

An Approach to Predict Student Results using J48 Algorithm and Data Mining Tool

Asha N^{1*}, Suma M²

^{1*} Asst Professor, Department of computer science, Nrupathunga University, Karnataka, India

²Asst Professor, Department of computer science, Nrupathunga University, Karnataka, India

Abstract: One of the most global challenges met by the data analyst is to analyze large amount of stored data and extract essential information. The task of extracting the data is required for future predictions in many sectors of that matter. The predictions are made using forecasting models to take necessary actions for decision policy making. J48 Decision tree algorithm is one of the best techniques to make predictions on nominal data. The paper focuses on implementing and proposes an approach to predict Student results using J48 Algorithm. The entropy of information is measured using student dataset. Based on the predictions made, the educational institutions can make decisions to improve the quality of results. The task is carried out using Data mining tool such as Weka.

Keywords: J48 Algorithm, Decision Tree, Forecasting models, Nominal data, Data mining tool, Weka.

1. INTRODUCTION

“A secret of success is to know something that nobody else knows” said by Aristotle Onassis [1]. It is a known factor that in every sector of Education, Business, Commerce there is large volume of data generated day by day. The Data collected occupies large storage and requires to extract useful information as the data is rich in information. The mining of useful Knowledge to take relevant actions to make Decisions is one of the characteristics of Data mining. There are many Techniques as such in Data mining. The Techniques not only draw out information but also helps in prediction. As the Data in Educational sectors requires clustering of similar data and classify data based on conditions, Classification and Clustering technique of data mining can be used.

Classification is a supervised learning technique that uses a model to assign a specified class label to a data item. The goal of categorization is to forecast future production based on the information available. As a result, educational institutions are attempting to forecast the future output of their enrolled students based on prior and current data. Making classification of Current Student Data one of the most tedious task. So, the paper focuses on the approaches that are better suited to educational research [2].

The outline of this paper works on educational data in a renowned university to make predictions of Student Performance. The courses need to be analysed have a significant impact on final SGPA (Semester Grade Point Average). The Results of VI Sem BSc Students are taken as a case study from Nrupathunga University, Bangalore.

2. RELATED WORK

In this [1], the paper focuses on Educational Data Mining as a technique for predicting a student's success based on their exam results. WEKA tools are utilized for this. The J48 decision tree approach is used to discover categorization rules once the data has been preprocessed.

In this [2] the study aims at gathering transcript data from students, which includes their final GPA and grades in all classes. The J48 decision tree algorithm is applied to discover classification rules. the extracted useful knowledge for final GPA, and identify the most important courses in the students' study plan based on their grades in the mandatory courses.

In this [3] the paper presents a new J48-based decision tree algorithm with improved error trimming. The tree obtained is a rapid decision tree learning tree that will be based on information gain or variance reduction.

3. METHODOLOGY

One of the best Classification techniques used for prediction of nominal data is Decision tree algorithm. The Decision tree algorithm in Weka tool is implemented by the name J48 Algorithm. In Fig 1.0 the last attribute Result column in the Data set is used as a target variable which contains Nominal data (other than Numerical data). Nominal Data in the Result column are Pass, Fail and Tal (To be announced Later). The Data set is supplied with training and test data set to predict



result. The performance of the j28 algorithm measures error difference between the Observed Result versus Predicted Result in terms of Mean Square Error (MSE), Mean Absolute Error (MAE), Root mean Square Error (RMSE), Relative Absolute Error (RAE), Co-relation Coefficient, Total Number of Instances, Correctly and Incorrectly classified Instances.

