

Ensemble Classifications for Student Academics Performance Data Set

Dr.G.Ayyappan,

Associate Professor, Bharath Institute of Higher Education and Research, Chennai
ayyappangmca@gmail.com

Abstract: The higher education institutes use data mining tools and techniques for academic improvement of the student performance and to prevent drop out. The data consists of socio-economic, demographic as well as academic information of three hundred students with 22 attributes. Five ensemble classification methods Attribute Selected Classifier, Bagging, Classification Via Regression, Weighted Instances Handler Wrapper and Multi Class Classifier were used. The Class Attendance Percentage attribute makes the highest impact in the final semester results of the students in our dataset. The results showed that MultiClassClassifier outperforms the other classifiers based on accuracy and classifier errors.

Keywords: Attribute Selected Classifier, Bagging, Classification Via Regression, Weighted Instances Handler Wrapper and Multi Class Classifier.

I. Introduction

Educational data mining is an emerging field in the area of data mining. In this competitive world, the educational setting also uses data mining tools to explore and analyze student performance, predict their results to prevent drop out and focus on both good and academically poor performers, feedback for the faculties and instructors, visualization of data and to have a better assessment of learning process. The quality of education needs to be improved and educational data mining is a tool for this improvement. Modern educational institutes need data mining for their strategy and future plans. Student's performance depends on various factors like personal, social, economic and other environmental ones[8, 9]. The top-level educational institutes' authorities may utilize the outcome of the experimental results to understand the trends and behaviors in students' performance which may lead to design new pedagogical strategies [10].

In this paper organizes section one has related works and brief introduction of this fields. In section two represents materials and methods, In section three describes results and discussions and the section four presents conclusion

II. Materials and Methods

The students' academic performance data set borrowed from <https://archive.ics.uci.edu/ml/datasets/Student+Academics+Performance> (It was evaluated based on academic and personal data collected from 3 different colleges from Assam, India). The dataset tried to find the end semester percentage prediction based on different social, economic and academic attributes

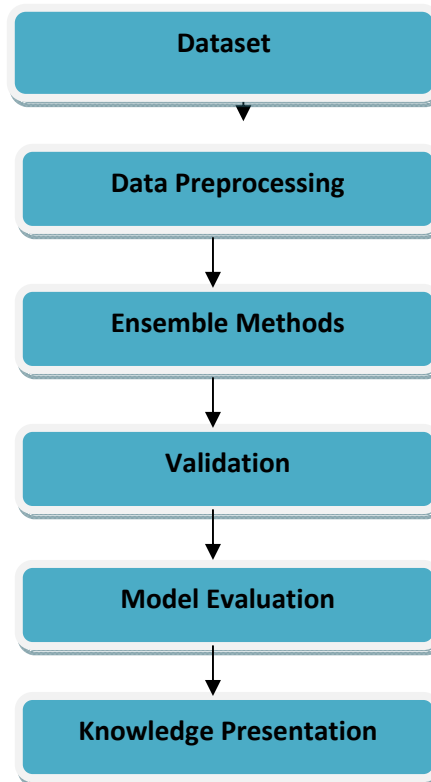


Figure 1 Architecture of Proposed Model

The stage is set for the experiments. WEKA has various classification algorithms. The authors had used AttributeSelectedClassifier, Bagging, ClassificationViaRegression, WeightedInstancesHandlerWrapper and MultiClassClassifier methods available in WEKA. These methods are supervised learning algorithms which use the training data to test the correctness of testing data.

