Enrollment Prediction Models Using Data Mining

Ashutosh Nandeshwar Subodh Chaudhari

April 22, 2009

1 Introduction

Following World War II, a great need for higher education institutions arose in the United States, and the higher education leaders built institutions on "build it and they will come" basis. After the World War II, enrollment in the public as well as the private institutions soared (Greenberg, 2004); however, this changed by 1990s, due to a significant drop in enrollment, universities were in a marketplace with "hypercompetition," and institutions faced the unfamiliar problem of receiving less applicants than they were used to receive (Klein, 2001).

Today higher education institutions are facing the problem of student retention, which is related to graduation rates; colleges with higher freshmen retention rate tend to have higher graduation rates within four years. The average national retention rate is close to 55% and in some colleges fewer than 20% of incoming student cohort graduate (Druzdzel and Glymour, 1994), and approximately 50% of students entering in an engineering program leave before graduation (Scalise et al., 2000).

Tinto (1982) reported national dropout rates and BA degree completions rates for the past 100 years to be constant at 45 and 52 percent respectively with the exception of the World War II period (see Figure 1 for the completion rates from 1880 to 1980). Tillman and Burns at Valdosta State University (VSU) projected lost revenues per 10 students, who do not persist their first semester, to be $326,811. Although gap between private institutions and public institutions in terms of first-year students returning to second year is closing, the retention rates have been constant for a long period for both types of institutions(ACT, 2007). National Center for Public Policy and Higher Education (NCPPHE) reported the U.S. average retention rate for the year 2002 to be 73.6% (NCPPHE, 2007). This problem is not only limited to the U.S. institutions, but also for the institutions in many countries such as U.K and Belgium. The U.K. national average freshmen retention for the year 1996 was 75% (Lau, 2003), and Vandamme (2007) found that 60% of the first generation first-year students in Belgium fail or dropout.

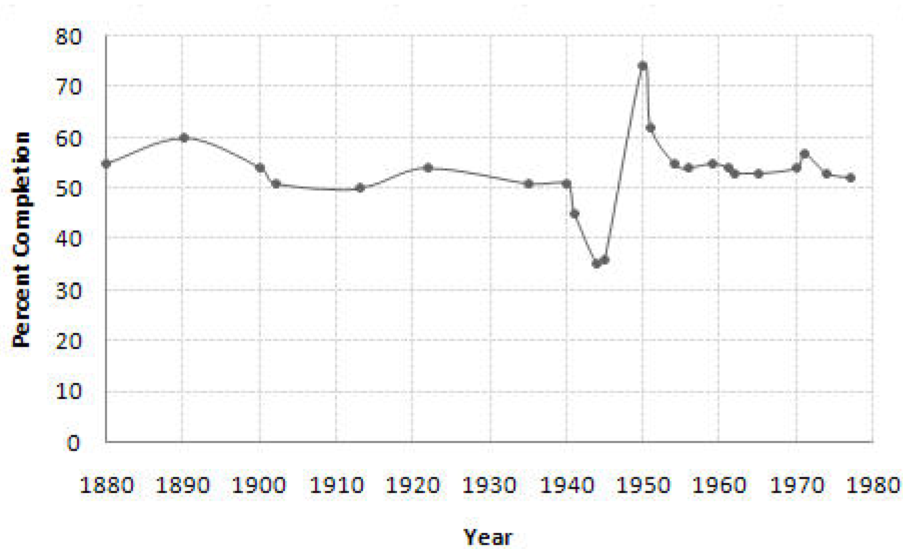


Figure 1: BA Degree Completion Rates for the period 1880 to 1980, where Percent Completion is the Number of BAs Divided by the Number of First-time Degree Enrollment Four Years Earlier (Tinto, 1982)