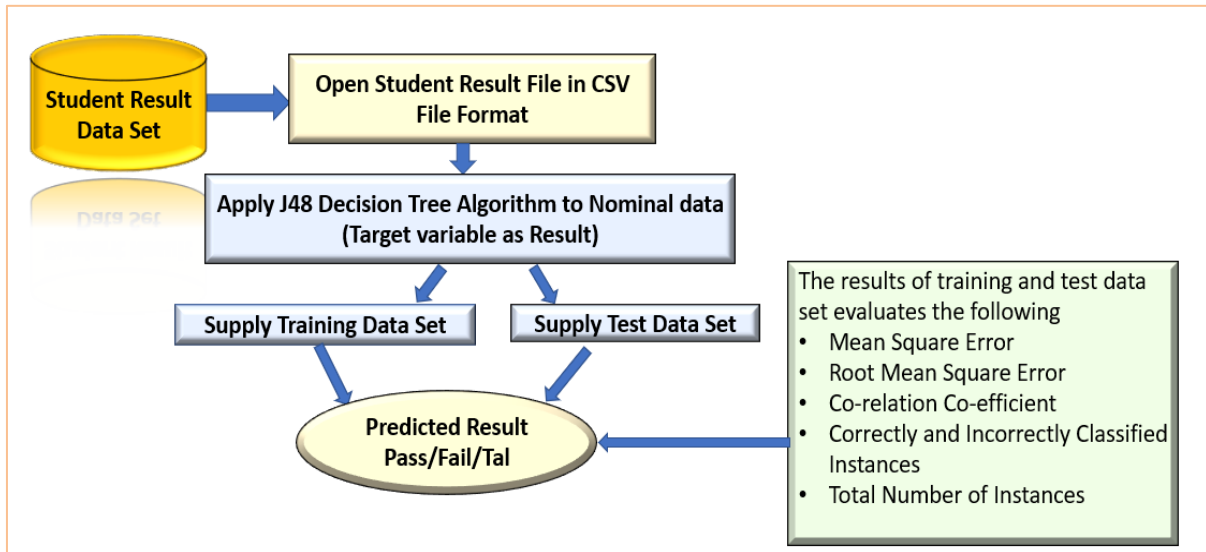


Fig 1.0 Workflow to Run Students Results data in Weka

4. RESULTS AND DISCUSSIONS:

The following Results are observed by giving input file in CSV (Comma Separated Values) file format to J48 Algorithm in Weka tool.

1. Student Data Set used for Prediction
2. Observations of Training and Test data
3. Observations of Observed vs Predicted result of Training and Test data set

4.1 Student Data set used for Prediction

Regno	Name	Sub													Total	SGPA	CGPA	Result
		1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8	Sub9	Sub10	Sub11	Sub12	Sub13				
S160047	Naveen Kumar L	29	41	53	42	56	39	58	48	45	31	54	41	28	2224	0	0	Fail
S160057	Prasanna Kumar K N	51	45	58	35	44	40	56	44	22	34	23	38	50	1410	0	0	Fail
S1600063	Sachin Kumar	48	42	43	37	44	40	51	47	30	34	46	43	50	1826	0	0	Fail
S1600081	Vidyashree	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Tal
S1600092	Gopi R	42	40	57	40	10	1	8	1	50	32	51	35	30	2125	0	0	Fail
S1600111	Sudharshan S	42	49	68	45	62	44	41	38	74	47	71	45	30	3006	6.96	6.37	Pass
S1600164	Swathi R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Tal
S1600241	Tejas M N	13	32	41	38	44	10	40	20	82	75	35	0	0	1937	0	0	Fail
S1600403	Srikanth N	44	37	48	43	63	41	51	47	45	34	45	45	50	3013	6.37	6.51	Pass
S1600468	Sumitra K J	51	43	56	43	49	44	62	39	68	44	69	44	50	2991	7.08	6.31	Pass
S1600475	Gaurav Kumar Sharma S		22	43	33	32	48	38	43	52	25	55	24	45	927	0	0	Fail
S1600478	Jagadeesh Madivalar	54	41	73	45	50	38	59	46	54	37	52	26	30	2248	6.46	0	Pass
S1710010	Charan H S	49	49	59	46	47	48	67	47	48	41	70	49	46	2502	7.04	0	Pass

Fig 1.1 Excel File Format of Student Data set

The Fig 1.1 Represents the student Data set of VI Sem BSc with 19 attributes and 75 instances, which is used as a source of input to Weka to run J48 Algorithm.

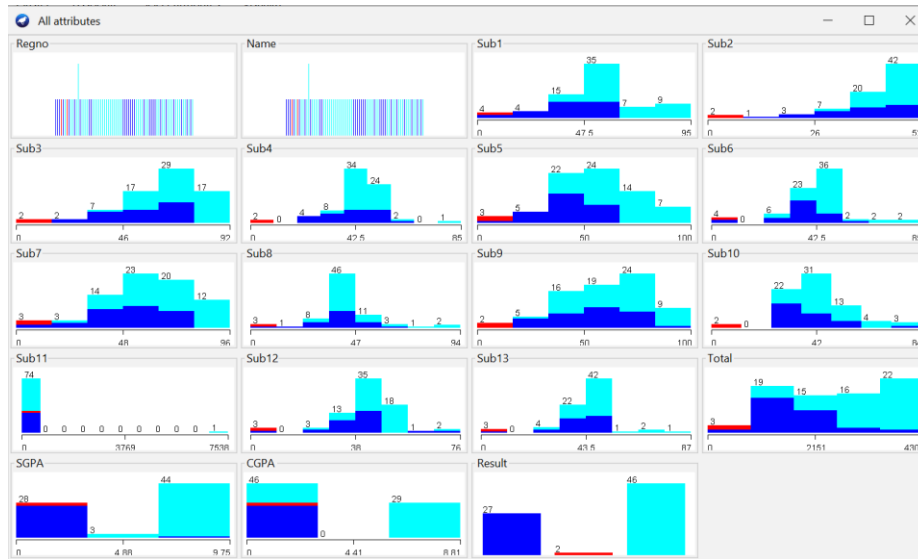


Fig 1.2. Visualization Representation of all instance and values of corresponding attributes

The Fig 1.2 Represents Visualization representation of all the attributes and the number of instances in it.

