Applying Attribute Selection Algorithms in Academic Performance Prediction

Mu Lin Wong(✉) and S. Senthil

School of Computer Science and Applications,
REVA University, Bengaluru, India
joshuawml@yahoo.com, dir.csa@reva.edu.in

Abstract. There is an essential need to identify effective and efficient attribute selection algorithms for developing predictive and descriptive classifiers to predict students' academic performance. Personal, Academic and Socio-economic data were collected from 100 MCA students through questionnaire and cleansed. Eight attribute selection algorithms available in WEKA were applied to identify the best attributes. These attribute sets were then employed by four rule-based classifiers and four tree-based classifiers to examine the rate of accuracy, recall and AUC. The results proved that dimension reduction improves classifier performance by a maximum margin of 9% in accuracy, 25% in recall and 45% in AUC, when attributes were selected by subset-based algorithms. Both academic and socio-economic data are required to develop effective classifiers.

Keywords: Educational data mining · Academic performance prediction
Attribute selection algorithm · Rule-based classifiers · Tree-based classifiers
Historical data

1 Introduction

In the past decade, many researches on academic performance prediction focused on dimension reduction to increase the accuracy of the prediction models. There are various attribute selection algorithms available in WEKA data mining tool to select either a subset of best attributes or to list all the attributes in series of ranking. It is common for researchers to aggregate the best attributes selected from a few attribute selection algorithms and choose those that appear with the highest frequency. Even though this is a quick method, it remains unclear why such a method wouldn't provide a bias perspective. This is because each algorithm applies a different formula in its calculation.

Therefore, the purpose of this study is to compare which attribute selection algorithm is the most efficient and most effective to select the best attributes for academic performance prediction model development. In addition, this study also aimed to compare whether such attributes will demonstrate improved accuracy, recall and area under curve (AUC) in the prediction models. Besides attaining the predictability factor of the models, the third aim of this study is to attain the descriptive factor of these

© Springer Nature Switzerland AG 2019
J. Hemanth et al. (Eds.): ICICI 2018, LNDECT 26, pp. 694–701, 2019.
https://doi.org/10.1007/978-3-030-03146-6_78

models. Therefore, we will use four rule-based algorithms and four tree-based algorithms to develop the classification models.

2 Literature Review

Students' academic performance prediction has been one of the core applications of Educational Data Mining (EDM). In the two renown review papers written on EDM by Romero and Ventura [1, 2], the focus of academic performance prediction include online education and conventional classroom environment. Comparatively, it is much easier to collect real time online data from digital repository than to collect real time classroom engagement data. Another extensive review paper [3] reveals that most academic performance prediction researches done in 2011 to 2012 were more predictive than descriptive in nature. However, the current need of prediction models is to excel in both predictive and descriptive factors. More recent surveys unveil that classification was the most common data mining technique in academic performance prediction [4], and WEKA was the most common data mining tool used [5]. Performance prediction is not only limited to students, but analyzing faculty members' performance is gaining popularity in research too [6].

Various historical attributes were used to develop prediction models. However, it was found that these attributes generally falls into either one of three categories: Academic, Personal or Socio-economic [7, 8]. It is clear that just depending on a single category of attributes will not provide the best prediction models [9]. Recent studies [10, 11] reported including online data and self-report data as attributes. In order to improve the comprehensibility of the prediction models developed, it was recommended to use rule-based and/or tree-based algorithms [12, 13]. Equipped with an overview understanding from literature mentioned above, the study was conducted with methods described in the next segment.

3 Methodology

3.1 Data Collection

In order to achieve novelty, 100 MCA students was requested to fill in a questionnaire with 56 attributes. The semester 1 final result was collected from the college website and used as 2-class attribute (Pass or Fail). These attributes represent the standard three categories of attributes focused in EDM: Personality, Social Economic and Academic [7]. Table 1 shows the attributes collected according to their categories and data types (categorical or numerical).

3.2 Data Cleaning

The total number of attributes was reduced to 51 as some attributes were irrelevant to model accuracy. Date of Birth (DoB) was redundant as Age was present. Country was irrelevant as all students are from India. Student Registration Number (SRN), Name

Table 1. Attributes arranged according to attribute categories and data types.