1.1 Previous Applications of Data Mining

Various researchers have applied data mining in different areas of education, such as enrollment management (Gonzlez and DesJardins, 2002; Chang, 2006; Antons and Maltz, 2006), graduation (Eykamp, 2006; Bailey, 2006), academic performance (Naplava and Snorek, 2001; Pardos et al., 2006; Vandamme, 2007; Ogor, 2007), gifted education (Ma et al., 2000; Im et al., 2005), web-based education (Minaei-Bidgoli et al., 2003), retention (Druzdzel and Glymour, 1994; Sanjeev and Zytkow, 1995; Massa and Puliafito, 1999; Stewart and Levin, 2001; Veitch, 2004; Barker et al., 2004; Salazar et al., 2004; Superby et al., 2006; Sujitparapitaya, 2006; Herzog, 2006; Atwell et al., 2006; Yu et al., 2007; DeLong et al., 2007), and other areas (Intrasai and Avatchanakorn, 1998; Baker and Richards, 1999; Thomas and Galambos, 2004). Luan and Serban (2002) listed some of the applications of data mining to higher education, and provided some case studies to showcase the application of data mining to the student retention problem.. Delavari and Beikzadeh (2004); Delavari et al. (2005) proposed a data mining analysis model to used in higher educational system, which identified various research areas in higher education that could use data mining.

2 Research Objective

Research objectives of this project were:

- To build models to predict enrollment using the student admissions data
- To evaluate the models using cross-validation, win-loss tables and quartile charts
- To present explainable theories to the business users

2.1 Tools Used

DaimlerChrysler (then Daimler-Benz), SPSS (then ISL), and NCR, in 1996, worked together to form the CRoss Industry Standard Process for Data Mining (CRISP-DM). Their philosophy behind creating this standard was to form non-propriety, freely available, and application-neutral standards for data mining. Figure 2 shows CRISP-DM version 1.0, and it illustrates the non-linear (cyclic) nature of data mining.

Standard's phases include, business understanding, data understanding, data preparation, modeling, evaluation, and deployment. This standard was the base of this research, and we created data mining models using Weka, which is a collection of machine learning algorithms for data mining tasks and an open source software. In addition, we used MS Access to import the flat files in database format, modifying and creating new fields, and converting Access tables to ARFF using VBA.

3 Classifiers

3.1 Decision Trees

Decisions tree are a collection of nodes, branches, and leaves. Each node represents an attribute; this node is then split into branches and leaves. Decision trees work on the "divide and conquer" approach; each node is divided, using purity information criteria, until the data are classified to meet a stopping condition. Gini index and information gain ratio are two common purity measurement criteria; Classification and Regression Tree (CART) algorithm uses Gini index, and C4.5 algorithm uses the information gain ratio (Quinlan, 1986, 1996). The Gini index is given by Equation 1, and the information gain is given by Equation 2.

$$I_G(i) = 1 - \sum_{j=1}^{m} f(i,j)^2 = \sum_{j \neq k} f(i,j) f(i,k) \tag{1}$$

$$I_E(i) = -\sum_{j=1}^{m} f(i,j) \log_2 f(i,j) \tag{2}$$

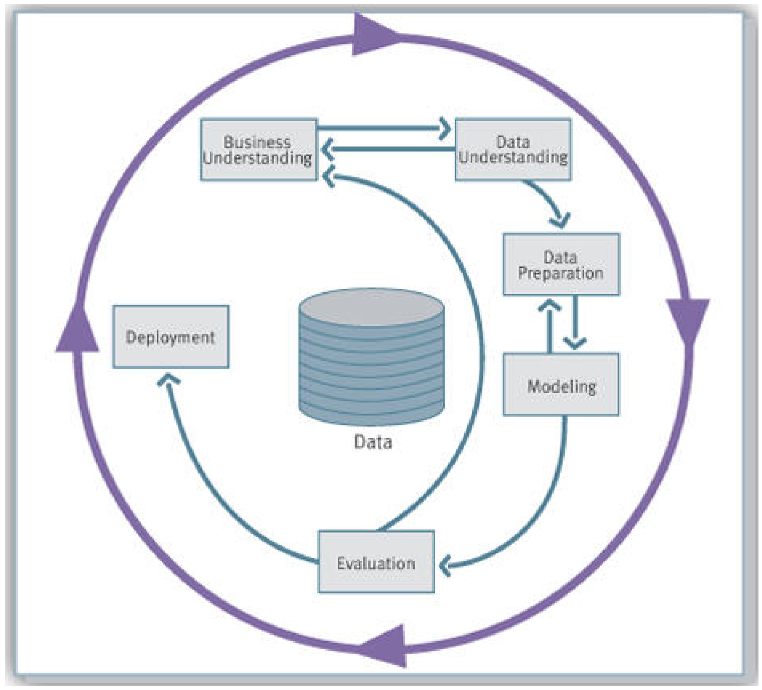


Figure 2: CRISP-DM Model Version 1.0

where, m is the number of values an attribute can take, and $f(i, j)$ is the proportion of class in i that belong to the $j^t h$ class.

