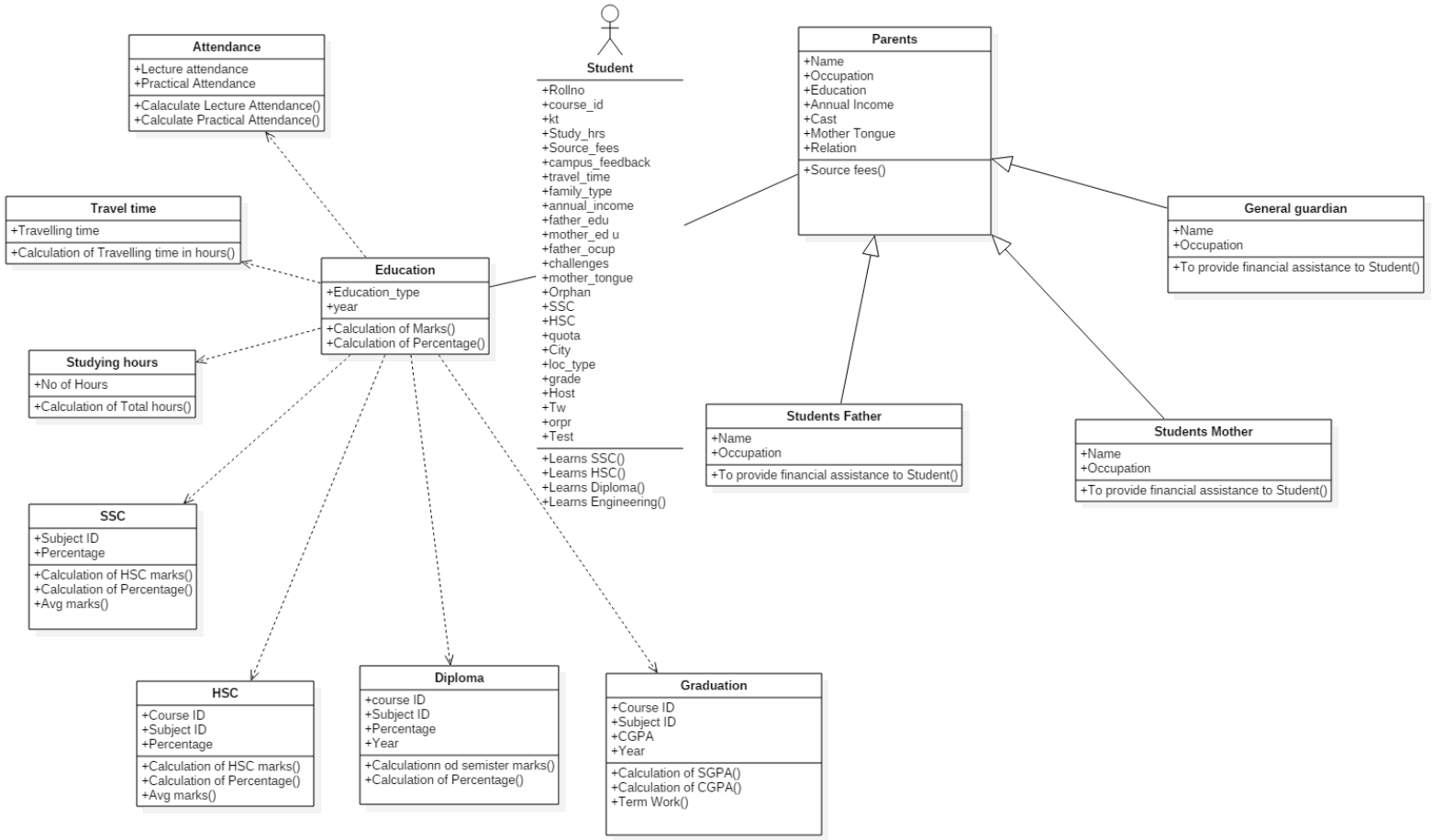

Calculation of Student kt
Measuring Attendance of Student's Academic
Measuring Studying hours of student
Students Father
Calculation of Travelling time
General Guardian
Calculation of precise annual Income
+travels
Student
+Rollno
+course_id
+kt
+Study_hrs
+Source_fees
+campus_feedback
+travel_time
+family_type
+annual_income
+father_edu
+mother_ed u
+father_ocup
+challenges
+mother_tongue
+Orphan
+SSC
+HSC
+quota
+City
+loc_type
+grade
+Host
+Tw
+orpr
+Test
+Operation1()
+Operation2()
+Operation3()
Students Mother
Calculation of SSC grade
+secures
Calculation of Term work marks
+secures
+secures
Calculation of HSC Grade
+secures
Calculation of Oral or Practical Marks
Use Case Diagram For Prediction Of Students Behaviour

Modelling

Figure: Sequence Diagram



Modelling



Class Diagram For Prediction Of Students Behaviour

Advantages

- Students will be able to know **their own area of interest**
- It will also be helpful for teachers as well as parents **To improve the weak point** of the student.
- It will help **to Improve the Result** of student and Colleges.
- It will help in Improving the student not only in studies but also in **extra curricular activies.**

Conclusion

- Student **dropout rates** in India are high.
- Data mining techniques can be able to predict the performance of students
- There is a need of Developing a tool of Data mining.
- Random Forest, Naive Bayes and JRip is found to be the Accurate Classification methods we are going to choose.

Reference

- Sonali Agarwal, G. N. Pandey, and M. D. Tiwari, Data Mining in Education: Data Classification and Decision Tree Approach, in the Journal of e-Education, e-Business, e-Management and e-Learning, Vol. 2, No. 2, April 2014, pp. 140-144
- Muhammed Salman Shamsi, J. Lakshmi, "Student performance prediction using classification data mining techniques ", in in the International Journal of Information and Education Technology, Vol. 6, No. 7, July 2016, pp. 528-533.
- M. Al-Razgan, A. S. Al-Khalifa, and H. S. Al-Khalifa, "Educational data mining: A systematic review of the published literature 20062013," in Proc. the 1st International Conference on Advanced Data and Information Engineering, 2013, pp. 711-719.