Category	Attributes
Personal	SRN (C), Name (C), DoB (C), Age (N), Gender (C), State (C), Country (C), Ambition (C), Personality (C), Romance (C), SocialMedia (C), Sports (C), Cocurriculum (C), AttendanceMotivation (C), StudySkill (C)
Social-Economic	FamilyIncome (C), SiblingNumber (C), FatherEducation (C), MotherEducation (C), FatherJob (C), MotherJob (C), Infrastructure (C), SiblingEducation (C), ParentsTime (C), Topic (C), FamilyAccommodation (C), StudentAccommodation (C), CommutingTime (C), TransportMode (C)
Academic	SchoolName (C), SchoolType (C), City (C), SchoolState (C), ExamBoard (C), Stream (C), Core (C), Language (C), Maths (C), UGCourse (C), AvePercentage (N), UGFailure (N), S1C1_Math (N), S1C1_CF (N), S1C1_C (N), S1C1_WT (N), S1C1_Eng (N), S1C1_MIS (N), S1C1 (N), S1C2_Math (N), S1C2_CF (N), S1C2_C (N), S1C2_WT (N), S1C2_Eng (N), S1C2_MIS (N), S1C2 (N), S1Failure (N)

and SchoolName were all unique data, therefore unable to assist in developing the performance prediction model. This step not only increase the accuracy of the model developed, but also the time efficiency in developing such model.

3.3 Attribute Sets

Eight attribute selection algorithms available in WEKA data mining tool were chosen to determine the best attributes required to develop a classifier model with high accuracy. Two algorithms produced a subset of best attributes while the remaining six algorithms ranked all the attributes according to a descending order. In this study, all attributes with zero or negative ranking were removed from being used in classifier model development.

CfsSubsetEval and FilteredSubsetEval, produced the same result of 3 attributes, whereby 2 attributes were categorical and one was numerical. Therefore, FilteredSubsetEval was removed during analysis as to reduce redundancy. Similarly, ChiSquareAttributeEval was chosen to represent a group of 5 algorithms that produced the same results. The other 4 algorithms removed were FilteredAttributeEval, GainRatioAttributeEval, InfoGainAttributeEval and SymmetricalUncertAttributeEval. The output result was 32 categorical attributes and 2 numerical attributes. OneRAttributeEval produces 30 categorical attributes and 14 numerical attributes. The results of these algorithms were summarized in Table 2.

In order to understand the importance of the different attributes, four rule-based algorithms and four tree-based algorithms, as listed in the same order below, were used to develop the classifier models. These algorithms were chosen due to their ability to provide rules or criteria in making prediction.

- JRip [14], implements a proportional rule learner called Repeated Incremental Pruning to Produce Error Reduction (RIPPER) algorithm.

Table 2. Output of each attribute selection algorithm

Attribute Selection Algorithms	Output	# of Attributes
CfsSubsetEval ~~FilteredSubsetEval~~	Subset	3 (C:1, N:2)
ChiSquaredAttributeEval ~~FilteredAttributeEval~~ ~~GainRatioAttributeEval~~ ~~InfoGainAttributeEval~~ ~~SymmetricalUncertAttributeEval~~	Ranking	34 (C:32, N:2)
OneRAttributeEval	Ranking	44 (C:30, N:14)

- NNge [15, 16], implements a nearest-neighbor-like algorithm using non-nested generalized exemplars (if-then rules).
- PART [17], builds a partial C4.5 decision tree in each iteration and makes the best leaf into a rule.
- Ridor [18], is a RIpple-DOwn Rule learner that generates a default rule first and then the exceptions for the default rule with the least error rate.
- ADTree [19], is an alternating decision tree that supports only two-class problems.
- DecisionStump [20], is a simple one-level decision tree classifier usually used in conjunction with a boosting algorithm.
- J48 [21], generates a pruned or unpruned C4.5 decision tree.
- LADTree [22], is a multi-class alternating decision tree using the LogitBoost strategy.