Figure 3 is an example of construction decision tree using the Titanic data and Clementine software. Based on the impurity, Clementine selected the attribute sex (male and female) as the root node, then for attribute value sex = male, Clementine created one more split on age (child and adult).

3.2 Rules

Construction of rules is quite similar to the construction of decision trees; however, rules first cover all the instances for each class, and exclude the instances, which do not have class in it. Therefore, these algorithms are called as covering algorithms, and pseudocode of such algorithm is given in Figure 4 reproduced from Witten and Frank (2005). A fictitious example of a rule learner is given in Figure 5.

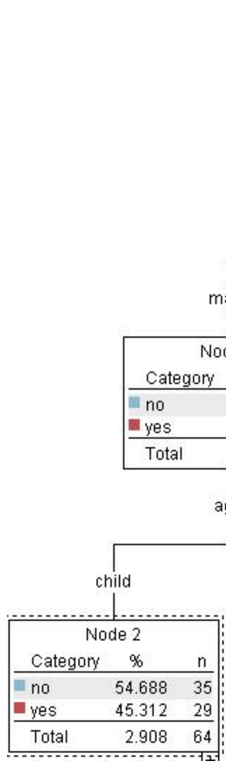


Figure 3: Construction of Decision Tree by Clementine

```
For each class C
  Initialize E to the instance set
  While E contains instances in class C
  Create a rule R with an empty left-hand side that predicts class C
  Until R is perfect (or there are no more attributes to use) do
  For each attribute A not mentioned in R, and each value v,
   Consider adding the condition A=v to the LHS of R
   Select A and v to maximize the accuracy p/t
   (break ties by choosing the condition with the largest p)
   Add A=v to R
   Remove the instances covered by R from E
```

Figure 4: Pseudocode for a Basic Rule Learner

```
IF FinancialAid = "Yes" AND HighSchoolGPA > 3.00
THEN Persistence="Yes"

IF FinancialAid="No" AND HighSchoolGPA < 2.5 AND HoursRegistered < 10
THEN Persistence="No"
```

Figure 5: A Fictitious Example of a Rule Learner

4 Data

The data for this research were from WVU's data warehouse. WVU used SunGard Banner as an Enterprise Resource Planning (ERP) system to run student operations. This system stores the data in a relational database form, which an unit in the Office of Information Technology (OIT) used to SQL queries to obtain data in flat files. This flow of data is represented in Figure 6.

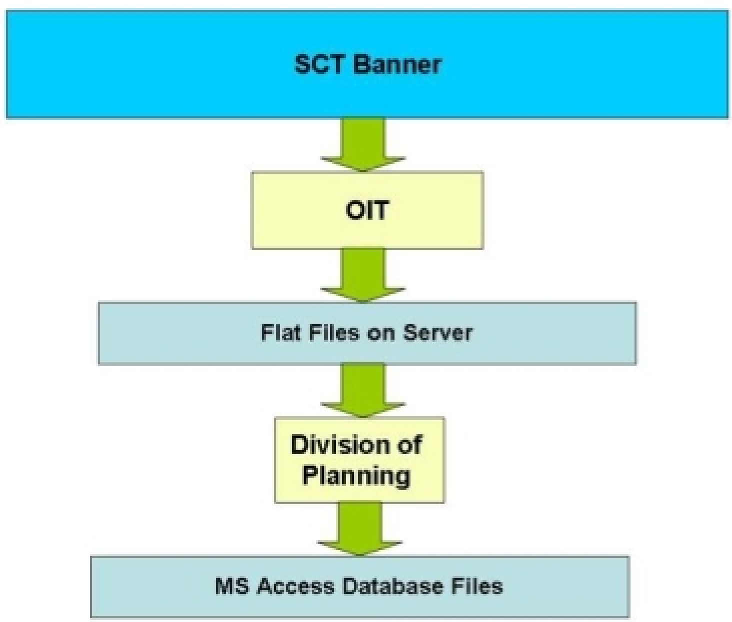


Figure 6: Data Flow of WVU's Admissions Data

4.1 Extraction and Preprocessing

We used admissions data from spring 1999 to fall 2006, and there were approximately 3,000 applications for spring and 25,000 applications for fall. These data contained 248 attributes with demographical and academic information of the applicants. We performed some preprocessing on these data to use them in the modeling process:

- All the data tables were joined to create a single table.
- Flag variables were modified —Enrollment indicator, First Generation indicator, Accepted indicator.
- ACT and SAT scores were combined using concordance tables.
- Permanent address Zip codes were used to create a field Median Family Income from using Zip code and Income data from census.gov website.
- Applications which were not accepted were removed, and the total number of instances reduced to 112,390.
- Domain knowledge and common sense was used to remove some attributes —email address, phone numbers, etc
- Access table was converted to ARFF using VBA script
- String variables were removed using Weka's preprocessing filter: RemoveType string

4.2 Data Visualization

Data visualization in Weka offered interesting insights on these data. For example, Figure 7 shows that 51% of accepted applicants enrolled at WVU, and Figure 8 shows that 92% of accepted applicants who received some form of financial aid enrolled at WVU. Figure 9 shows that 66% of accepted WV residents enrolled at WVU and 62% of accepted non-residents did not enroll at WVU.

5 Experiment

5.1 Feature Subset Selection (FSS)

Feature subset selection is a method to select relevant attributes (or features) from the full set of attributes as a measure of dimensionality reduction. Although some of the data mining techniques, such as decision trees, select relevant attributes, their performance can be improved, as the experiments have shown(Witten and Frank, 2005, p. 288).

Two main approaches of feature or attribute selection are the filters and the wrappers (Witten and Frank, 2005). A filter is an unsupervised attribute selection method, which conducts an independent assessment on general characteristics of the data. It is called as a filter because the attributes are filtered before the learning procedure starts. A wrapper is a supervised attribute selection method, which uses data mining algorithms to evaluate the attributes. It is

Figure 7: Enrolled Indicator

Figure 8: Financial Aid Indicator

Figure 9: Residency Indicator

called as a wrapper because the learning method is wrapped in the attribute selection technique. In an attribute selection method, different search algorithms are employed, such as, genetic algorithm, greedy step-wise, rank search, and others.

For this research, we used Wrapper and InfoGain; Wrapper included J48 tree learner and Naive Bayes learner as part of the attribute selection process. We used these FSS techniques to generate rankings of attributes in order of importance. We then used these rankings for adding attributes in the dataset to evaluate the changes in accuracy on three different learners: J48, Naive Bayes, and RIDOR. To avoid any learning bias, we cross-validated each learning procedure 10 times.

5.2 xval

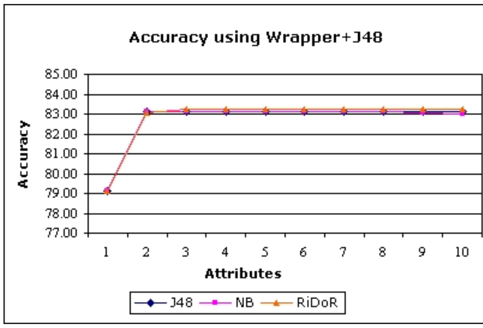
We ran a script called xval, which performed following actions:

- Randomly divided data in two parts: training and testing
- Applied specified discretizers to the datasets
- Applied specified learners to given datasets $< repeat >$ number of times

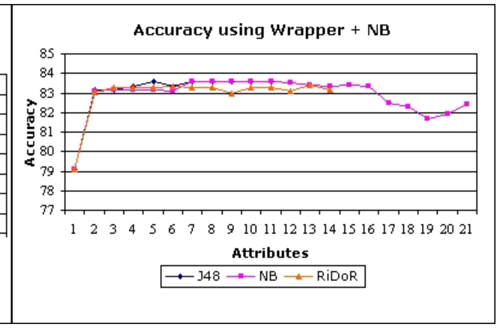
For this experiment, we set value of $< repeat >$ to 10, and we used Nbins and Fayyad-Irani's discretizers. We used five learners: JRip, J48, Aode, Bayes, and OneR.