4.2 Observations of Training and Test data

From the Fig 1.3, it is understood that the entire data set is divided into 2 parts, where 75% of Data set is used for Training and 25% of Data set is used for Testing. The following are the observations made with Mean Absolute Error (MAE), Root mean Square Error (RMSE), Relative Absolute Error (RAE) for 75% of data set for training data set with 45 instances and 25% of data set for testing with 30 instances. The Co-relation coefficient is evaluated which shows 95.5556% for training data set with 2 Incorrectly classified instances and 96.66667% for test data with 1 Incorrectly Classified instance. From the Fig 1.3 it is examined that the J48 algorithm is giving greater value of Co-relation Co-efficient during testing.

Data Set	Correctly Classified Instances	Incorrectly Classified Instances	Mean absolute error	Root mean Squared Error	Relative Absolute Error	Total Number of Instances	Co-relation Co-efficient
Training Data Set	43	2	0.0405	0.1688	11.3889%	45	95.5556%
Test Data Set	29	1	0.0307	0.1466	9.4708	30	96.66667%

Fig 1.3 Performance Evaluation of Training and Test Data sets of Student Data Set

4.3 Observations of Observed vs Predicted result of Training and Test data set

Data Set	Regno	SGPA	Total	Observed Result	Predicted Result
Training Data Set	S1820145	6.83	1125	Fail	Pass
	S1600164	-0.866	0	Tal	Fail
Test Data Set	S1820125	7.71	2649	Pass	Pass
	S1600176	-0.866	0	Tal	Fail

Fig 1.4 Observation of Observed and Predicted Result for two instances with respect to Training and Test Data set

The Fig 1.4 shows the Observed and Predicted result of training and test data set for two instances. The fig shows that in training data set though the observed result is fail, it is showing Pass, Tal is showing as Fail. These are the 2 incorrectly classified instances of J48 algorithm during training, whereas in Test data set one instance is correctly classified as pass and one more instance is incorrectly classified as Fail instead of Tal.

**5.CONCLUSION**

- An Approach of Predicting Student Results is proposed from the methodology
- Also, this methodology could be applied to large data set in educational institutions to take decisions to improve the quality of teaching.
- The management can make a thorough study of student's prediction and adopt new teaching techniques to encourage students by giving extra coaching to fetch good SGPA.
- J48 Decision tree algorithm could be proposed as Best Classification Predictive algorithm.

REFERENCES

- [1] Mehta Smruti Hemantkumar, Dr. Ashish Adholiya Gunasekaran S, Cooper T, Berlage A, Krishnan P. Predicting Students' Performance using J48 Decision Tree. In: Volume 4, Issue 4, March-April -2019, International Journal of Scientific Research in Computer Science, Engineering and Information Technology.
- [2] Mashael A. Al-Barrak and Muna Al-Razgan. Predicting Students Final GPA Using Decision Trees: A Case Study. In: International Journal of Information and Education Technology, Vol. 6, No. 7, July 2016.
- [3] Purna Kapoor, Reena Rani. In: Efficient Decision Tree Algorithm Using J48 and Reduced Error Pruning. In: International Journal of Engineering Research and General Science Volume 3, Issue 3, May-June, 2015 ISSN 2091-2730
- [2] Dinesh Kumar B Vaghela, Priyanka Sharma, "Students' Admission Prediction using GRBST with Distributed Data Mining". In: Communications on Applied Electronics (CAE), Volume 2 – No.1, June 2015.
- [3] Dr. Pranav Patil, "A Study of Student's Academic Performance Using Data Mining Techniques", International Journal Of Research In Computer Applications And Robotics.In: Vol.3, Issue 9, ISSN 2320-7345, September 2015.
- [4] Sen, Umesh Kumar, "A Brief Review Status of Educational Data Mining". In: International Journal of Advanced Research in Computer Science & Technology (IJARCST 2015), Vol. 3, Issue 1 (Jan. – Mar. 2015).
- [5] Jai Ruby, Dr. K. David, "Analysis of Influencing Factors in Predicting Students Performance Using MLP – A Comparative Study". In: Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 2, February 2015.