For each classification algorithm mentioned above, a set of attributes selected based on each attribute selection algorithm is used in classifier development and comparison, resulting in the following 4 attribute sets:

- All 51 attributes
- 3 attributes selected by CfsSubsetEval
- 34 attributes selected by ChiSquaredAttributeEval
- 44 attributes selected by OneRAttributeEval

3.4 Parameters for Analysis

For the purpose of this study, we will focus on 3 comparison parameters, namely accuracy, recall and area under curve (AUC). Accuracy is the combined percentage of identifying pass students and fail students accurately. Recall is defined by dividing the number of students predicted to fail and actually failed by the total number of students predicted to fail. AUC is a method commonly used to compare the effectiveness of different classifier, with a higher percentage being the better choice. WEKA provides these 3 parameters for us to perform analysis in our study, which will be discussed in detail in the next segment.

4 Results and Discussion

From Table 3 we can observe that ADTree produces the best accuracy (90.6%), recall (25.0%) and AUC (92.0%) when it uses 3 attributes selected by CfsSubsetEval. PART produces the best accuracy (84.4%), recall (50.0%) and AUC (79.5%) when it uses 34 attributes selected by ChiSquaredAttributeEval. Across all classification algorithms, when all attributes are used, accuracy ranges from 75.0% to 81.3%, recall remains at 0%, and AUC ranges from 35.3% to 67.4%. Using attributes selected by CfsSubsetEval, there is an obvious increase in accuracy (78.1%–90.6%), recall (0%–25.0%) and AUC (36.6%–92.0%). Even similar results occurred with attributes selected by ChiSquaredAttributeEval with accuracy (75.0%–84.4%), recall (0%–50.0%) and AUC (36.6%–79.5%). However, attributes selected by OneRAttributeEval (accuracy: 71.9%–81.3%; recall: 0%; AUC: 32.1%–71.9%) displays either stagnation or decrement in their performances. Therefore, the results proved that dimension reduction does improved classifier performance, but it requires choosing the right attributes.

Table 3. Performance of classifiers using attributes selected by different attribute selection algorithms

	All Attributes			**CfsSubsetEval**			**ChiSquaredAttributeEval**			**OneRAttributeEval**		
	Acc	Rec	AUC	Acc	Rec	AUC	Acc	Rec	AUC	Acc	Rec	AUC
Jrip	**81.3**	0.0	37.9	78.1	0.0	36.6	78.1	0.0	37.5	75.0	0.0	36.2
NNge	**81.3**	0.0	46.4	**90.6**	**25.0**	62.5	81.3	0.0	46.4	**81.3**	0.0	46.4
PART	78.1	0.0	58.5	78.1	**25.0**	71.9	**84.4**	**50.0**	**79.5**	75.0	0.0	67.0
Ridor	**81.3**	0.0	46.4	84.4	0.0	48.2	78.1	0.0	44.6	**81.3**	0.0	46.4
ADTree	78.1	0.0	47.3	**90.6**	**25.0**	**92.0**	78.1	0.0	63.4	75.0	0.0	53.6
DecisionStump	78.1	0.0	51.3	78.1	0.0	74.1	78.1	0.0	74.1	78.1	0.0	51.3
J48	78.1	0.0	**67.4**	87.5	0.0	80.4	81.3	0.0	69.6	78.1	0.0	**71.9**
LAD Tree	75.0	0.0	35.3	**90.6**	**25.0**	67.9	75.0	0.0	36.6	71.9	0.0	32.1

Upon evaluating the predictive factors of the classifiers, we now move to the descriptive factors of the classifiers. Since some of the descriptions are rather long, therefore we summarized them in Table 4.

It can be seen that JRip classifier provides different best attribute with 30% to 40% error rate. Therefore, JRip is a not a suitable classifier for descriptive analysis. NNge on the other hand provides description that is too long to be meaningful, unless the number of attributes is 5 or less, as in the attribute set provided by CfsSubsetEval.