6 Results

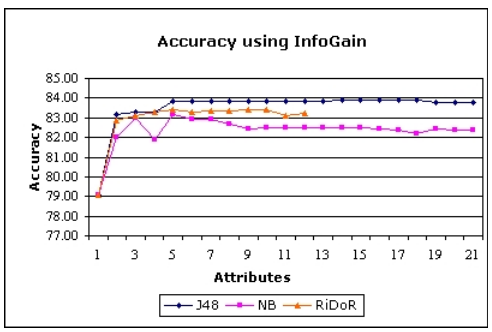
Using the rankings obtained from the FSS experiment, we added each attribute sequentially in the dataset and ran the learning procedure to observe the changes in the accuracy. As shown in Figure 10, accuracy was between 83%-84% for all combinations after adding the variable: FinancialAid Indicator.



(a) Accuracy Using Wrapper and J48



(b) Accuracy Using Wrapper and Bayes



(c) Accuracy Using InfoGain

Figure 10: Results Comparison for FSS, Accuracy, and Number of Attributes

Dataset, Data_WRP_NB_J48, was created with two attributes selected using wrapper, and dataset, Data_IG, was created with seven attributes selected using InfoGain, because the tree size was small with seven attributes (see Figure 11). Figure 12 shows the results obtained by using different learners on the datasets created using Wrapper and InfoGain. RiDor with nbins discretizer was the best for Data_WRP_NB_J48 dataset (highlighted in Figure 12a), and J48 with Fayyad-Irani's discretizer was the best for Data_IG dataset (highlighted in Figure 12b).

There was no significant difference found between these two datasets by any of the learners; however, statistically, by means of t-test with 95% confidence, J48 with Fayyad-Irani was the best and OneR with nbins was the worst, as

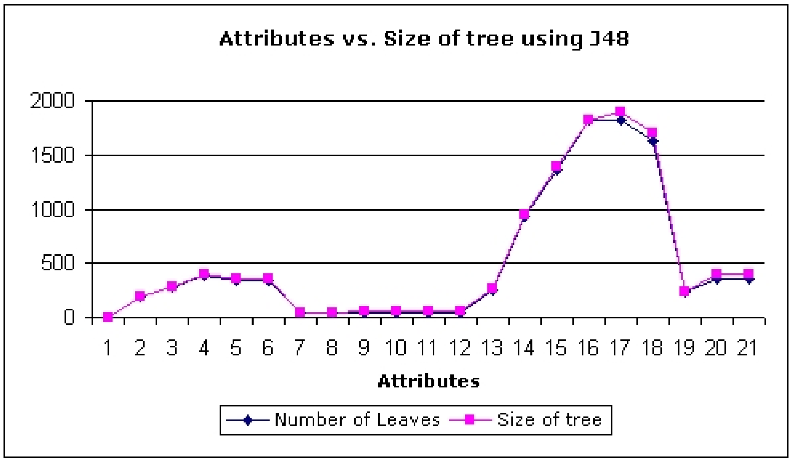


Figure 11: Number of Attributes vs. Tree Size using J48

shown in the win-loss table (Figure 13a). The quartile charts show the margin of win or a loss on another learners, as shown in Figure 13b.

6.1 Learned Theory

Using Win-Loss tables and Quartile charts as guideline for selecting learners and attributes, we found the rules using J48, given in Figure 14, and using RiDor, given in Figure 15.

7 Conclusions

Overall, financial aid was the most important factor that attracted students to enroll. Student enrolled at this institution if they received some form of financial aid, regardless of their high school GPA and ACT/SAT scores. Therefore, financial aid can be used as a controlling factor for increasing the quality of incoming student.

#data	learner	prep	Average of acc	Average of pd	Average of pf
Data_WRP_NB_J48	Aode	fayyadIrani	83.17	87.41	20.27
		myNbins	83.17	87.43	20.28
	bayes	fayyadIrani	83.18	90.18	22.00
		myNbins	83.18	90.18	22.00
	j48	fayyadIrani	83.17	87.41	20.27
		myNbins	83.17	87.43	20.28
	JRip	fayyadIrani	83.17	87.41	20.27
		myNbins	83.17	87.43	20.28
	oneR	fayyadIrani	79.17	92.42	28.15
		myNbins	79.16	92.42	28.16
	Ridor	fayyadIrani	83.17	87.62	20.40
		myNbins	83.18	87.23	20.14
Data_WRP_NB_J48 Average			82.51	88.71	21.88