Table 1 Attribute Description in the dataset

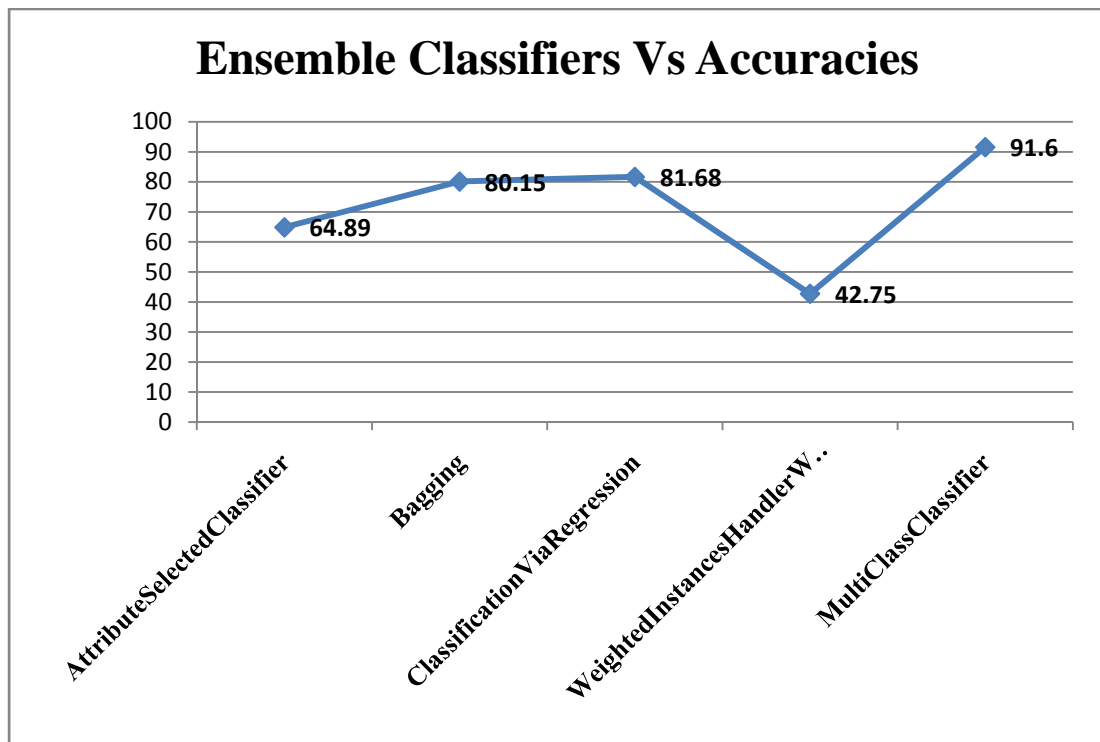
S.No	Attribute	Description	Values
1	GE	Gender	Male, Female
2	CST	Caste	General,SC,ST,OBC,MOBC
3	TNP	Class X Percentage	Best, Very Good, Good, Pass, Fail
4	TWP	Class XII Percentage	Best, Very Good, Good, Pass, Fail
5	IAP	Internal Assessment Percentage	Best, Very Good, Good, Pass, Fail
6	ESP	End Semester Percentage	Best, Very Good, Good, Pass, Fail
7	ARR	Whether the student has back or arrear papers	Yes, No
8	MS	Marital Status	Married, Unmarried
9	LS	Lived in Town or Village	Town, Village
10	AS	Admission Category	Free, Paid
11	FMI	Family Monthly Income in INR	Very High, High, Above Medium, Medium, Low
12	FS	Family Size	Large, Average, Small
13	FQ	Father Qualification	IL, UM, 10, 12 , Degree, PG (IL= Illiterate UM= Under Class X)
14	MQ	Mother Qualification	IL, UM, 10, 12 , Degree, PG
15	FO	Father Occupation	Service, Business, Retired, Farmer, Others
16	MO	Mother Occupation	Service, Business, Retired, Farmer, Others
17	NF	Number of Friends	Large, Average, Small
18	SH	Study Hours	Good, Average, Poor

19	SS	Student School attended at Class X level	Govt., Private
20	ME	Medium	Eng,Asm,Hin,Ben
21	TT	Home to College Travel Time	Large, Average, Small
22	ATD	Class Attendance Percentage	Good, Average, Poor

III. Results and Discussions

In this section presents the ensemble classifiers applied in weka tool by using the student academic performance dataset. In this research work class value consider Class Attendance Percentage attribute. Testing options applied for Training set validation.

S.No	Ensemble Classifier	Accuracy
1	AttributeSelectedClassifier	64.89%
2	Bagging	80.15%
3	ClassificationViaRegression	81.68%
4	WeightedInstancesHandlerWrapper	42.75%
5	MultiClassClassifier	91.60%



The above diagram clearly represents the meta classifications selected few algorithms namely AttributeSelectedClassifier has 64.89% accuracy level, Bagging has 80.15% accuracy level, ClassificationViaRegression has 81.68% accuracy level, WeightedInstancesHandlerWrapper has 42.75% accuracy level and MultiClassClassifier has 91.60% accuracy level. The Bagging, ClassificationViaRegression and MultiClassClassifier have more than 80% accuracy level.

IV. Conclusion

The total numbers of records were 300 with 22 attributes. They were AttributeSelectedClassifier, Bagging, ClassificationViaRegression, WeightedInstancesHandlerWrapper and MultiClassClassifier. The data mining tool used in the experiment was WEKA 3.8.1 Based on the accuracy may conclude that the MultiClassClassifier Classification method was the most suited algorithm for the dataset.

References

- [1] Hussain S, Dahan N.A, Ba-Alwi F.M, Ribata N. Educational Data Mining and Analysis of StudentsTM Academic Performance Using WEKA. Indonesian Journal of Electrical Engineering and Computer Science. 2018; Vol. 9, No. 2. February. pp. 447~459
- [2] Amrieh, E. A., Hamtini, T., &Aljarah, I. 2016. Mining Educational Data to Predict Student's academic Performance using Ensemble Methods. International Journal of Database Theory and Application, 98, 119-136.
- [3] Amrieh, E. A., Hamtini, T., &Aljarah, I. 2015, November. Preprocessing and analyzing educational data set using X-API for improving student's performance. In Applied Electrical Engineering and Computing Technologies AEECT, 2015 IEEE Jordan Conference on pp. 1-5. IEEE.
- [4] Hijazi, S.T. and S.M.M.R. Naqvi, Factors Affecting Students' Performance, A Case Of Private Colleges. Bangladesh e-Journal of Sociology, 2006. 3(1): p. 10.
- [5] Bhardwaj, B.K. and S. Pal, Data Mining: A prediction for performance improvement using classification. International Journal of Computer Science and Information Security, 2011. 9(4): p. 5.
- [6] Strecht, P., et al., A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance. Proceedings of the 8th International Conference on Educational Data Mining, 2015: p. 3.