Table 4. Analysis of description produced by various classifiers across attribute sets

<table>
<tr><th>Classifiers</th><th>Detailed Analysis</th></tr>
<tr><td>Jrip</td><td>Best attribute selected varied across attribute sets. 2 attribute set didn't contain best attribute. (S1C1_Eng <= 7) => Class = Fail (5.0/2.0).
(Infrastructure = CBA) => Class = Fail (3.0/1.0)</td></tr>
<tr><td>Nnge</td><td>Description provided is too long to be meaningful. The best description is when there are only 3 attributes. class Fail IF: SiblingEducation in {Nil, Diploma,PUC} ^ 5.0 <=S1C1_CF <=10.0 ^ 2.0 <=S1C2_C <=5.0 (4)</td></tr>
<tr><td>PART</td><td>Two main attributes selected are consistent across attribute sets, with a variation that 2 attribute sets provided a tertiary attribute, as observed below.
S1C2_C > 5: Pass (24.0) SiblingEducation = Nil: Fail (2.0)
UGFailure <= 1: Pass (4.0): Fail (2.0)</td></tr>
<tr><td>Ridor</td><td>No best attribute selected. Class = Pass (32.0/0.0)</td></tr>
<tr><td>ADTree</td><td>Between 8 to 10 best attributes are chosen with a numerical factor highlighting the critical level of each rule. All attribute sets nominate S1C2_C as the best attribute. Common rules identified for failure are:
S1C2_C < 5.5: 0.693 S1C1_Math < 12.5: 0.249 S1C2_Math < 9: 0.331
S1C1_C < 7.5: 0.205 SiblingNumber = One: 0.345
SiblingEducation = Nil: 0.205
S1C1_CF < 10.5 and Sports! = Strong and State! = Gujarat and
S1C2_C < 6; 0.408
S1C1_C < 7.5 and S1C2_T < 8.5 and FatherJob! = Private and
S1C1_Math < 12.5: 0.359</td></tr>
<tr><td>DecisionStump</td><td>Consistently identifies S1C2_C as the best attribute in all attribute sets.
S1C2_C <= 5.5 Pass (0.5) Fail (0.5) S1C2_C > 5.5 Pass (1.0) Fail (0.0)
S1C2_C is missing Pass (0.875) Fail (0.125)</td></tr>
<tr><td>J48</td><td>S1C2_C <= 5| SiblingEducation = Nil: Fail (2.0)|
SiblingEducation = Degree: Pass (2.0)| SiblingEducation = Master: Pass (2.0)| SiblingEducation = Diploma: Fail (1.0)| SiblingEducation = PUC: Fail (1.0)
S1C2_C > 5: Pass (24.0)</td></tr>
<tr><td>LADTree</td><td>Between 3 to 9 best attributes are chosen with an immediate numerical factor indicating the significance of the rule. All attribute sets nominate S1C2_C as the best attribute. Common rules identified for failure are:
S1C2_C < 5.5 and S1C2_Eng < 10.5 and S1C2_Math < 9: -1.244
SiblingEducation = Nil: -0.369 S1C2_Math < 9 and State! = Jharkhand: -0.553 S1C2_Math > = 9 and AttendanceMotivation! = APF: -1.8
S1C2_C < 5.5 and Sports! = Strong and SiblingEducation! = Master: -2.833 S1C1_Math < 12.5: -0.214 S1C2_C < 5.5 and
SiblingEducation = Nil: -1.1 S1C1_WT > = 4.5 and
SiblingNumber = One: -0.408</td></tr>
</table>

PART classifier produces rules that are similar across attribute sets. Ridor is the worst classifier as it fails in providing any meaningful description. ADTree consistently promotes S1C2_C (Academic data) as the best attribute for classification, even though it also nominates other attributes or other sets of rules for classification.

DecisionStump consistently identifies S1C2_C as the main attribute in classification throughout the 4 attribute sets. Similarly, J48 also consistently identifies S1C2_C and SiblingEducation (Socio-economic data) as the main attributes in classification in all attribute sets. S1C2_C is the marks obtained in the subject C Programming during the second Internal Assessment test. SiblingEducation refers to the highest academic qualification attained by the sbling of the student. In the final classifier, LADTree, which is very similar to ADTree, projects S1C2_C as the best attribute, with another 2 to 8 other attributes as supporting attributes. Further research with more student records should provide greater detail and accuracy on the rules that define a student who is at risk of failing.