(a) Results for Wrapper Dataset

#data	learner	prep	Average of acc	Average of pd	Average of pf
Data_IG	Aode	fayyadIrani	83.64	85.63	18.24
		myNbins	83.61	86.82	19.24
	bayes	fayyadIrani	82.98	83.38	17.45
		myNbins	82.96	84.73	18.73
	j48	fayyadIrani	83.89	87.27	19.08
		myNbins	83.64	87.69	19.79
	JRip	fayyadIrani	83.82	86.87	18.89
		myNbins	83.86	86.74	18.73
	oneR	fayyadIrani	79.17	92.41	28.14
		myNbins	79.16	92.42	28.16
	Ridor	fayyadIrani	83.54	85.77	18.40
		myNbins	83.56	85.37	18.11
Data_IG Average			82.82	87.09	20.25

(b) Results for InfoGain Dataset

Figure 12: Pivot Table for Datasets, Learners, and Discretizers

Learner|Discretizer	Wins-Losses	Wins	Losses	Ties
j48|fayyadIrani|2	96	96	0	4
JRip|myNbins|2	91	91	0	9
j48|fayyadIrani|1	83	87	4	9
Aode|myNbins|2	74	83	9	9
JRip|fayyadIrani|2	70	83	13	4
JRip|fayyadIrani|1	52	74	22	4
j48|myNbins|2	43	65	22	13
JRip|myNbins|1	39	65	26	9
Aode|fayyadIrani|2	39	65	26	9
Aode|fayyadIrani|1	17	57	39	4
Ridor|myNbins|2	13	52	39	9
Ridor|myNbins|1	9	52	43	4
j48|myNbins|1	-9	43	52	4
Ridor|fayyadIrani|1	-13	39	52	9
Aode|myNbins|1	-17	39	57	4
bayes|myNbins|2	-35	30	65	4
bayes|fayyadIrani|2	-35	30	65	4
bayes|fayyadIrani|1	-48	26	74	0
bayes|myNbins|1	-57	22	78	0
Ridor|fayyadIrani|2	-65	17	83	0
oneR|fayyadIrani|2	-74	13	87	0
oneR|myNbins|2	-83	9	91	0
oneR|fayyadIrani|1	-91	4	96	0
oneR|myNbins|1	-100	0	100	0

(a) Win-Loss Table

Learner|Discretizer	Median	Quartile Chart
JRip|myNbins|2	0.2	....................|....................
JRip|fayyadIrani|2	0.1	....................|....................
j48|fayyadIrani|2	0.2	....................|....................
j48|fayyadIrani|1	0.2	....................|....................
Aode|myNbins|2	0.2	....................|....................
JRip|myNbins|1	0.1	....................|....................
JRip|fayyadIrani|1	0.1	....................|....................
j48|myNbins|2	0.2	....................|....................
Aode|fayyadIrani|2	0.1	....................|....................
Ridor|myNbins|2	0.2	....................|....................
Ridor|myNbins|1	0	....................|....................
Ridor|fayyadIrani|1	0	....................|....................
j48|myNbins|1	0	....................|....................
Aode|myNbins|1	-0.1	....................|....................
Aode|fayyadIrani|1	0	....................|....................
bayes|myNbins|2	0.1	....................|....................
bayes|fayyadIrani|2	0.1	....................|....................
bayes|fayyadIrani|1	0	....................|....................
bayes|myNbins|1	-0.2	....................|....................
Ridor|fayyadIrani|2	-0.2	....................|....................
oneR|fayyadIrani|2	-3.9	...................|.....................
oneR|myNbins|2	-3.9	...................|.....................
oneR|fayyadIrani|1	-4	...................|.....................
oneR|myNbins|1	-4.2	...................|.....................

(b) Quartile Chart

Figure 13: Win-Loss Table and Quartile Chart

```
FinancialAidIndicator = N
  | ApplicationStypCode = O: N
  | ApplicationStypCode = A: N
  | ApplicationStypCode = B: Y
  | ApplicationStypCode = C: N
  | ApplicationStypCode = D: Y
  | ApplicationStypCode = E: Y
FinancialAidIndicator = Y: Y
Number of Leaves : 7
Size of the tree : 9
Correctly Classified Instances 93448 83.1462%
```

Figure 14: J48 tree with two attributes and accuracy of 83.15%

```
EnrolledIndicator = Y
 Except (FinancialAidIndicator = N) and (ApplicationStypCode = A) =>
EnrolledIndicator = N
 Except (FinancialAidIndicator = N) and (ApplicationStypCode = C) =>
EnrolledIndicator = N
Total number of rules (incl. the default rule): 3
Correctly Classified Instances 93349 83.0581 %
```

Figure 15: Ridor rules with two attributes and accuracy of 83.05%

8 Future Work

Attributes, such as, distance from the campus and first method of contact, should be created to see their effect. Although financial aid was the most significant factor resulting in enrollment, amount of financial aid offered should also be included in the data. So that "bins" can be created on the amount of financial aid offered and then learners can be used for classification using those bins.

Even though financial aid helps recruiting students, it does not necessarily help in retaining the students. In order to find attributes affecting retention, enrolled indicator and "persistence indicator" should be combined. Similar experiments would be necessary to find a relationship between the student demographic, academic information, and retention

9 Acknowledgments

Authors sincerely thank Dr Tim Menzies, our faculty adviser, and Roberta Dean, director of institutional research, both from West Virginia University, for the efforts, the expertise, and the help offered to us for this project.

References

ACT. ACT National Collegiate Retention and Persistence to Degree Rates, 2007. http://www.act.org/research/policymakers/reports/retain.html.

C.M. Antons and E.N. Maltz. Expanding the role of institutional research at small private universities: A case study in enrollment management using data mining. *New Directions for Institutional Research*, 2006(131):69, 2006.

R. H. Atwell, W. Ding, M. Ehasz, S. Johnson, and M. Wang. Using data mining techniques to predict student development and retention. In *Proceedings of the National Symposium on Student Retention*, 2006.

B.L. Bailey. Let the data talk: Developing models to explain IPEDS graduation rates. *New Directions for Institutional Research*, 2006(131):101–11515, 2006.

Bruce D. Baker and Craig E. Richards. A comparison of conventional linear regression methods and neural networks for forecasting educational spending. *Economics of Education Review*, 18(4):405–415, 1999.

K. Barker, T. Trafalis, and T. R. Rhoads. Learning from student data. *Systems and Information Engineering Design Symposium*, pages 79–86, 2004.

L. Chang. Applying data mining to predict college admissions yield: A case study. *New Directions for Institutional Research*, 2006(131), 2006.

N. Delavari and M. R. Beikzadeh. A new analysis model for data mining processes in higher educational systems, 2004.

N. Delavari, M.R. Beikzadeh, and S. Phon-Amnuaisuk. Application of enhanced analysis model for data mining processes in higher educational system. *ITHET* 6^{th}*Annual International Conference* , pages 7–9, July 2005.

C. DeLong, P. M. Radcliffe, and L. S. Gorny. Recruiting for retention: Using data mining and machine learning to leverage the admissions process for improved freshman retention. In *Proceedings of the National Symposium on Student Retention*, 2007.

M. J. Druzdzel and C. Glymour. Application of the TETRAD II program to the study of student retention in u.s. colleges. In *Working notes of the AAAI-94 Workshop on Knowledge Discovery in Databases (KDD-94)*, pages 419–430, Seattle, WA, 1994.

P.W. Eykamp. Using data mining to explore which students use advanced placement to reduce time to degree. *New Directions for Institutional Research*, 2006(131):83, 2006.

J. M. B. Gonzlez and S. L. DesJardins. Artificial neural networks: A new approach to predicting application behavior. *Research in Higher Education*, 43(2):235–258, 2002.

M. Greenberg. How the GI bill changed higher education, June 18, 2004 2004.

S. Herzog. Estimating student retention and degree-completion time: Decision trees and neural networks vis--vis regression. *New Directions for Institutional Research*, 131(2006), 2006.

K. H. Im, T. H. Kim, S. Bae, and S. C. Park. Conceptual modeling with neural network for giftedness identification and education. In *Advances in Natural Computation*, volume 3611, page 530. Springer, 2005.

C. Intrasai and V. Avatchanakorn. Genetic data mining algorithm with academic planning application. In *IASTED International Conference on Applied Modeling and Simulation*, pages 286–129, Alberta, Canada, 1998.

TA. Klein. A fresh look at market segments in higher education. *Planning for Higher Education*, 30(1):5, 2001.

L. K. Lau. Institutional factors affecting student retention. *Education*, 124(1):126–137, 2003.

J. Luan and A. M. Serban. Data mining and its application in higher education. In *Knowledge Management: Building a Competitive Advantage in Higher Education: New Directions for Institutional Research*. Jossey-Bass, 2002.