5 Conclusion

With the results from the previous segment, we can conclude that subset-based attribute selection algorithms (CfsSubsetEval, FilteredSubsetEval) are the best and most efficient algorithms to choose significant attributes for classifier development. Simultaneously, this study proves that both academic and socio-economic data are critical in developing effective classifiers. This experiment recognizes that PART and ADTree are the most predictive classifiers when dimension reduction occurs, while ADTree and LADTree are the most descriptive classifiers. However, J48 is the only classifier that provides simple, clear and robust description in all circumstances. In the near future, we look forward to validating these findings with a larger dataset and with ensemble classifiers.

References

1. Romero, C., Ventura, S.: Educational data mining: a survey from 1995 to 2005. Expert Syst. Appl. **33**, 135–146 (2007)
2. Romero, C., Ventura, S.: Educational data mining: a review of the state of the art. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **40**(6), 601–618 (2010)
3. Pena-Ayala, A.: Educational data mining: a survey and a data mining-based analysis of recent works. Expert Syst. Appl. **41**, 1432–1462 (2014)
4. Ganesh, S.H., Christy, A.J.: Applications of educational data mining: a survey. In: 2nd IEEE International Conference on Innovations in Information Embedded and Communication Systems, pp. 1–6. IEEE Xplore (2015)
5. Sukhija, K., Jindal, M., Aggarwal, N.: The recent state of educational data mining: a survey and future visions. In: 3rd IEEE International Conference on MOOCs, Innovation and Technology in Education, pp. 354–359. IEEE Xplore (2015)
6. Anoopkumar, M., Rahman, A.M.J.M.Z.: A review on data mining techniques and factors used in educational data mining to predict student amelioration. In: International Conference on Data Mining and Advanced Computing, pp. 122–133. IEEE Xplore (2016)

7. Lei, C., Li, K.F.: Academic performance predictors. In 29th International Conference on Advanced Information Networking and Applications Workshops, pp. 577–581. IEEE Xplore (2015)
8. Vuttipittayamongkol, P.: Predicting factors of academic performance. In: 2nd Asian Conference on Defence Technology, pp. 161–166. IEEE Xplore (2016)
9. Hasbun, T., Araya, A., Villalon, J.: Extracurricular activities as dropout prediction factors in higher education using decision trees. In: 16th IEEE International Conference on Advanced Learning Technologies, pp. 242–244. IEEE Xplore (2016)
10. Conijn, R., Snijders, C., Kleingeld, A., Matzat, U.: Predicting student performance from LMS data: a comparison of 17 blended courses using moodle LMS. IEEE Trans. Learn. Technol. **10**(1), 17–29 (2017)
11. Pardo, A., Han, F., Ellis, R.A.: Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance. IEEE Trans. Learn. Technol. **10**(1), 82–92 (2017)
12. Cano, A., Zafra, A., Ventura, S.: An ınterpretable classification rule mining algorithm. Inf. Sci. **240**, 1–20 (2013)
13. Huysmans, J., Dejaeger, K., Mues, C., Vanthienen, J., Baesens, B.: An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models. Decis. Support Syst. **51**, 141–154 (2011)
14. Cohen, W..: Fast effective rule induction. In: 12th International Conference on Machine Learning, pp. 115–123 (1995)
15. Martin, B.: Instance-Based learning: Nearest Neighbor with Generalization. Hamilton, New Zealand (1995)
16. Roy, S.: Nearest Neighbor with Generalization. Christchurch, New Zealand (2002)
17. Frank, E., Witten, I.H.: Generating accurate rule sets without global optimization. In: 15th International Conference on Machine Learning, pp. 144–151 (1998)
18. Gaines, B.R., Compton, P.: Induction of ripple-down rules applied to modeling large databases. J. Intell. Inf. Syst. **5**(3), 211–228 (1995)
19. Freund, Y., Mason, L.: The alternating decision tree learning algorithm. In: 16th International Conference on Machine Learning, pp. 124–133. Bled, Slovenia (1999)
20. Witten, I.H., Frank, E., Hall, M.: Data Mining: Practical Machine Learning Tools and Techniques, 3rd edn. Morgan Kaufmann, Burlington (2011)
21. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Francisco (1993)
22. Holmes, G., Pfahringer, B., Kirkby, R., Frank, E., Hall, M.: Multiclass alternating decision trees. In: ECML, pp. 161–172 (2001)