Y. Ma, B. Liu, C. K. Wong, P. S. Yu, and S. M. Lee. Targeting the right students using data mining. In *Conference on Knowledge Discovery and Data mining*, pages 457–464, Boston, Massachusetts, 2000. ACM Press New York, NY, USA.

S. Massa and P.P. Puliafito. An application of data mining to the problem of the university students' dropout using markov chains. In *Principles of Data Mining and Knowledge Discovery. Third European Conference, PKDD'99*, pages 51–60, Prague, Czech Republic, 1999.

B. Minaei-Bidgoli, D.A. Kashy, G. Kortmeyer, and W.F. Punch. Predicting student performance: an application of data mining methods with an educational web-based system. In *33rd Annual Frontiers in Education*, pages T2A–13–18 Vol.1, Westminster, CO, USA, 2003. IEEE.

P. Naplava and N. Snorek. Modeling of student's quality by means of GMDH algorithms. In *Modelling and Simulation 2001. 15th European Simulation Multiconference 2001. ESM'2001*, pages 696–700, Prague, Czech Republic, 2001.

NCPPHE. Retention rates - first-time college freshmen returning their second year (ACT), 2007.

E.N. Ogor. Student academic performance monitoring and evaluation using data mining techniques. *Electronics, Robotics and Automotive Mechanics Conference, 2007. CERMA 2007*, pages 354–359, 2007.

Z.A. Pardos, N.T. Heffernan, B. Anderson, and C.L. Heffernan. Using fine grained skill models to fit student performance with bayesian networks. In *8th International Conference on Intelligent Tutoring Systems (ITS 2006)*, pages 5–12, Jhongli, Taiwan, 2006.

J. R. Quinlan. Induction of decision trees. *Machine Learning*, 1(1):81–106, 1986.

J. R. Quinlan. Improved use of continuous attributes in C4. 5. *Journal of Artificial Intelligence Research*, 4:77–90, 1996.

A. Salazar, J. Gosalbez, I. Bosch, R. Miralles, and L. Vergara. A case study of knowledge discovery on academic achievement, student desertion and student retention. *Information Technology: Research and Education, 2004. ITRE 2004. 2nd International Conference on*, pages 150–154, 2004.

A.P. Sanjeev and J.M. Zytkow. Discovering enrolment knowledge in university databases. In *First International Conference on Knowledge Discovery and Data Mining*, pages 246–51, Montreal, Que., Canada, 1995.

A. Scalise, M. Besterfield-Sacre, L. Shuman, and H. Wolfe. First term probation: models for identifying high risk students. In *30th Annual Frontiers in Education Conference*, pages F1F/11–16 vol.1, Kansas City, MO, USA, 2000. Stripes Publishing.

D. L. Stewart and B. H. Levin. A model to marry recruitment and retention: A case study of prototype development in the new administration of justice program at blue ridge community college, 2001.

S. Sujitparapitaya. Considering student mobility in retention outcomes. *New Directions for Institutional Research*, 2006(131), 2006.

J. F. Superby, J. P. Vandamme, and N. Meskens. Determination of factors influencing the achievement of the first-year university students using data mining methods. In *8th International Conference on Intelligent Tutoring Systems (ITS 2006)*, pages 37–44, Jhongli, Taiwan, 2006.

E. H. Thomas and N. Galambos. What satisfies students? mining student-opinion data with regression and decision tree analysis. *Research in Higher Education*, 45 (3):251–269, 2004.

C. Tillman and P. Burns. Presentation on First Year Experience. `http://www.valdosta.edu/~cgtillma/powerpoint.ppt`.

V. Tinto. Limits of Theory and Practice in Student Attrition. *The Journal of Higher Education*, 53(6):687–700, 1982.

J.P. Vandamme. Predicting Academic Performance by Data Mining Methods. *Education Economics*, 15(4):405–419, 2007.

W. R. Veitch. Identifying characteristics of high school dropouts: Data mining with a decision tree model, 2004.

I.H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers, San Francisco, 2 edition, 2005.

Chong Ho Yu, Samuel DiGangi, Angel Jannasch-Pennell, Wenjuo Lo, and Charles Kaprolet. A data-mining approach to differentiate predictors of retention between online and traditional students, 2007.